

การเปรียบเทียบวิธีการระบุค่าพารามิเตอร์ในวิธี DBSCAN

นางสาววรรณัญ โกวินทวนิชกานนท์



จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2558

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

THE COMPARISON OF PARAMETER SPECIFICATION METHODS IN DBSCAN ALGORITHM

Miss Woranan Kowitwanichkanont



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2015
Copyright of Chulalongkorn University

วรรณัญ โกวิทวนิชกานนท์ : การเปรียบเทียบวิธีการระบุค่าพารามิเตอร์ในวิธี DBSCAN (THE COMPARISON OF PARAMETER SPECIFICATION METHODS IN DBSCAN ALGORITHM) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: อ. ดร. อัครินทร์ ไพบูลย์พานิช, 222 หน้า.

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการระบุค่าพารามิเตอร์ Eps ในเทคนิคการจัดกลุ่มแบบ DBSCAN ซึ่งวิธีที่ใช้ในการเปรียบเทียบมี 11 วิธี ได้แก่ วิธีของ Daszykowski วิธี mean วิธี median วิธี P75 วิธี P95 วิธีของ Xia วิธีที่ประยุกต์จากวิธีของ Xia (Xia mean Xia median Xia P75 และ Xia P95) และวิธีของ Karami โดยพิจารณาประสิทธิภาพกับข้อมูล 2 มิติ ซึ่งเป็นข้อมูลที่มีการแจกแจงไม่ปกติ 4 รูปแบบ ได้แก่ ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มหลากหลาย ข้อมูลวงแหวนรูปแบบที่ 1 ข้อมูลวงแหวนรูปแบบที่ 2 และข้อมูลมีลักษณะเป็นมุม และข้อมูลที่มีการแจกแจงปกติ 24 รูปแบบ ผู้วิจัยใช้เกณฑ์วัดประสิทธิภาพการวิเคราะห์กลุ่มที่ประยุกต์มาจากค่า purity เรียกว่า modified purity ในการเปรียบเทียบประสิทธิภาพการจัดกลุ่มของ 11 วิธีดังกล่าว ผลการศึกษาพบว่าทุกชุดข้อมูลจะมีวิธีที่เหมาะสมแตกต่างกันตามช่วงของ MinPts ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มหลากหลาย วิธีการที่เหมาะสมคือวิธี Daszykowski วิธี Karami และวิธี P95 ส่วนข้อมูลวงแหวนรูปแบบที่ 1 และรูปแบบที่ 2 ทุกวิธีการเหมาะสมแต่แตกต่างกันตามช่วงของ MinPts ส่วนข้อมูลมีลักษณะมุม วิธีการที่เหมาะสมคือวิธี Daszykowski วิธี Karami วิธี P95 และวิธี Xia P95 นอกจากนี้ข้อมูลที่มีการแจกแจงปกติ วิธีการที่เหมาะสมคือ K-means นอกจากนี้เมื่อตรวจสอบผลวิธีการคำนวณค่า MinPts จากวิธี Daszykowski พบว่าสำหรับข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน เมื่อใช้ค่า MinPts ที่ได้จากวิธี Daszykowski มีวิธีที่เหมาะสมได้แก่ P75 และ Xia P75 แต่สำหรับข้อมูลวงแหวนรูปแบบที่ 1 ข้อมูลวงแหวนรูปแบบที่ 2 และข้อมูลมีลักษณะเป็นมุม เมื่อใช้ค่า MinPts ที่ได้จากวิธี Daszykowski พบว่าผลการจัดกลุ่ม DBSCAN ไม่เหมาะสม ส่วนข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร พบว่าโดยส่วนใหญ่วิธีระบุค่าพารามิเตอร์ Eps ที่เหมาะสมได้แก่วิธี mean และ Xia mean อย่างไรก็ตามผลการจัดกลุ่ม DBSCAN ก็ยังไม่เหมาะสมเท่ากับวิธี K-means

ภาควิชา สถิติ

ลายมือชื่อนิสิต

สาขาวิชา สถิติ

ลายมือชื่อ อ.ที่ปรึกษาหลัก

ปีการศึกษา 2558

5681581526 : MAJOR STATISTICS

KEYWORDS: CLUSTERING ANALYSIS / DBSCAN / EPS / MINPTS

WORANAN KOWITWANICHKANONT: THE COMPARISON OF PARAMETER SPECIFICATION METHODS IN DBSCAN ALGORITHM. ADVISOR: AKARIN PHAIBULPANICH, Ph.D., 222 pp.

The purpose of this research is to compare the performances of 11 Eps identification methods in DBSCAN algorithm. The 11 methods are the Daszykowski method, the mean method, the median method, the P75 method, the P95 method, the method of Xia, the modified method of Xia (Xia mean, Xia median, Xia P75 and Xia P95), and the method of Karami. We apply these methods to 4 different simulated 2 dimensional data sets: one with complex shapes, where the shapes and number of samples in each group are different, ring data type I, ring data type II and corner data, and 24 data sets generated from bivariate normal distributions. We use “modified purity” adapted from purity as a performance measure. The result shows that for all data sets, there are some proper methods at different range of MinPts. For the data with different shapes and sizes, the preferred approach is the method of Daszykowski, the method of Karami and the method of P95. For ring data type I and II, all methods are suitable but differ according range of MinPts. For corner data, the proper approach is the method of Daszykowski, the method of Karami, the method of P95 and the method of Xia P95. For normal distributed data, the proper approach is the method of K-means. Moreover, when calculating the MinPts from Daszykowski method found for the data with different shapes and sizes, the preferred approaches are the method of P75 and the method of Xia P75. But for ring data type I, ring data type II and corner data, the MinPts from Daszykowski method does not produce clustering that fits the data. Furthermore, for bivariate normal data, the preferred Eps identification methods are the method of mean and the method of Xia mean. However, the results of DBSCAN clustering are inferior to K-means.

Department: Statistics

Student's Signature

Field of Study: Statistics

Advisor's Signature

Academic Year: 2015

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ลงได้ด้วยดี ด้วยความช่วยเหลือและเอาใจใส่จาก อาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณ ท่านอาจารย์เป็นอย่างสูงที่กรุณาให้คำปรึกษา อบรมสั่งสอน และให้ข้อคิดเห็นต่าง ๆ ตลอดจน ให้ความช่วยเหลือ คำแนะนำเพื่อปรับปรุงแก้ไขวิทยานิพนธ์ และเป็นกำลังใจในการทำงาน จนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์ด้วยดี

ผู้วิจัยขอกราบขอบพระคุณท่าน รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์ ประธานกรรมการสอบวิทยานิพนธ์ อาจารย์ ดร. นัท กุลวานิช และอาจารย์ ดร. อรุณี กำลัง กรรมการสอบวิทยานิพนธ์เป็นอย่างสูงที่ท่านอาจารย์ทั้งสามท่านได้เสียสละเวลาเพื่อสอบ ตรวจสอบ และให้คำแนะนำเพื่อแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น อีกทั้งขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์ มหาวิทยาลัยทุกท่านที่ให้โอกาสทางการศึกษา และอบรมสั่งสอนให้ความรู้ทั้งในการเรียนและการ ดำรงชีวิตให้แก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษาในครั้งนี้

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณครอบครัว ที่คอยให้กำลังใจและความห่วงใย ส่งเสริม และสนับสนุนมาโดยตลอด และขอขอบคุณเพื่อน ๆ ทุกคนที่คอยช่วยเหลือ ให้คำแนะนำ และเป็นกำลังใจให้กับผู้วิจัยตลอดมา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฌ
สารบัญรูป.....	ฐ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	5
1.3 ขอบเขตของการวิจัย.....	5
1.4 คำจำกัดความของงานวิจัย.....	7
1.5 วิธีการดำเนินงานวิจัย.....	9
1.6 ประโยชน์ที่คาดว่าจะได้รับ.....	9
บทที่ 2 ทฤษฎีและตัวสถิติที่เกี่ยวข้อง.....	10
2.1 วิธีการจัดกลุ่ม DBSCAN.....	10
2.2 วิธีของ Daszykowski.....	22
2.3 วิธีของ Xia และ Jing.....	23
2.4 วิธีของ Karami และ Johansson.....	24
2.5 วิธี K-means.....	25
2.6 การวัดประสิทธิภาพการจัดกลุ่ม.....	26
บทที่ 3 วิธีดำเนินการศึกษา.....	28
3.1 ขอบเขตของการศึกษา.....	28

3.2 ขั้นตอนในการดำเนินการศึกษา.....	33
3.3 ขั้นตอนการทำงานของโปรแกรม.....	35
บทที่ 4 ผลการวิจัย.....	38
4.1 ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน.....	39
4.2 ข้อมูลวงแหวนรูปแบบที่ 1	45
4.3 ข้อมูลวงแหวนรูปแบบที่ 2	52
4.4 ข้อมูลมีลักษณะเป็นมุม	58
4.5 ข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร	63
4.6 การเปรียบเทียบภายใต้สถานการณ์จำนวนกลุ่มและขนาดตัวอย่างเท่ากันสำหรับข้อมูลที่มี การแจกแจงปกติ	182
4.7 การตรวจสอบผลวิธีการคำนวณค่า MinPts ที่ได้จากวิธี Daszykowski.....	188
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ	191
5.1 สรุปผลการวิจัย.....	191
5.2 ข้อเสนอแนะ	194
รายการอ้างอิง.....	195
ภาคผนวก.....	197
ประวัติผู้เขียนวิทยานิพนธ์	222

สารบัญตาราง

	หน้า
ตารางที่ 2.1.1	ค่าระยะห่างจากจุดเริ่มต้น (1,6) ไปยังจุดปลาย..... 13
ตารางที่ 2.1.2	ค่าระยะห่างจากจุดเริ่มต้น (2,5) ไปยังจุดปลาย..... 14
ตารางที่ 2.1.3	ค่าระยะห่างจากจุดเริ่มต้น (4,4) ไปยังจุดปลาย..... 15
ตารางที่ 2.1.4	ค่าระยะห่างจากจุดเริ่มต้น (6,3) ไปยังจุดปลาย..... 16
ตารางที่ 2.1.5	ค่าระยะห่างจากจุดเริ่มต้น (7,2) ไปยังจุดปลาย..... 17
ตารางที่ 2.1.6	ค่าระยะห่างจากจุดเริ่มต้น (9,2) ไปยังจุดปลาย..... 18
ตารางที่ 2.1.7	ค่าระยะห่างจากจุดเริ่มต้น (11,12) ไปยังจุดปลาย..... 19
ตารางที่ 2.1.8	ค่าระยะห่างจากจุดเริ่มต้น (11,16) ไปยังจุดปลาย..... 19
ตารางที่ 2.1.9	ค่าระยะห่างจากจุดเริ่มต้น (12,10) ไปยังจุดปลาย..... 20
ตารางที่ 2.1.10	ค่าระยะห่างจากจุดเริ่มต้น (12,14) ไปยังจุดปลาย..... 20
ตารางที่ 2.1.11	ค่าระยะห่างจากจุดเริ่มต้น (14,15) ไปยังจุดปลาย..... 21
ตารางที่ 2.1.12	ค่าระยะห่างจากจุดเริ่มต้น (18,1) ไปยังจุดปลาย..... 21
ตารางที่ 3.1.1.1	จำนวนข้อมูลแต่ละกลุ่มสำหรับขนาดตัวอย่างเท่ากับ 300 เมื่อจำนวนกลุ่มเท่ากับ 3..... 30
ตารางที่ 3.1.1.2	จำนวนข้อมูลแต่ละกลุ่มสำหรับขนาดตัวอย่างเท่ากับ 1,200 เมื่อจำนวนกลุ่มเท่ากับ 3..... 30
ตารางที่ 3.1.1.3	จำนวนข้อมูลแต่ละกลุ่มสำหรับขนาดตัวอย่างเท่ากับ 300 เมื่อจำนวนกลุ่มเท่ากับ 5..... 31
ตารางที่ 3.1.1.4	จำนวนข้อมูลแต่ละกลุ่มสำหรับขนาดตัวอย่างเท่ากับ 1,200 เมื่อจำนวนกลุ่มเท่ากับ 5..... 31
ตารางที่ 4.1.1	ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.1..... 43
ตารางที่ 4.1.2	ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.1..... 44
ตารางที่ 4.2.1	ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.2..... 50
ตารางที่ 4.2.2	ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.2..... 51
ตารางที่ 4.3.1	ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.3..... 56

ตารางที่ 4.5.2.7.2	ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูล กรณีนี้ที่ 4.5.2.7	157
ตารางที่ 4.5.2.8.1	ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้ที่ 4.5.2.8..	161
ตารางที่ 4.5.2.8.2	ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูล กรณีนี้ที่ 4.5.2.8	162
ตารางที่ 4.5.2.9.1	ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้ที่ 4.5.2.9..	166
ตารางที่ 4.5.2.9.2	ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูล กรณีนี้ที่ 4.5.2.9	167
ตารางที่ 4.5.2.10.1	ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้ที่ 4.5.2.10	171
ตารางที่ 4.5.2.10.2	ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูล กรณีนี้ที่ 4.5.2.10	172
ตารางที่ 4.5.2.11.1	ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้ที่ 4.5.2.11	176
ตารางที่ 4.5.2.11.2	ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูล กรณีนี้ที่ 4.5.2.11	177
ตารางที่ 4.5.2.12.1	ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้ที่ 4.5.2.12	181
ตารางที่ 4.5.2.12.2	ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูล กรณีนี้ที่ 4.5.2.12	182
ตารางที่ 4.7.1	วิธีการระบุค่าพารามิเตอร์ Eps ที่เหมาะสมและค่า modified purity เมื่อ ใช้ค่า MinPts ที่ได้จากวิธี Daszykowski	188

สารบัญรูป

	หน้า
รูปที่ 1.1.1	ข้อจำกัดของ K-means ในกรณีที่รูปร่างไม่ใช่ทรงกลม 1
รูปที่ 1.1.2	จุดแกนกลาง จุดขอบ และจุดรบกวน..... 3
รูปที่ 1.1.3	การ fit กราฟฮิสโทแกรมของระยะห่างแต่ละจุดไปยังจุดที่ใกล้กันเป็นลำดับที่ k ด้วยการแจกแจงเกาส์เซียนแบบผกผัน 4
รูปที่ 1.4.4.1	Eps-neighborhood ของจุด p..... 7
รูปที่ 1.4.7.1	ความหนาแน่นของการเชื่อมต่อ..... 8
รูปที่ 1.4.8.1	ลักษณะ Center-based density..... 8
รูปที่ 2.1.1	ตัวอย่างการกระจายของข้อมูล..... 11
รูปที่ 2.1.2	แต่ละจุดของข้อมูลมีการกวาดรัศมีไปเป็นระยะ Eps..... 12
รูปที่ 2.1.3	ผลจากการแบ่งกลุ่มข้อมูล 22
รูปที่ 3.1.1.1.1	แสดงข้อมูลไม่ปกติ 2 มิติ 4 ลักษณะ..... 29
รูปที่ 4.1.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.1..... 39
รูปที่ 4.1.2	(A) กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.1 และ (B) กราฟขยายส่วนต้นของรูป (A) โดยแสดงค่า modified purity ด้วยความสว่างของพื้นหลังของข้อมูลกรณีที่ 4.1 40
รูปที่ 4.1.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.1 41
รูปที่ 4.1.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.1..... 41
รูปที่ 4.2.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.2..... 45
รูปที่ 4.2.2	(A) กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.2 และ (B) กราฟขยายส่วนต้นของรูป (A) โดยแสดงค่า modified purity ด้วยความสว่างของพื้นหลังของข้อมูลกรณีที่ 4.2..... 46
รูปที่ 4.2.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.2..... 47
รูปที่ 4.2.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.2..... 47

รูปที่ 4.3.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.3.....	52
รูปที่ 4.3.2	(A) กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.3 และ (B) กราฟขยายส่วนต้นของรูป (A) โดยแสดงค่า modified purity ด้วยความสว่างของ พื้นหลังของข้อมูลกรณีที่ 4.3.....	53
รูปที่ 4.3.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.3.....	54
รูปที่ 4.3.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.3.....	54
รูปที่ 4.4.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.4.....	58
รูปที่ 4.4.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.4	58
รูปที่ 4.4.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.4.....	59
รูปที่ 4.4.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.4.....	59
รูปที่ 4.5.1.1.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.1.....	63
รูปที่ 4.5.1.1.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.1	63
รูปที่ 4.5.1.1.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.1.....	64
รูปที่ 4.5.1.1.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.1	64
รูปที่ 4.5.1.2.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.2.....	68
รูปที่ 4.5.1.2.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.2.....	68
รูปที่ 4.5.1.2.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.2.....	69
รูปที่ 4.5.1.2.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.2	69
รูปที่ 4.5.1.3.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.3.....	73

รูปที่ 4.5.1.3.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.1.3.....	73
รูปที่ 4.5.1.3.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.3.....	74
รูปที่ 4.5.1.3.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.3	74
รูปที่ 4.5.1.4.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.4.....	78
รูปที่ 4.5.1.4.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.1.4.....	78
รูปที่ 4.5.1.4.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.4.....	79
รูปที่ 4.5.1.4.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.4	79
รูปที่ 4.5.1.5.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.5.....	83
รูปที่ 4.5.1.5.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.1.5.....	83
รูปที่ 4.5.1.5.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.5.....	84
รูปที่ 4.5.1.5.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.5	84
รูปที่ 4.5.1.6.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.6.....	88
รูปที่ 4.5.1.6.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.1.6.....	88
รูปที่ 4.5.1.6.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.6.....	89
รูปที่ 4.5.1.6.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.6	89
รูปที่ 4.5.1.7.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.7.....	93
รูปที่ 4.5.1.7.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.1.7.....	93

รูปที่ 4.5.1.7.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.7.....	94
รูปที่ 4.5.1.7.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.7	94
รูปที่ 4.5.1.8.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.8.....	98
รูปที่ 4.5.1.8.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.1.8.....	98
รูปที่ 4.5.1.8.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.8.....	99
รูปที่ 4.5.1.8.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.8	99
รูปที่ 4.5.1.9.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.9.....	103
รูปที่ 4.5.1.9.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.1.9	103
รูปที่ 4.5.1.9.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.9.....	104
รูปที่ 4.5.1.9.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.9	104
รูปที่ 4.5.1.10.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.10.....	108
รูปที่ 4.5.1.10.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.1.10.....	108
รูปที่ 4.5.1.10.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.10	109
รูปที่ 4.5.1.10.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.10	109
รูปที่ 4.5.1.11.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.11.....	113
รูปที่ 4.5.1.11.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.1.11.....	113
รูปที่ 4.5.1.11.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.11	114

รูปที่ 4.5.1.11.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.11	114
รูปที่ 4.5.1.12.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.12.....	118
รูปที่ 4.5.1.12.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.12.....	118
รูปที่ 4.5.1.12.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.12	119
รูปที่ 4.5.1.12.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.12	119
รูปที่ 4.5.2.1.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.1.....	123
รูปที่ 4.5.2.1.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.1	123
รูปที่ 4.5.2.1.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.1.....	124
รูปที่ 4.5.2.1.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.1.....	124
รูปที่ 4.5.2.2.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.2.....	128
รูปที่ 4.5.2.2.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.2	128
รูปที่ 4.5.2.2.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.2.....	129
รูปที่ 4.5.2.2.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.2	129
รูปที่ 4.5.2.3.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.3.....	133
รูปที่ 4.5.2.3.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.3.....	133
รูปที่ 4.5.2.3.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.3.....	134
รูปที่ 4.5.2.3.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.3	134
รูปที่ 4.5.2.4.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.4.....	138

รูปที่ 4.5.2.4.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.2.4.....	138
รูปที่ 4.5.2.4.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.4.....	139
รูปที่ 4.5.2.4.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.4.....	139
รูปที่ 4.5.2.5.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.5.....	143
รูปที่ 4.5.2.5.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.2.5.....	143
รูปที่ 4.5.2.5.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.5.....	144
รูปที่ 4.5.2.5.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.5.....	144
รูปที่ 4.5.2.6.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.6.....	148
รูปที่ 4.5.2.6.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.2.6.....	148
รูปที่ 4.5.2.6.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.6.....	149
รูปที่ 4.5.2.6.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.6.....	149
รูปที่ 4.5.2.7.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.7.....	153
รูปที่ 4.5.2.7.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.2.7.....	153
รูปที่ 4.5.2.7.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.7.....	154
รูปที่ 4.5.2.7.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.7.....	154
รูปที่ 4.5.2.8.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.8.....	158
รูปที่ 4.5.2.8.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.5.2.8.....	158

รูปที่ 4.5.2.8.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.8.....	159
รูปที่ 4.5.2.8.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.8	159
รูปที่ 4.5.2.9.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.9.....	163
รูปที่ 4.5.2.9.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.9	163
รูปที่ 4.5.2.9.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.9.....	164
รูปที่ 4.5.2.9.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.9	164
รูปที่ 4.5.2.10.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.10.....	168
รูปที่ 4.5.2.10.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.10.....	168
รูปที่ 4.5.2.10.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.10	169
รูปที่ 4.5.2.10.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.10	169
รูปที่ 4.5.2.11.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.11.....	173
รูปที่ 4.5.2.11.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.11.....	173
รูปที่ 4.5.2.11.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.11	174
รูปที่ 4.5.2.11.4	แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.11	174
รูปที่ 4.5.2.12.1	ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.12.....	178
รูปที่ 4.5.2.12.2	กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.12.....	178
รูปที่ 4.5.2.12.3	กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.12	179

รูปที่ 4.5.2.12.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีศึกษาที่ 4.5.2.12 179



บทที่ 1

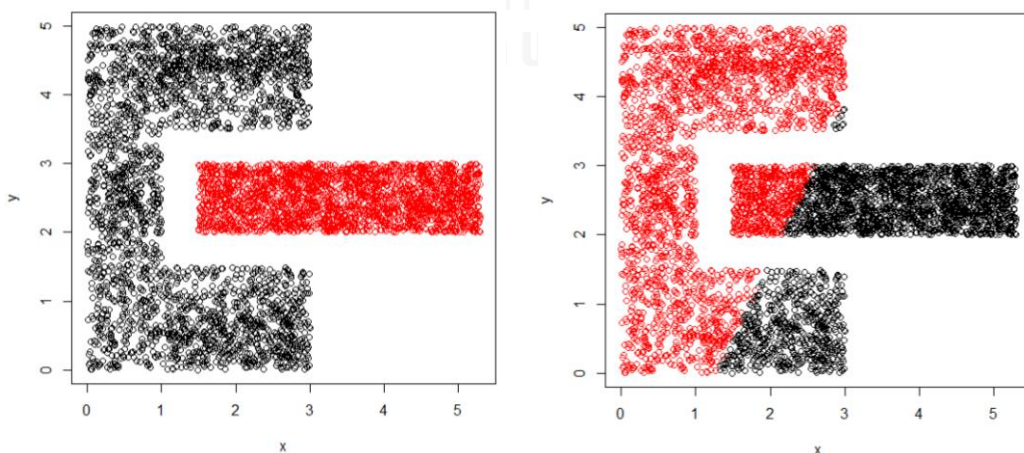
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การวิเคราะห์กลุ่ม (Cluster Analysis) เป็นเทคนิคการแบ่งข้อมูลออกเป็นกลุ่มย่อย โดยหลักการทั่วไปในการแบ่งคือหน่วยที่อยู่ในกลุ่มเดียวกันควรมีลักษณะที่คล้ายกัน หน่วยที่อยู่ต่างกลุ่มกันควรมีลักษณะที่แตกต่างกัน การวิเคราะห์กลุ่มนั้นมีบทบาทมากในการทำเหมืองข้อมูลและมีการใช้งานอย่างกว้างขวางในด้านต่าง ๆ

ในปัจจุบันเทคนิคการวิเคราะห์กลุ่มข้อมูลมีหลากหลายวิธี เช่น การวิเคราะห์กลุ่มแบบขั้นตอน (Hierarchical Cluster Analysis) ด้วยเทคนิค Agglomeration ยังมีข้อจำกัดบางประการ นั่นคือ ถ้าข้อมูลมีขนาดใหญ่ การที่ลดจำนวนกลุ่มลงครั้งละ 1 กลุ่ม ซึ่งใช้เวลาในการคำนวณมาก และผู้วิเคราะห์ต้องตัดสินใจเองว่าควรมีจำนวนกลุ่มเท่าใด (กัลยา วานิชย์บัญชา, 2552)

นอกจากนี้ K-means ซึ่งเป็นหนึ่งในตัวอย่างของการวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน (Nonhierarchical Cluster Analysis) หรือการแบ่งส่วน (Partitioning) โดยผู้วิเคราะห์ต้องกำหนดเองว่าต้องการแบ่งออกเป็นกี่กลุ่ม อัลกอริทึมนี้เป็นอัลกอริทึมที่ง่าย แต่ก็มีปัญหาบางประการ กล่าวคือวิธีการนี้ไม่สามารถจัดการกลุ่มที่มีรูปร่างไม่ใช่ทรงกลมหรือกลุ่มที่มีขนาดหรือความหนาแน่นแตกต่างกันได้ (ชินพัฒน์ แก้วชินพร, 2553)



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 1.1.1 ข้อจำกัดของ K-means ในกรณีที่รูปร่างไม่ใช่ทรงกลม

ตามความเป็นจริงแล้วข้อมูลไม่ได้มีรูปร่างที่เป็นทรงกลมเสมอไป หรือขนาดของกลุ่มข้อมูล แต่แต่ละกลุ่มมีความแตกต่างกันเป็นอย่างมาก ดังตัวอย่างในรูปที่ 1.1.1 ซึ่งแสดงข้อจำกัดของ K-means ในกรณีที่มีรูปร่างไม่ใช่ทรงกลม ด้านซ้ายแสดงการแบ่งกลุ่มของข้อมูลที่ถูกตัด ส่วนด้านขวาแสดงการแบ่งกลุ่มของข้อมูลโดยใช้ K-means ซึ่งจัดกลุ่มได้ไม่ถูกต้องตามความเป็นจริง ในทางปฏิบัติกลุ่มข้อมูลมักจะมีรูปร่างที่ซับซ้อน หรือมีการแจกแจงข้อมูลไม่ปกติ ด้วยเหตุนี้ทำให้ผู้วิจัยมีความสนใจที่จะทำการศึกษาค้นคว้าวิธี Density Based Spatial Clustering of Applications with Noise (DBSCAN) ซึ่งเป็นขั้นตอนวิธีการจัดกลุ่มข้อมูลที่น่าเสนอโดย Martin Ester, Hans-Peter Kriegel, Jörg Sander และ Xiaowei Xu ในปี 1996 DBSCAN นี้เป็นวิธีการจัดกลุ่มที่พิจารณาจากความหนาแน่น ซึ่งมีการประยุกต์ใช้ในด้านต่าง ๆ เช่น ภาพจากดาวเทียม การวิเคราะห์เชิงภูมิศาสตร์ การวิเคราะห์การเดินทางรับส่งบริการรถแท็กซี่ เป็นต้น

DBSCAN เป็นวิธีการจัดกลุ่มซึ่งอาศัยพารามิเตอร์สองตัว ได้แก่ Eps คือรัศมีที่ใช้กำหนด neighborhood ของจุด ๆ หนึ่ง และ MinPts คือจำนวนจุดที่น้อยที่สุดที่อยู่ภายใน neighborhood ของจุด ๆ หนึ่ง ในรัศมี Eps ซึ่งใช้บ่งชี้ว่าจุด ๆ นั้นเป็นจุดชนิดใด

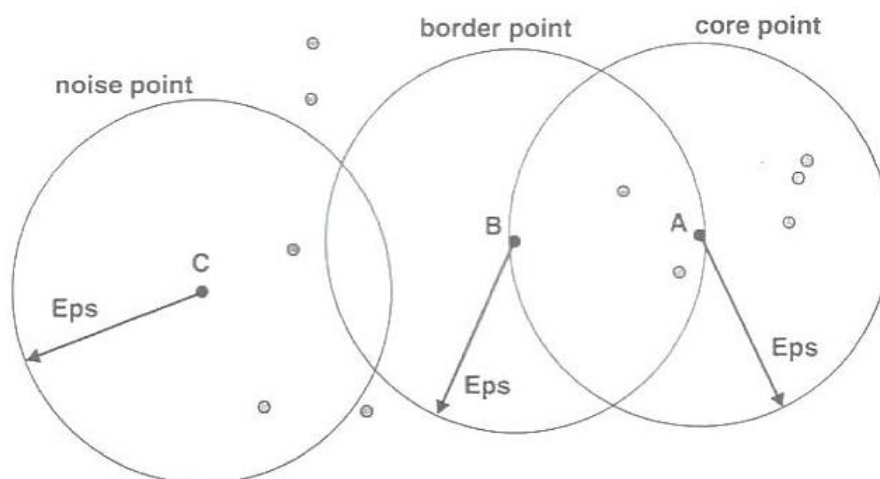
ค่าพารามิเตอร์ทั้งสองดังกล่าวมีความสำคัญเป็นอย่างมาก (ขั้นตอนสำหรับการทำ DBSCAN ในรายละเอียดจะอยู่ในบทที่ 2) กล่าวคือ ถ้า Eps มีขนาดใหญ่เกินไป ชุดข้อมูลนั้น ๆ ก็อาจมีจำนวนกลุ่มน้อยกว่าที่ควรจะเป็น แต่ถ้าหาก Eps มีขนาดเล็กเกินไป ก็จะทำให้มีจำนวนกลุ่มมากเกินไปที่ควรจะเป็น ในกรณีที่ Eps คงที่ ถ้า MinPts มีค่าสูง จะทำให้จุดข้อมูลยากต่อการรวมเป็นกลุ่มเดียวกัน อีกทั้งอาจมีจุดรบกวนเกิดขึ้นมาก แต่ถ้า MinPts มีค่าน้อย จะทำให้จุดข้อมูลสามารถรวมเป็นกลุ่มเดียวกันได้ง่าย

ชนิดของจุดที่เกี่ยวข้องกับ DBSCAN (Tan, Steinbach, & Kumar, 2013) มีดังนี้

จุดแกนกลาง (Core points) เป็นจุดที่มีจำนวนจุดภายในรัศมี Eps มีค่ามากกว่าหรือเท่ากับ MinPts และทุกจุดที่อยู่ภายใน Eps เดียวกันจะให้อยู่กลุ่มเดียวกัน ในข้อมูลชุดหนึ่ง ๆ อาจมีจุดแกนกลางหลายจุดได้

จุดขอบ (Border points) ไม่ใช่จุดแกนกลาง เป็นจุดที่อยู่ภายในรัศมี Eps ของจุดแกนกลาง จุดขอบสามารถตกอยู่ภายในบริเวณใกล้เคียง (neighborhood) ของจุดแกนกลางหลาย ๆ จุดได้ นอกจากนี้ทุกจุดขอบให้อยู่ในกลุ่มเดียวกับจุดแกนกลาง

จุดรบกวน (Noise points) ไม่ใช่ทั้งจุดแกนกลางและจุดขอบ เป็นจุดที่ข้อมูลมีความผิดปกติ จึงไม่มีการระบุว่าอยู่กลุ่มใด



รูปที่ 1.1.2 จุดแกนกลาง จุดขอบ และจุดรบกวน
(Tan et al., 2013)

จากรูปที่ 1.1.2 ถ้า $MinPts = 5$ และ Eps มีค่าตามรูป จะได้ว่าจุด A เป็นจุดแกนกลาง จุด B เป็นจุดขอบ และ จุด C เป็นจุดรบกวน

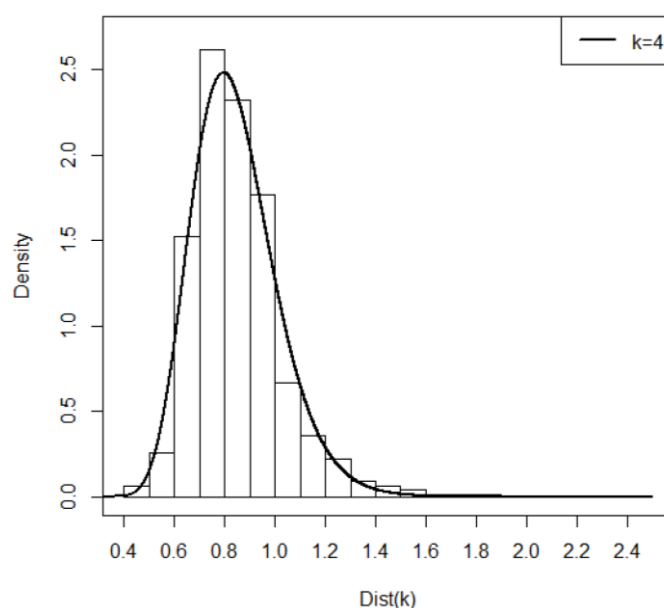
การดำเนินการของ DBSCAN ทำโดยกำหนดค่าพารามิเตอร์ Eps และ $MinPts$ จากนั้นทำการพิจารณาว่าแต่ละจุดเป็นจุดแกนกลาง จุดขอบ หรือจุดรบกวน แล้วดำเนินการเดียวกับกับจุดแกนกลางอื่น ๆ ที่อยู่ภายในรัศมี Eps ของกันและกัน เพื่อกำหนดเป็นกลุ่มใหม่ขึ้นมา ส่วนจุดขอบที่อยู่ภายในรัศมี Eps ของจุดแกนกลางใด ๆ ก็จะถูกจัดอยู่ในกลุ่มเดียวกับกลุ่มของจุดแกนกลางนั้น ๆ ซึ่งงานวิจัยที่ผ่านมาได้มีนักวิจัยนำเสนอวิธีการกำหนดค่าพารามิเตอร์ Eps และ $MinPts$ ดังนี้

Daszykowski, Walczak, and Massart (2001) ได้กำหนดค่า $MinPts$ ให้เป็น

$$MinPts = integer\left(\frac{m}{25}\right)$$

เมื่อ m คือจำนวนจุดในชุดข้อมูล แต่ถ้าข้อมูลมีจำนวนข้อมูลเกิน 500 ควรใช้ $MinPts$ เท่ากับ 20

XIA and JING (2009) ได้นำเสนอ SA-DBSCAN (A self-adaptive density-based clustering algorithm) โดยมีวิธีการหา Eps ด้วยการประมาณการแจกแจงของระยะห่างแต่ละจุดไปยังจุดที่ใกล้กันเป็นลำดับที่ k โดยใช้การแจกแจงเกาส์เซียนแบบผกผัน เพื่อหาค่าฐานนิยมที่ได้นั้นคือ Eps



รูปที่ 1.1.3 การ fit กราฟฮิสโทแกรมของระยะห่างแต่ละจุดไปยังจุดที่ใกล้กันเป็นลำดับที่ k ด้วยการแจกแจงเกาส์เซียนแบบผกผัน

Karami and Johansson (2014) ได้ทำการศึกษาความมีประสิทธิภาพของวิธีจัดกลุ่มลูกผสม เรียกว่า BDE-DBSCAN ซึ่งเป็นการรวมกันระหว่าง Binary Differential Evolution (BDE) และขั้นตอนวิธี DBSCAN การดำเนินการของ BDE-DBSCAN เป็นการหาค่าโดยใช้ชุดข้อมูลที่หลากหลายด้วยความหนาแน่นและรูปร่างที่แตกต่างกันซึ่งมีการกล่าวถึงการระบุค่าพารามิเตอร์ Eps ที่คำนวณได้จากสูตรอีกด้วย

ในงานวิจัยนี้ ผู้วิจัยสนใจศึกษาการหาค่า Eps ซึ่งดัดแปลงมาจากการวิเคราะห์วิธีการของ Xia โดยหา Eps จากค่าเฉลี่ย ค่ามัธยฐาน ค่า quantile ที่ 75 หรือ 95 จากการประมาณฟังก์ชันการแจกแจงของระยะห่างระหว่างข้อมูลโดยใช้การแจกแจงเกาส์เซียนแบบผกผัน โดยวัดประสิทธิภาพจากเกณฑ์ modified purity ซึ่งเป็นมาตรวัดประสิทธิภาพที่ผู้วิจัยได้คิดค้นขึ้นและเปรียบเทียบกับผลที่ได้จากการทำ K-means

1.2 วัตถุประสงค์ของการวิจัย

เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของวิธีที่ใช้ในการระบุค่าพารามิเตอร์ Eps นั้นคือวิธีของ Daszykowski วิธี mean วิธี median วิธี P75 วิธี P95 วิธีของ Xia วิธีที่ประยุกต์จากวิธีของ Xia (Xia mean Xia median Xia P75 และ Xia P95) และวิธีของ Karami ในขั้นตอนวิธี DBSCAN

1.3 ขอบเขตของการวิจัย

1.3.1 ลักษณะข้อมูลที่ศึกษา

1.3.1.1 ข้อมูลไม่ปกติ 2 มิติ 4 ลักษณะ ได้แก่

1. ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน
2. ข้อมูลวงแหวนรูปแบบที่ 1
3. ข้อมูลวงแหวนรูปแบบที่ 2
4. ข้อมูลมีลักษณะเป็นมุม

1.3.1.2 ข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร แบ่งออกเป็นกรณีจำนวนกลุ่มเท่ากับ 3 และจำนวนกลุ่มเท่ากับ 5

ในการศึกษานี้จะทำการศึกษาภายใต้ขอบเขต ดังต่อไปนี้

1. จำนวนกลุ่มเท่ากับ 3 ($K = 3$)

กำหนดให้ $X \sim N_2(\mu_i, \Sigma_i)$, $i = 1, 2, 3$ โดยที่

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \text{ และ } \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

และพิจารณาเมทริกซ์ความแปรปรวนในสองกรณีย่อย ได้แก่ กรณีความเมทริกซ์แปรปรวนเท่ากัน และเมทริกซ์แปรปรวนไม่เท่ากัน

กรณี 1 เมทริกซ์ความแปรปรวนเท่ากัน

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

กรณี 2 เมทริกซ์ความแปรปรวนไม่เท่ากัน

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \text{ และ } \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

รวมถึงพิจารณาขนาดตัวอย่างในสองกรณีย่อย ได้แก่ $n=300$ และ

$n=1,200$

2. จำนวนกลุ่มเท่ากับ 5 ($K = 5$)

กำหนดให้ $X \sim N_2(\mu_i, \Sigma_i)$, $i = 1, 2, 3, 4, 5$ โดยที่

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix} \text{ และ } \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$$

และพิจารณาเมทริกซ์ความแปรปรวนในสองกรณีย่อย ได้แก่ กรณีความเมทริกซ์แปรปรวนเท่ากัน และเมทริกซ์แปรปรวนไม่เท่ากัน

กรณี 1 เมทริกซ์ความแปรปรวนเท่ากัน

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

กรณี 2 เมทริกซ์ความแปรปรวนไม่เท่ากัน

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \text{ และ } \Sigma_4 = \Sigma_5 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

รวมถึงพิจารณาขนาดตัวอย่างในสองกรณีย่อย ได้แก่ $n=300$ และ $n=1,200$

1.3.2 วิธีการระบุค่าพารามิเตอร์ Eps มีดังนี้

1.3.3.1 วิธี Daszykowski วิธีที่ใช้ระยะทางจากทุกจุดในข้อมูลไปยังจุดที่ห่างออกไปเป็นลำดับที่ k มาประกอบ โดยสุ่มให้มีการกระจายแบบสม่ำเสมอทั่วทั้งบริเวณ (Daszykowski et al., 2001)

1.3.3.2 วิธี Xia วิธีที่ได้จากการ fit กราฟฮิสโทแกรมระยะห่างด้วยการแจกแจงเกาส์เซียนแบบผกผัน ด้วยค่าเฉลี่ย μ และพารามิเตอร์รูปร่าง (shape parameter) λ (XIA & JING, 2009)

1.3.3.3 วิธีที่ประยุกต์จากวิธีของ Xia ปรับหาค่า Eps นอกเหนือจากสูตรผ่านกระบวนการ fit ด้วยการแจกแจงเกาส์เซียนแบบผกผัน โดยใช้เป็นค่าเฉลี่ย ค่ามัธยฐาน ค่า quantile ที่ 75 และ 95

1.3.3.4 วิธี Karami เป็นวิธีที่ใช้การคำนวณจากสูตร (Karami & Johansson, 2014)

1.3.3.5 วิธีที่คำนวณจากระยะทางของแต่ละจุดไปยังทุกจุดมาคำนวณเป็นค่าเฉลี่ย ค่ามัธยฐาน ค่า quantile ที่ 75 และ 95 เพื่อนำค่าดังกล่าวเป็นค่า Eps

1.3.3 จำลองข้อมูลเชิงตัวเลข และประมวลผลโดยใช้โปรแกรม R เวอร์ชัน 3.1.3

1.3.4 กำหนดการจำลองข้อมูล 500 ชุดต่อ 1 กรณี

1.4 คำจำกัดความของงานวิจัย

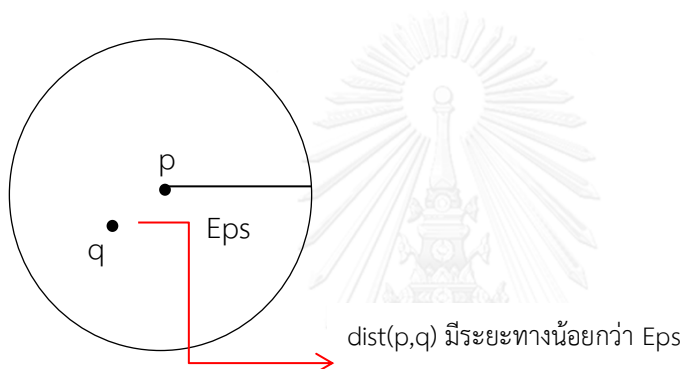
1.4.1 เอพซิลอน (Eps) คือรัศมีที่ใช้กำหนด neighborhood ของจุด ๆ หนึ่ง

1.4.2 จำนวนจุดที่น้อยที่สุด (MinPts) คือจำนวนจุดที่น้อยที่สุดที่อยู่ภายใน neighborhood ของจุด ๆ หนึ่ง ในรัศมี Eps ซึ่งใช้บ่งชี้ว่าจุด ๆ นั้นเป็นจุดแกนกลางหรือไม่

1.4.3 ความหนาแน่น (density) คือจำนวนของจุดภายในรัศมีที่กำหนด (Eps)

1.4.4 Eps-neighborhood ของจุด p แทนด้วย $N_{Eps}(p)$ ซึ่งกำหนดโดย

$$N_{Eps}(p) = \{q \in D \mid \text{dist}(p, q) \leq Eps\}$$



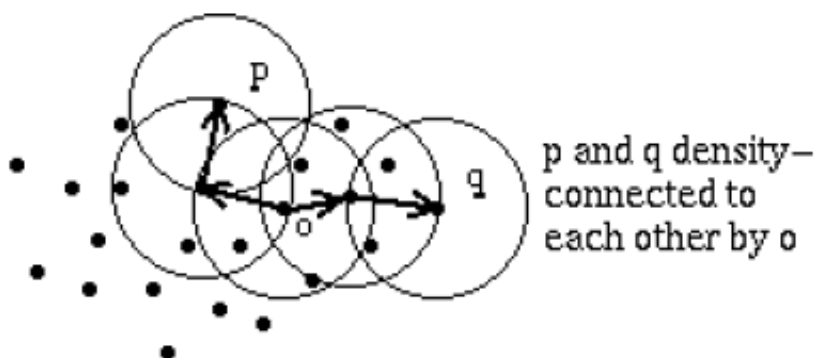
รูปที่ 1.4.4.1 Eps-neighborhood ของจุด p

1.4.5 ความหนาแน่นสามารถเข้าถึงได้โดยตรง (directly density-reachable) จุด p เป็นจุดที่มีความหนาแน่นสามารถเข้าถึงได้โดยตรงจากจุด q ถ้า

- 1) $p \in N_{Eps}(q)$ และ
- 2) $|N_{Eps}(q)| \geq \text{MinPts}$ (เงื่อนไขจุดแกนกลาง)

1.4.6 ความหนาแน่นสามารถเข้าถึงได้ (density-reachable) จุด p เป็นจุดที่มีความหนาแน่นสามารถเข้าถึงได้จากจุด q ถ้ามีลูกโซ่ของจุด p_1, \dots, p_n กับ $p_1=q$ และ $p_n=p$ ดังกล่าวว่า p_{i+1} เป็นจุดที่มีความหนาแน่นสามารถเข้าถึงได้โดยตรงจากจุด p_i

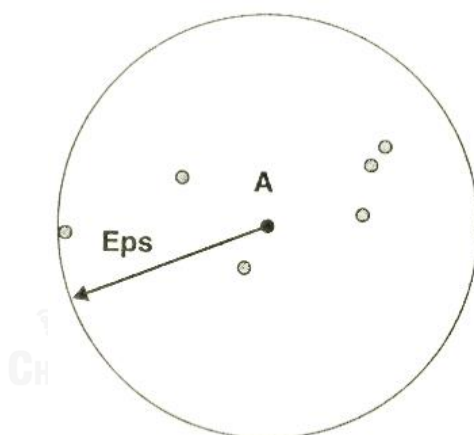
1.4.7 ความหนาแน่นของการเชื่อมต่อ (density-connected) จุด p เป็นจุดที่มีความหนาแน่นของการเชื่อมต่อไปยังจุด q โดยพิจารณา Eps และ MinPts ถ้ามีจุด o ดังกล่าวว่าทั้ง p และ q เป็นจุดที่มีความหนาแน่นสามารถเข้าถึงได้จากจุด o โดยพิจารณา Eps และ MinPts



รูปที่ 1.4.7.1 ความหนาแน่นของการเชื่อมต่อ

(Ester, Kriegel, Sander, & Xu, 1996)

1.4.8 จุดแกนกลาง (Core points) เป็นจุดที่มีจำนวนจุดภายในรัศมี Eps มีค่ามากกว่าหรือเท่ากับ $MinPts$ และทุกจุดที่อยู่ภายใน Eps เดียวกันจะให้อยู่กลุ่มเดียวกัน ในข้อมูลชุดหนึ่ง ๆ อาจมีจุดแกนกลางหลายจุดได้



รูปที่ 1.4.8.1 ลักษณะ Center-based density

(Tan et al., 2013)

จากรูปที่ 1.4.8.1 จุด A เป็นจุดแกนกลางสำหรับรัศมีที่แสดง (Eps) ถ้า $MinPts \leq 7$

1.4.9 จุดขอบ (Border points) ไม่ใช่จุดแกนกลาง เป็นจุดที่อยู่ภายในรัศมี Eps ของจุดแกนกลาง จุดขอบสามารถตกอยู่ภายในบริเวณใกล้เคียง (neighborhood) ของจุดแกนกลางหลาย ๆ จุดได้ นอกจากนี้ทุกจุดขอบให้อยู่ในกลุ่มเดียวกับจุดแกนกลาง

1.4.10 จุดรบกวน (Noise points) ไม่ใช่ทั้งจุดแกนกลางและจุดขอบ เป็นจุดที่ข้อมูลมีความผิดปกติ จึงไม่มีการระบุว่าจะอยู่กลุ่มใด

1.5 วิธีการดำเนินงานวิจัย

1.5.1 กำหนดข้อมูลไม่ปกติใน 4 ลักษณะได้แก่ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน ข้อมูลวงแหวนรูปแบบที่ 1 ข้อมูลวงแหวนรูปแบบที่ 2 ข้อมูลมีลักษณะเป็นมุม และข้อมูลที่มีการแจกแจงแบบปกติ 2 ตัวแปร 24 ลักษณะ

1.5.2 นำข้อมูลแบบต่าง ๆ มาคำนวณหาค่า Eps ด้วยวิธีต่อไปนี้

1.5.2.1 วิธี Daszykowski วิธีที่ใช้ระยะทางจากทุกจุดในข้อมูลไปยังจุดที่ห่างออกไปเป็นลำดับที่ k มาประกอบ โดยสุ่มให้มีการกระจายแบบสม่ำเสมอทั่วทั้งบริเวณ

1.5.2.2 วิธี Xia วิธีที่ได้จากการ fit กราฟฮิสโทแกรมระยะห่างด้วยการแจกแจงเกาส์เซียนแบบผกผัน ด้วยค่าเฉลี่ย μ และพารามิเตอร์รูปร่าง (shape parameter) λ ฐานนิยมที่ได้ นั่นคือ Eps

1.5.2.3 วิธีที่ประยุกต์จากวิธีของ Xia ปรับปรุงหาค่า Eps นอกเหนือจากสูตรผ่านกระบวนการ fit ด้วยการแจกแจงเกาส์เซียนแบบผกผัน โดยใช้เป็นค่าเฉลี่ย ค่ามัธยฐาน ค่า quantile ที่ 75 หรือ 95 ในที่นี้เรียกว่าวิธี Xia mean, Xia median, Xia P75 และ Xia P95 ตามลำดับ

1.5.2.4 วิธี Karami เป็นวิธีที่ใช้การคำนวณจากสูตร

1.5.2.5 นำระยะทางของแต่ละจุดไปยังทุกจุดมาคำนวณเป็นค่าเฉลี่ย ค่ามัธยฐาน ค่า quantile ที่ 75 หรือ 95 เพื่อนำค่าดังกล่าวเป็นค่า Eps ในที่นี้เรียกว่าวิธี mean, median, P75 และ P95 ตามลำดับ

1.5.3 นำค่า MinPts และ Eps ที่ได้จากข้างต้น มาเข้าสู่วิธีการ DBSCAN

1.5.4 เปรียบเทียบประสิทธิภาพของวิธีในการระบุค่าพารามิเตอร์ด้วยค่า modified purity โดยนำเสนอในรูปแบบของกราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จากวิธีต่าง ๆ ด้วยค่า MinPts ที่ต่างกัน และกราฟแสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ด้วยค่า MinPts ที่ต่างกัน รวมถึงเปรียบเทียบกับ K-means ด้วย

1.6 ประโยชน์ที่คาดว่าจะได้รับ

เพื่อเป็นแนวทางในการเลือกใช้วิธีในการระบุค่าพารามิเตอร์ในขั้นตอนวิธี DBSCAN ได้อย่างเหมาะสม

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

การระบุค่าพารามิเตอร์ Eps ในขั้นตอนวิธี DBSCAN ได้แก่ วิธีของ Daszykowski วิธี mean วิธี median วิธี P75 วิธี P95 วิธีของ Xia วิธีที่ประยุกต์จากวิธีของ Xia (Xia mean Xia median Xia P75 และ Xia P95) และวิธีของ Karami มีทฤษฎีที่เกี่ยวข้องและเกณฑ์ในการวัดประสิทธิภาพ ดังต่อไปนี้

2.1 วิธีการจัดกลุ่ม DBSCAN

Ester et al. (1996) ได้ทำการศึกษาขั้นตอนวิธีการจัดกลุ่มแบบใหม่ซึ่งอาศัยความหนาแน่นที่มีรากฐานจากแนวคิดของกลุ่ม วิธีดังกล่าวนี้คือ DBSCAN เป็นวิธีที่ถูกรออกแบบเพื่อค้นพบกลุ่มที่มีรูปร่างไม่ปกติ ในงานวิจัยนี้ดำเนินการประเมินผลโดยใช้ข้อมูลสังเคราะห์ขึ้นมาและข้อมูลจริง ผลจากการทดลองพบว่า DBSCAN มีประสิทธิภาพอย่างมีนัยสำคัญกว่าขั้นตอนวิธีการจัดกลุ่ม CLARANS และ DBSCAN ดำเนินการได้ดีกว่า CLARANS ด้วยปัจจัยอย่างน้อย 100 ตัวในเทอมของประสิทธิภาพ

DBSCAN (Density Based Spatial Clustering of Applications with Noise) เป็นการจัดกลุ่มโดยใช้ความหนาแน่น ด้วยเหตุนี้ทำให้สามารถใช้ได้กับรูปร่างที่ไม่แน่นอน ขนาดต่างๆ ได้ และสามารถแยกจุดที่ข้อมูลมีความผิดปกติ ไม่สามารถรวมกับกลุ่มใดๆ ได้อีกด้วย อีกทั้งไม่ต้องกำหนดจำนวนกลุ่มที่ต้องการ วิธีการนี้เหมาะสำหรับข้อมูลเชิงตัวเลข

DBSCAN ประกอบด้วย 2 พารามิเตอร์ คือ Eps แสดงรัศมีที่ใช้กำหนด neighborhood ของจุด ๆ หนึ่ง และ MinPts คือจำนวนจุดที่น้อยที่สุดที่อยู่ภายใน neighborhood ของจุด ๆ หนึ่ง ในรัศมี Eps ซึ่งใช้บ่งชี้ว่าจุด ๆ นั้นเป็นจุดแกนกลางหรือไม่

จากงานวิจัยของพงค์เลิศ สังกะเพศ (2554) ได้อธิบายเนื้อหาในส่วนของขั้นตอน DBSCAN ดังนี้

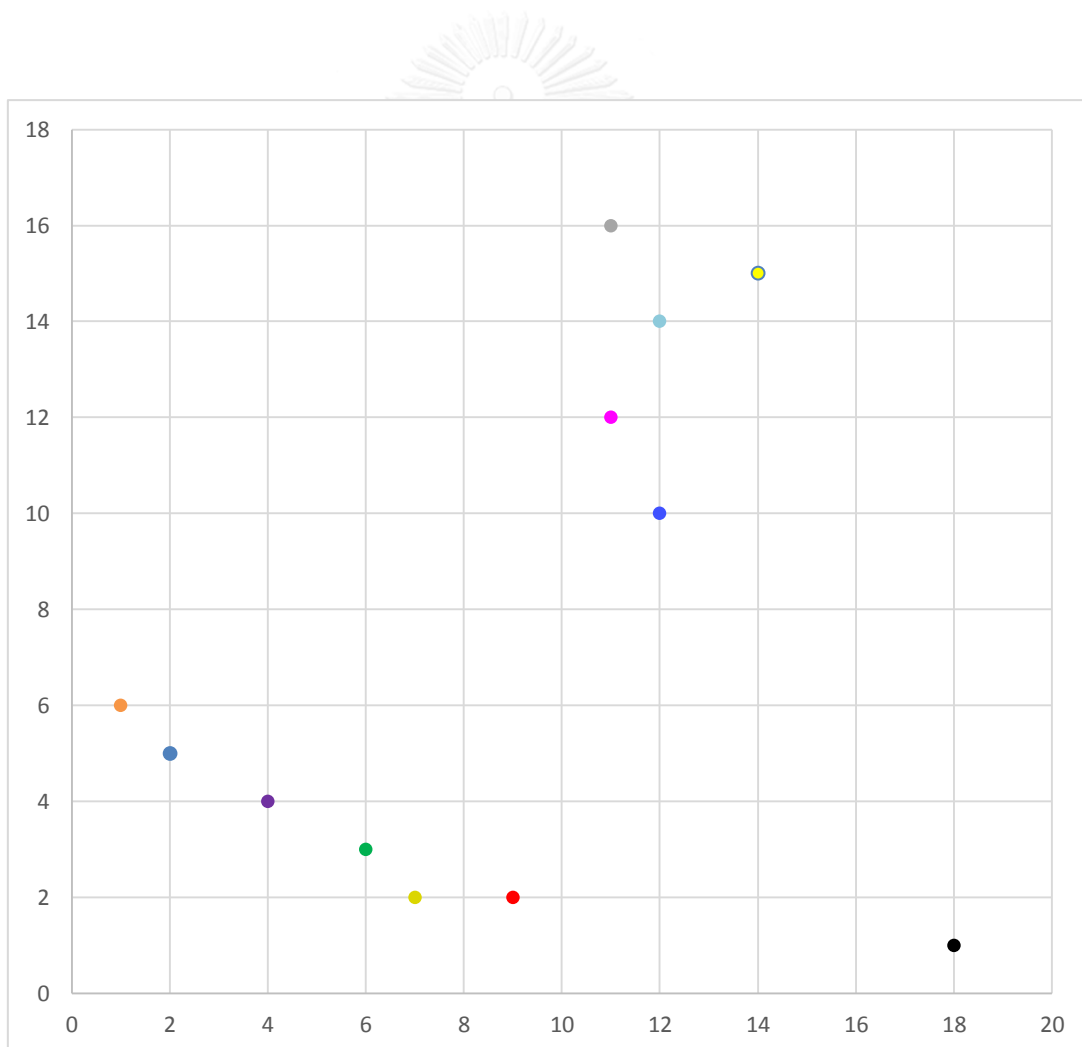
1. ระบุค่าพารามิเตอร์ ค่า Eps และค่า MinPts
2. สุ่มเลือกจุดเริ่มต้น โดยกำหนดให้เป็นจุด P
3. หาจุดที่อยู่ใกล้จุด P เมื่อได้จุดที่อยู่ใกล้จุด P (เรียกว่าจุด q) แล้ว ให้ทำการพิจารณาว่าจุด q นั้นอยู่ภายใต้เงื่อนไขของ Eps หรือไม่ นั่นคือ $\text{dist}(p,q) \leq \text{Eps}$ จากนั้นทำการรวมเป็นกลุ่มใหม่

4. ค้นหาทุกจุดที่ใกล้จุดอื่นที่อยู่ภายในกลุ่ม ทำการเลือกจุดอื่นที่อยู่ภายในกลุ่มเป็นจุด P จากนั้นทำตามขั้นตอน 3 และ 4 วนจนครบ

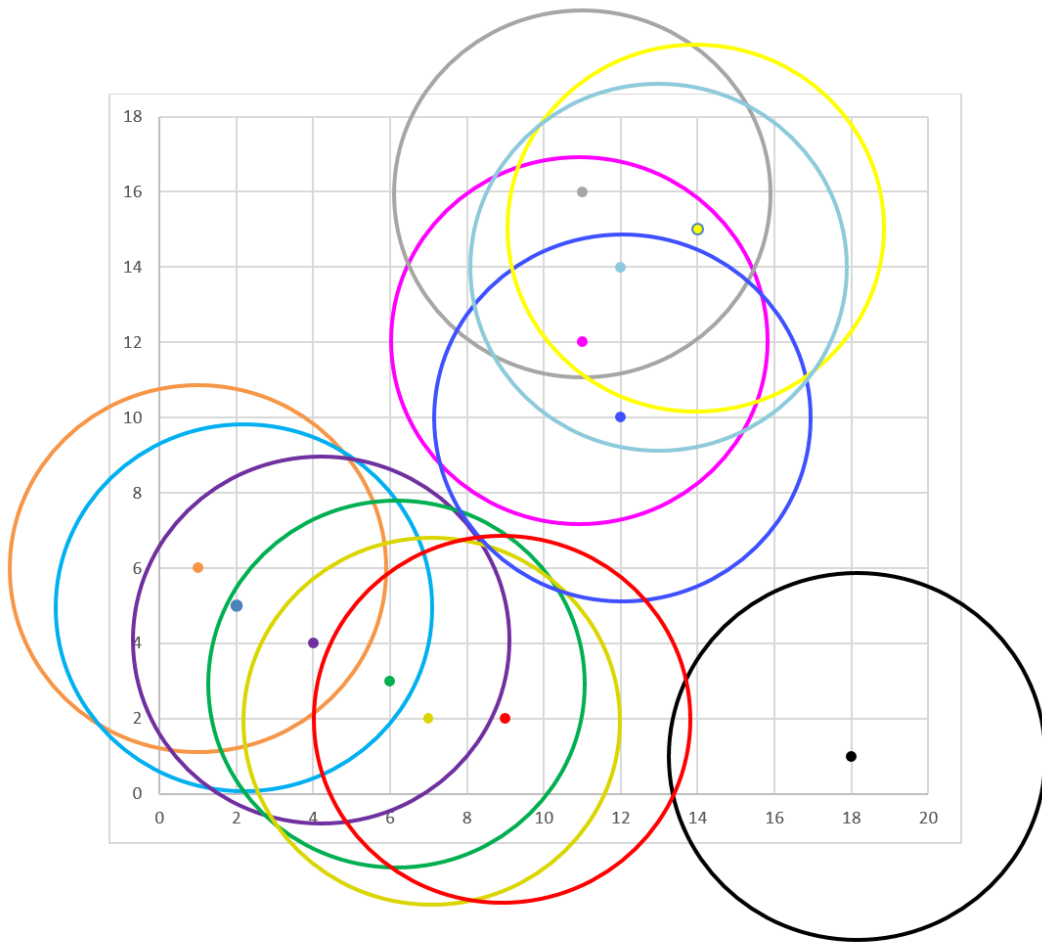
5. ตรวจสอบจุดครบถ้วน โดยพิจารณาจากจำนวนจุดภายในกลุ่มนั้นๆ น้อยกว่า MinPts หรือไม่ ถ้าน้อยกว่าให้กำหนดจุดในกลุ่มนั้นๆ เป็น Noise

6. วนซ้ำ วนซ้ำจากข้อ 2. ไปยังข้อ 5. ตรวจสอบว่าทุกจุดในเซตของข้อมูลถูกระบุเป็นจุดในการจัดกลุ่มครบหรือไม่ ถ้ายังไม่ครบให้กลับไปทำการสุ่มเลือกจุดเริ่มต้น โดยจะสุ่มเลือกตัวที่ไม่เคยถูกสุ่มและไม่ได้เป็นจุดที่อยู่ในกลุ่มใดๆ แต่ถ้าครบแล้วให้หยุดการทำงาน

ต่อไปเป็นตัวอย่างแสดงการดำเนินการของ DBSCAN



รูปที่ 2.1.1 ตัวอย่างการกระจายของข้อมูล



รูปที่ 2.1.2 แต่ละจุดของข้อมูลมีการกวาดรัศมีไปเป็นระยะ Eps

CHULALONGKORN UNIVERSITY

กำหนดค่าพารามิเตอร์ MinPts = 4 และ Eps = 5

จากรูปที่ 2.1.2 พบว่า

(2,5), (4,4), (6,3), (7,2), (11,12), (11,16), (12,14) และ (14,15) เป็นจุดที่มีจำนวนจุดภายในรัศมี Eps มีค่ามากกว่าหรือเท่ากับ MinPts ดังนั้น จุดข้างต้นเป็นจุดแกนกลาง

(1,6), (9,2) และ (12,10) เป็นจุดที่อยู่ภายในรัศมี Eps ของจุดแกนกลาง ดังนั้น จุดข้างต้นเป็นจุดขอบ

(18,1) ไม่ใช่ทั้งจุดแกนกลางและจุดขอบ ดังนั้น จุดนี้จึงเป็นจุดรบกวน กล่าวคือ ไม่มีการระบุว่ายู่กลุ่มใด

ดำเนินการต่อไปกับจุดแกนกลางที่อยู่ภายในรัศมี Eps ของกันและกัน ทำให้เกิดกลุ่มใหม่ขึ้นมา ส่วนจุดขอบที่อยู่ภายในรัศมี Eps ของจุดแกนกลางใด ๆ ก็จะอยู่ในกลุ่มเดียวกับกลุ่มของจุดแกนกลางนั้น ๆ

หรือดำเนินการ DBSCAN อีกแบบหนึ่ง

เลือกจุดเริ่มต้น คือ จุด (1,6) และเมื่อกวาดรัศมีจากจุดเริ่มต้นไปเป็นระยะ Eps ดังตารางที่ 2.1.1

ตารางที่ 2.1.1 ค่าระยะห่างจากจุดเริ่มต้น (1,6) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(1,6)	0.000
(2,5)	1.414
(4,4)	3.606
(6,3)	5.831
(7,2)	7.211
(9,2)	8.944
(11,12)	11.662
(11,16)	14.142
(12,10)	11.705
(12,14)	13.601
(14,15)	15.811
(18,1)	17.720

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 3 จุด นั่นคือ (1,6), (2,5) และ (4,4) ที่อยู่ในระยะ Eps

เลือกจุดเริ่มต้น คือ จุด (2,5) ดังตารางที่ 2.1.2

ตารางที่ 2.1.2 ค่าระยะห่างจากจุดเริ่มต้น (2,5) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(1,6)	1.414
(2,5)	0.000
(4,4)	2.236
(6,3)	4.472
(7,2)	5.831
(9,2)	7.616
(11,12)	11.402
(11,16)	14.213
(12,10)	11.180
(12,14)	13.454
(14,15)	15.621
(18,1)	16.492

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 4 จุด นั่นคือ (1,6), (2,5), (4,4) และ (6,3) ที่อยู่ในระยะ Eps (จุด (6,3) เพิ่มเข้ามาอยู่ในกลุ่มนี้)

เลือกจุดเริ่มต้น คือ จุด (4,4) ดังตารางที่ 2.1.3

ตารางที่ 2.1.3 ค่าระยะห่างจากจุดเริ่มต้น (4,4) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(1,6)	3.606
(2,5)	2.236
(4,4)	0.000
(6,3)	2.236
(7,2)	3.606
(9,2)	5.385
(11,12)	10.630
(11,16)	13.892
(12,10)	10.000
(12,14)	12.806
(14,15)	14.866
(18,1)	14.318

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 5 จุด นั่นคือ (1,6), (2,5), (4,4), (6,3) และ (7,2) ที่อยู่ในระยะ Eps (จุด (7,2) เพิ่มเข้ามาอยู่ในกลุ่มนี้)

เลือกจุดเริ่มต้น คือ จุด (6,3) ดังตารางที่ 2.1.4

ตารางที่ 2.1.4 ค่าระยะห่างจากจุดเริ่มต้น (6,3) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(1,6)	5.831
(2,5)	4.472
(4,4)	2.236
(6,3)	0.000
(7,2)	1.414
(9,2)	3.162
(11,12)	10.296
(11,16)	13.928
(12,10)	9.220
(12,14)	12.530
(14,15)	14.422
(18,1)	12.166

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 5 จุด นั่นคือ (2,5), (4,4), (6,3), (7,2) และ (9,2) ที่อยู่ในระยะ Eps (จุด (9,2) เพิ่มเข้ามาอยู่ในกลุ่มนี้)

กำหนดจุดเริ่มต้น คือ จุด (7,2) ดังตารางที่ 2.1.5

ตารางที่ 2.1.5 ค่าระยะห่างจากจุดเริ่มต้น (7,2) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(1,6)	7.211
(2,5)	5.831
(4,4)	3.606
(6,3)	1.414
(7,2)	0.000
(9,2)	2.000
(11,12)	10.770
(11,16)	14.560
(12,10)	9.434
(12,14)	13.000
(14,15)	14.765
(18,1)	11.045

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 4 จุด นั่นคือ (4,4), (6,3), (7,2) และ (9,2) ที่อยู่ในระยะ Eps

กำหนดจุดเริ่มต้น คือ จุด (9,2) ดังตารางที่ 2.1.6

ตารางที่ 2.1.6 ค่าระยะห่างจากจุดเริ่มต้น (9,2) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(1,6)	8.944
(2,5)	7.616
(4,4)	5.385
(6,3)	3.162
(7,2)	2.000
(9,2)	0.000
(11,12)	10.198
(11,16)	14.142
(12,10)	8.544
(12,14)	12.369
(14,15)	13.928
(18,1)	9.055

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 3 จุด นั่นคือ (6,3), (7,2) และ (9,2) ที่อยู่ในระยะ Eps

กำหนดจุดเริ่มต้น คือ จุด (11,12) ดังตารางที่ 2.1.7

ตารางที่ 2.1.7 ค่าระยะห่างจากจุดเริ่มต้น (11,12) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(1,6)	11.662
(2,5)	11.402
(4,4)	10.630
(6,3)	10.296
(7,2)	10.770
(9,2)	10.198
(11,12)	0.000
(11,16)	4.000
(12,10)	2.236
(12,14)	2.236
(14,15)	4.243
(18,1)	13.038

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 5 จุด นั่นคือ (11,12), (11,16), (12,10), (12,14) และ (14,15) ที่อยู่ในระยะ Eps

กำหนดจุดเริ่มต้น คือ จุด (11,16) ดังตารางที่ 2.1.8

ตารางที่ 2.1.8 ค่าระยะห่างจากจุดเริ่มต้น (11,16) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(11,12)	4.000
(11,16)	0.000
(12,10)	6.083
(12,14)	2.236
(14,15)	3.162
(18,1)	16.553

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 4 จุด นั่นคือ (11,12), (11,16), (12,14) และ (14,15) ที่อยู่ในระยะ Eps (จุด (12,14) เพิ่มเข้ามาอยู่ในกลุ่มนี้)

กำหนดจุดเริ่มต้น คือ จุด (12,10) ดังตารางที่ 2.1.9

ตารางที่ 2.1.9 ค่าระยะห่างจากจุดเริ่มต้น (12,10) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(11,12)	2.236
(11,16)	6.083
(12,10)	0.000
(12,14)	4.000
(14,15)	5.385
(18,1)	10.817

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 3 จุด นั่นคือ (11,12), (12,10) และ (12,14) ที่อยู่ในระยะ Eps

กำหนดจุดเริ่มต้น คือ จุด (12,14) ดังตารางที่ 2.1.10

ตารางที่ 2.1.10 ค่าระยะห่างจากจุดเริ่มต้น (12,14) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(11,12)	2.236
(11,16)	2.236
(12,10)	4.000
(12,14)	0.000
(14,15)	2.236
(18,1)	14.318

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 5 จุด นั่นคือ (11,12), (11,16), (12,10), (12,14) และ (14,15) ที่อยู่ในระยะ Eps

กำหนดจุดเริ่มต้น คือ จุด (14,15) ดังตารางที่ 2.1.11

ตารางที่ 2.1.11 ค่าระยะห่างจากจุดเริ่มต้น (14,15) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(11,12)	4.243
(11,16)	3.162
(12,10)	5.385
(12,14)	2.236
(14,15)	0.000
(18,1)	14.560

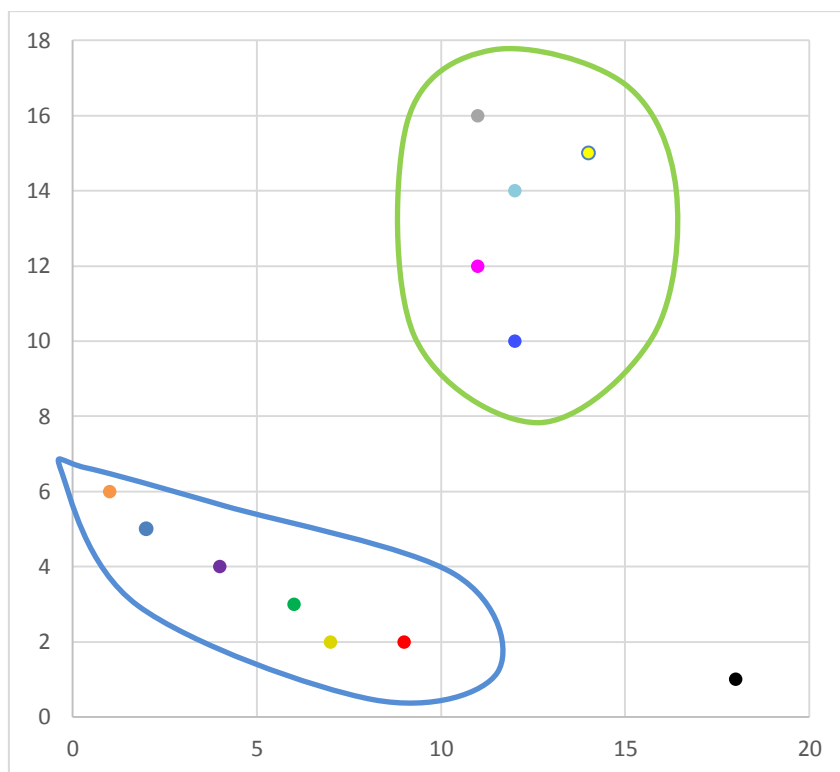
เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่ามี 4 จุด นั่นคือ (11,12), (11,16), (12,14) และ (14,15) ที่อยู่ในระยะ Eps

กำหนดจุดเริ่มต้น คือ จุด (18,1) ดังตารางที่ 2.1.12

ตารางที่ 2.1.12 ค่าระยะห่างจากจุดเริ่มต้น (18,1) ไปยังจุดปลาย

จุด	ระยะห่างจากจุดเริ่มต้น
(11,12)	13.038
(11,16)	16.553
(12,10)	10.817
(12,14)	14.318
(14,15)	14.560
(18,1)	0.000

เมื่อพิจารณาระยะห่างจากจุดเริ่มต้น พบว่าไม่มีจุดใดเลยที่อยู่ใกล้กับจุด (18,1) ภายในระยะ Eps ดังนั้น จุด (18,1) เป็นจุดรบกวน



รูปที่ 2.1.3 ผลจากการแบ่งกลุ่มข้อมูล

สรุปได้ว่า ข้อมูลนี้แบ่งออกเป็น 2 กลุ่ม

(1,6), (2,5), (4,4), (6,3), (7,2) และ (9,2) อยู่กลุ่มที่ 1

(11,12), (11,16), (12,10), (12,14) และ (14,15) อยู่กลุ่มที่ 2

จุด (18,1) เป็นจุดรบกวน

2.2 วิธีของ Daszykowski

Daszykowski et al. (2001) ได้ทำการศึกษาเกี่ยวกับการมองหารูปแบบโดยธรรมชาติในชุดข้อมูล ซึ่งมีการกล่าวถึงพารามิเตอร์ทั้งสองดังนี้

จากงานวิจัยนี้มีการกล่าวถึงพารามิเตอร์ทั้งสองดังนี้

วิธีที่ใช้หา MinPts

คำนวณจาก rule of thumb

$$MinPts = integer\left(\frac{m}{25}\right)$$

เมื่อ m คือจำนวนจุดในชุดข้อมูล แต่ถ้าข้อมูลมีจำนวนข้อมูลเกิน 500 ควรใช้ MinPts เท่ากับ 20

วิธีที่ใช้หา Eps

กรณีที่ข้อมูลนั้น ๆ มี $m \geq 50$

1. หาขอบเขตค่าสูงสุดและค่าต่ำสุดในแต่ละมิติ
2. สร้างรูปเหลี่ยมมุมฉาก n มิติ ที่ครอบคลุมค่าสูงสุดและค่าต่ำสุดของข้อมูลเดิมในทุกมิติ
3. จำลองข้อมูลโดยสุ่มจุดจำนวน m ตัวให้มีการกระจายแบบสม่ำเสมอทั่วบริเวณ
4. คำนวณค่า MinPts จากสูตร $MinPts = integer\left(\frac{m}{25}\right)$ หากข้อมูลมีจำนวนข้อมูลเกิน 500 ควรใช้ MinPts เท่ากับ 20 และกำหนดค่าที่คำนวณได้เป็น k
5. ในแต่ละจุดให้หาระยะทางจากจุด ๆ นั้น ไปยังจุดที่ห่างออกไปเป็นลำดับที่ k
6. เรียงลำดับระยะทาง จะได้ว่า Eps คือ 95% quantile ของระยะทางจากแต่ละจุดของข้อมูลไปยังจุดที่ห่างออกไปเป็นลำดับที่ k

กรณีที่ข้อมูลนั้น ๆ มี $m < 50$

ทำเหมือนขั้นตอนในกรณีที่ข้อมูลนั้น ๆ มี $m \geq 50$ แต่ทำต่อด้วยการจำลองข้อมูล t ครั้ง จะได้ว่า Eps คือค่าเฉลี่ยของ 95% quantile ของทั้ง t ครั้ง

2.3 วิธีของ Xia และ Jing

XIA and JING (2009) ได้ทำการศึกษา SA-DBSCAN (A self-adaptive density-based clustering algorithm) โดยมีวิธีการหา Eps ผ่านกระบวนการ fit กราฟฮิสโทแกรมระยะห่างด้วยการแจกแจงเกาส์เซียนแบบผกผัน เพื่อหาจุดสูงสุดภายใต้การแจกแจง

จากงานวิจัยนี้มีการกล่าวถึงพารามิเตอร์ดังนี้

วิธีที่ใช้หา Eps

1. เลือกค่า k (k ดังกล่าวคือ MinPts)
2. แต่ละจุดให้หาระยะทางจากจุด ๆ นั้น ไปยังจุดที่ห่างออกไปเป็นลำดับที่ k
3. เรียงลำดับระยะทางจากน้อยไปมาก

4. fit ข้อมูลระยะห่างจากแต่ละจุดของข้อมูลไปยังจุดที่ใกล้ที่สุดเป็นลำดับที่ k ด้วยการแจกแจงเกาส์เซียนแบบผกผัน

5. นำค่าเฉลี่ย μ และพารามิเตอร์รูปร่าง (shape parameter) λ ที่ได้จากการ fit กราฟฮิสโทแกรมมาทำการคำนวณ ฐานนิยมที่ได้ นั่นคือ Eps

$$Eps_k = \frac{\mu_k \sqrt{9\mu_k^2 + 4\lambda_k^2 - 3\mu_k^2}}{2\lambda_k}$$

จากข้อ 5 ผู้ศึกษาเล็งเห็นว่าสามารถปรับการหาค่า Eps นอกเหนือจากสูตรข้างต้นได้ โดยใช้เป็นค่าเฉลี่ย ค่ามัธยฐาน ค่า quantile ที่ 75 หรือ 95 (ค่ากลางค่าอื่นที่ไม่ใช่ฐานนิยม)

2.4 วิธีของ Karami และ Johansson

Karami and Johansson (2014) ได้ทำการศึกษาความมีประสิทธิภาพของวิธีจัดกลุ่มลูกผสม นั่นคือ BDE-DBSCAN ซึ่งเป็นการรวมกันระหว่าง Binary Differential Evolution (BDE) และขั้นตอนวิธี DBSCAN การดำเนินการของ BDE-DBSCAN เป็นการหาค่าโดยใช้ชุดข้อมูลที่หลากหลายด้วยความหนาแน่นและรูปร่างที่แตกต่างกัน

จากงานวิจัยนี้มีการกล่าวถึงพารามิเตอร์ดังนี้

วิธีที่ใช้หา Eps

คำนวณโดยใช้วิธีการวิเคราะห์ (analytical way) ได้ดังนี้

$$Eps = \left(\frac{(\prod_{d=1}^D (\max_{j \in \{1,2,\dots,n\}} x_{jd} - \min_{j \in \{1,2,\dots,n\}} x_{jd})) \times k \times \gamma(0.5 \times D + 1)}{n \times \sqrt{\pi^D}} \right)^{1/D}$$

$$= \left(\frac{(\prod_{d=1}^D (\max_{j \in \{1,2,\dots,n\}} x_{jd} - \min_{j \in \{1,2,\dots,n\}} x_{jd})) \times k \times (0.5 \times D)!}{n \times \sqrt{\pi^D}} \right)^{1/D}$$

เมื่อ X คือ เมตริกซ์ข้อมูล $\begin{bmatrix} x_{11} & \cdots & x_{1D} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nD} \end{bmatrix}$

D คือ จำนวนมิติของเซตข้อมูลหรือจำนวนตัวแปร

n คือ จำนวนข้อมูล

k คือ MinPts

γ คือ ฟังก์ชันแกมมา

ซึ่งมีที่มาของสูตรมาจาก (Ankerst, Breunig, Kriegel, & Sander, 1999) ดังนี้

เราใช้ระยะห่าง k-nearest-neighbor ที่คาดหวัง สำหรับ $k = \text{MinPts}$ ภายใต้ข้อสมมติฐาน ข้อมูลมีการแจกแจงอย่างสม่ำเสมอ กล่าวคือ ข้อมูลไม่มีกลุ่ม ค่านี้สามารถกำหนดสำหรับ data space DS ที่ประกอบด้วย n จุด ระยะห่างจะมีค่าเท่ากับรัศมี r ของ hypersphere S ที่มี d มิติ ใน DS โดยที่ S ประกอบด้วย k จุดอย่างแน่ชัด

$$\text{Volume}_S = \frac{\text{Volume}_{DS}}{n} \times k$$

ปริมาตรของ hypersphere S ที่มี d มิติ และรัศมี r

$$\text{Volume}_{S(r)} = \frac{\sqrt{\pi^d}}{\gamma(\frac{d}{2} + 1)} \times r^d$$

โดยที่ γ คือฟังก์ชันแกมมา

ดังนั้น รัศมี r สามารถคำนวณได้ดังนี้

$$r = \sqrt{\frac{\text{Volume}_{DS} \times k \times \gamma(\frac{d}{2} + 1)}{n \times \sqrt{\pi^d}}}$$

เมื่อเทียบกับสูตร Eps ข้างต้น จะเห็นว่า ในส่วนของ $\prod_{d=1}^D (\max_{j \in \{1,2,\dots,n\}} x_{jd} - \min_{j \in \{1,2,\dots,n\}} x_{jd})$ เป็นการหาปริมาตรของข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.5 วิธี K-means

ขั้นตอนการวิเคราะห์ของวิธี K-means ผู้ศึกษาต้องกำหนดจำนวนกลุ่มเอง และตัวแปรที่ใช้เป็นตัวแปรเชิงปริมาณ

จาก กัลยา วานิชย์บัญชา (2552) ได้อธิบายเนื้อหาในส่วนของขั้นตอน ดังนี้

1. แบ่งข้อมูลเป็น K กลุ่ม
2. คำนวณหาจุดกลางกลุ่มของแต่ละกลุ่ม
3. มีวิธีการพิจารณา 2 แบบ

3.1 คำนวณหาระยะห่างจากแต่ละหน่วยไปยังจุดกลางกลุ่มของทุกกลุ่ม ถ้าระยะห่างจากหน่วยที่ i ใกล้กับจุดกลางกลุ่มของกลุ่มใดน้อยที่สุด ก็ย้ายให้หน่วยที่ i ไปอยู่กลุ่มนั้น

3.2 คำนวณระยะห่างกำลังสองของแต่ละหน่วยไปยังจุดกลางกลุ่มที่หน่วยนั้นอยู่ โดยให้ ESS (Error Sum Square) เท่ากับระยะห่างกำลังสองของแต่ละหน่วยไปยังจุดกลางกลุ่ม

$$ESS = \sum_{i=1}^n (x_i - \bar{x}_{C(i)})^2$$

โดยที่ $C(i)$ หมายถึงกลุ่มของหน่วยที่ i

ESS หมายถึง ผลบวกของระยะห่างจากแต่ละหน่วยในกลุ่มไปยังจุดกลางกลุ่มรวมทุกกลุ่ม กลุ่มใดที่มีค่า ESS ต่ำ แสดงว่าหน่วยที่อยู่ในกลุ่มนั้นมีความคล้ายคลึงกัน

4. การพิจารณาการย้ายกลุ่ม จะใช้เกณฑ์การย้ายตามค่าที่คำนวณได้ในขั้นที่ 3

4.1 จะทำการย้ายหน่วยที่ i ไปยังกลุ่มที่ทำให้ระยะห่างจากหน่วยที่ i ไปยังจุดกลางกลุ่มมีค่าต่ำสุด

4.2 จะทำการย้ายหน่วยที่ i ไปยังกลุ่มที่ทำให้ค่า ESS มีค่าต่ำสุด

ถ้าในขั้น 4 ไม่มีการย้ายกลุ่มอีกแล้ว แสดงว่ากลุ่มที่แบ่งได้นั้นเหมาะสมแล้ว แต่ถ้าในขั้นที่ 4 มีการย้ายกลุ่ม กลุ่มที่มีหน่วยย้ายเข้าหรือย้ายออกจะต้องทำการคำนวณหาจุดกลางกลุ่มใหม่ นั่นคือ ต้องกลับไปทำขั้นที่ 2

2.6 การวัดประสิทธิภาพการจัดกลุ่ม

การวัดประสิทธิภาพการจัดกลุ่มจะทำให้ทราบว่าการระบุพารามิเตอร์วิธีใดมีความเหมาะสมมากกว่ากัน ในที่นี้จะใช้เกณฑ์ในการตัดสินใจ ดังนี้

2.6.1 Purity เป็นค่าที่ชี้วัดว่าแต่ละกลุ่มที่แบ่งได้ประกอบด้วยปัจจัยที่ใช้แบ่งแยกข้อมูลที่มีรูปแบบเดียวกันหรือไม่ ซึ่งคำนวณได้จากสูตร (Manning, Raghavan, & Schütze, 2008)

$$purity(\Omega, C) = \frac{1}{N} \sum_k \max_j |\omega_k \cap c_j|$$

โดยที่ N คือ จำนวนข้อมูลทั้งหมด

Ω คือ $\{\omega_1, \omega_2, \dots, \omega_k\}$ เป็นเซตของกลุ่มที่ได้จากการแบ่งข้อมูล

C คือ $\{c_1, c_2, \dots, c_j\}$ เป็นเซตของกลุ่มข้อมูลที่แท้จริง

วิธีการจัดกลุ่มแบบใดมีค่า purity เป็น 1 แสดงว่ามีการจัดกลุ่มที่สมบูรณ์ แต่ถ้าวิธีการจัดกลุ่มแบบใดมีค่า purity ใกล้ 0 แสดงว่ามีการจัดกลุ่มที่ไม่เหมาะสม

2.6.2 Modified Purity เป็นค่าวัดประสิทธิภาพการจำแนกกลุ่มข้อมูล que ผู้ศึกษาดัดแปลงมาจากค่า Purity เพื่อให้สามารถชี้วัดความสามารถในการจัดจำแนกกลุ่มได้ถูกต้องตรงกับกลุ่มที่แท้จริงของข้อมูล เนื่องจากค่า Purity เป็นค่าวัดที่ไม่ได้พิจารณาถึงจำนวนกลุ่มที่ถูกจำแนกออกมาได้ ในบางกรณีอาจจะได้ค่า Purity สูง แต่ก็ไม่ใช่การจำแนกกลุ่มที่ถูกต้อง เช่น การจำแนกข้อมูลออกเป็นกลุ่มย่อยจำนวนมากเกินกว่าจำนวนกลุ่มที่แท้จริงที่กำหนดไว้ (คลาส) ผู้ศึกษาจึงได้กำหนดค่าวัด Modified Purity ขึ้น โดยมีการนำจำนวนกลุ่มที่แท้จริงของข้อมูล (คลาส) มาพิจารณาประกอบด้วย ซึ่งจะวัดความถูกต้องในการจำแนกกลุ่มโดยเลือกผลลัพธ์การจับคู่กลุ่มกับคลาสแบบ 1 ต่อ 1 (โดยการทำให้ combination) ที่มีค่า purity สูงที่สุดโดยใช้สูตรดังนี้

$$\text{modified purity}(\Omega, C) = \frac{1}{N} \max_{\substack{k_1, k_2, \dots, k_{\min(K,J)} \in \{1, 2, \dots, K\} \\ k_1 \neq k_2 \neq \dots \neq k_{\min(K,J)} \\ j_1, j_2, \dots, j_{\min(K,J)} \in \{1, 2, \dots, J\} \\ j_1 \neq j_2 \neq \dots \neq j_{\min(K,J)}}} \sum_{i=1}^{\min(K,J)} |\omega_{k_i} \cap c_{j_i}|$$

โดยที่ N คือ จำนวนข้อมูลทั้งหมด

K คือ จำนวนกลุ่มที่ได้จากการแบ่งข้อมูล (cluster)

J คือ จำนวนกลุ่มข้อมูลที่แท้จริง (class)

Ω คือ $\{\omega_1, \omega_2, \dots, \omega_K\}$ เป็นเซตของกลุ่มที่ได้จากการแบ่งข้อมูล K กลุ่ม

C คือ $\{c_1, c_2, \dots, c_J\}$ เป็นเซตของกลุ่มข้อมูลที่แท้จริง J กลุ่ม

บทที่ 3

วิธีดำเนินการศึกษา

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของวิธีที่ใช้ในการระบุค่าพารามิเตอร์ Eps นั่นคือวิธีของ Daszykowski วิธี mean วิธี median วิธี P75 วิธี P95 วิธีของ Xia วิธีที่ประยุกต์จากวิธีของ Xia (Xia mean Xia median Xia P75 และ Xia P95) และวิธีของ Karami ในขั้นตอนวิธี DBSCAN โดยใช้เกณฑ์วัดประสิทธิภาพการวิเคราะห์กลุ่มที่ประยุกต์มาจากค่า purity เรียกว่า Modified purity ในแต่ละชุดข้อมูล การวิเคราะห์ข้อมูลทั้งหมดใช้โปรแกรม R เวอร์ชัน 3.2.1 ภายใต้ขอบเขตและวิธีการดำเนินงานดังนี้

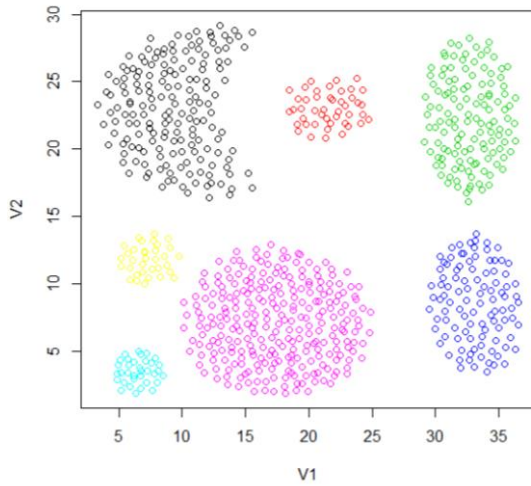
3.1 ขอบเขตของการศึกษา

ในการศึกษาครั้งนี้จะทำการศึกษาภายใต้ขอบเขตดังนี้

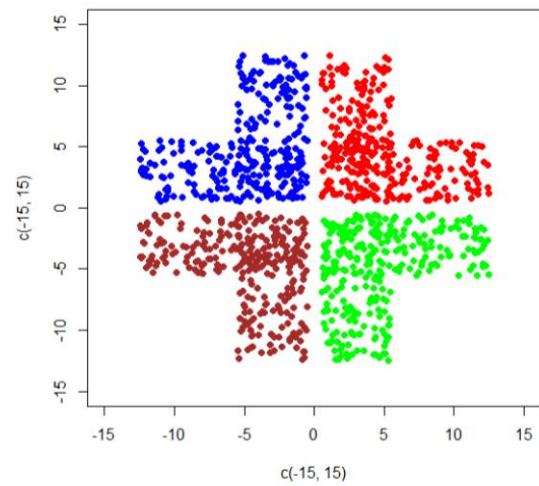
3.1.1 ลักษณะข้อมูลที่ศึกษา

3.1.1.1 ข้อมูลไม่ปกติ 2 มิติ 4 ลักษณะ ได้แก่

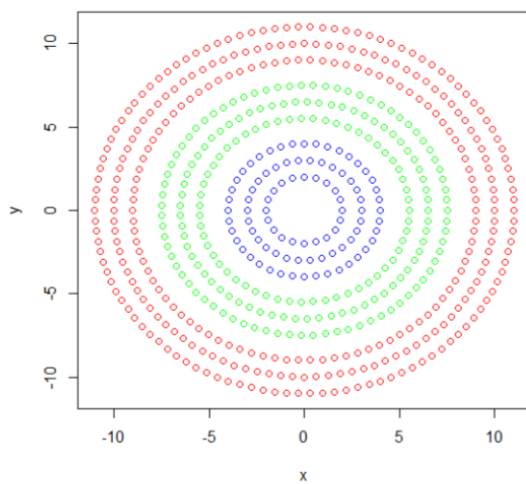
1. ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน
2. ข้อมูลวงแหวนรูปแบบที่ 1
3. ข้อมูลวงแหวนรูปแบบที่ 2
4. ข้อมูลมีลักษณะเป็นมุม



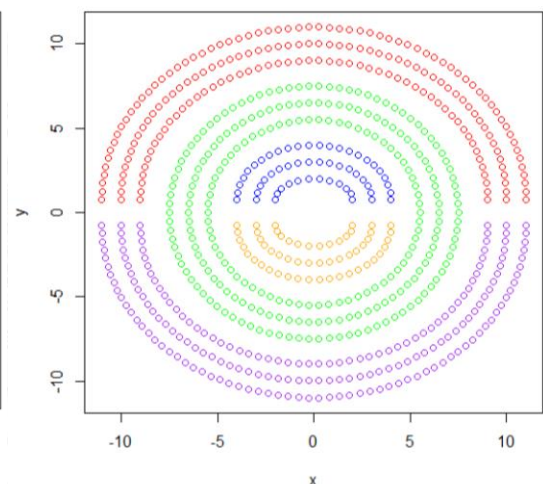
ข้อมูลที่มีรูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน



ข้อมูลมีลักษณะเป็นมุม



ข้อมูลวงแหวนรูปแบบที่ 1



ข้อมูลวงแหวนรูปแบบที่ 2

รูปที่ 3.1.1.1 แสดงข้อมูลไม่ปกติ 2 มิติ 4 ลักษณะ

(Karami & Johansson, 2014)

3.1.1.2 ข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร แบ่งออกเป็นกรณีจำนวนกลุ่มเท่ากับ 3 และจำนวนกลุ่มเท่ากับ 5 ในการศึกษานี้จะทำการศึกษาภายใต้ขอบเขต ดังต่อไปนี้

1. จำนวนกลุ่มเท่ากับ 3 (K = 3)

กำหนดให้ $X \sim N_2(\mu_i, \Sigma_i)$, $i = 1, 2, 3$ โดยที่

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \text{ และ } \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

และพิจารณาเมทริกซ์ความแปรปรวนในสองกรณีย่อย ได้แก่ กรณีความเมทริกซ์แปรปรวนเท่ากัน และเมทริกซ์แปรปรวนไม่เท่ากัน

กรณี 1 เมทริกซ์ความแปรปรวนเท่ากัน

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

กรณี 2 เมทริกซ์ความแปรปรวนไม่เท่ากัน

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \text{ และ } \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

รวมถึงพิจารณาขนาดตัวอย่างในสองกรณีย่อย ได้แก่ $n=300$ และ $n=1,200$

กำหนดขนาดตัวอย่าง $n = 300$ แบ่งออกเป็น 3 กรณี ดังตารางที่ 3.1.1.1

ตารางที่ 3.1.1.1 จำนวนข้อมูลแต่ละกลุ่มสำหรับขนาดตัวอย่างเท่ากับ 300 เมื่อจำนวนกลุ่มเท่ากับ 3

กรณี	จำนวนข้อมูลแต่ละกลุ่ม		
	n_1	n_2	n_3
1	50	100	150
2	100	100	100
3	150	100	50

และสำหรับ $n = 1,200$ แบ่งออกเป็น 3 กรณี ดังตารางที่ 3.1.1.2

ตารางที่ 3.1.1.2 จำนวนข้อมูลแต่ละกลุ่มสำหรับขนาดตัวอย่างเท่ากับ 1,200 เมื่อจำนวนกลุ่มเท่ากับ 3

กรณี	จำนวนข้อมูลแต่ละกลุ่ม		
	n_1	n_2	n_3
1	200	400	600
2	400	400	400
3	600	400	200

2. จำนวนกลุ่มเท่ากับ 5 ($K = 5$)

กำหนดให้ $X \sim N_2(\mu_i, \Sigma_i)$, $i = 1, 2, 3, 4, 5$ โดยที่

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix} \text{ และ } \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$$

และพิจารณาเมทริกซ์ความแปรปรวนในสองกรณีย่อย ได้แก่ กรณีความเมทริกซ์แปรปรวนเท่ากัน และเมทริกซ์แปรปรวนไม่เท่ากัน

กรณี 1 เมทริกซ์ความแปรปรวนเท่ากัน

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

กรณี 2 เมทริกซ์ความแปรปรวนไม่เท่ากัน

$$\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \text{ และ } \Sigma_4 = \Sigma_5 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

รวมถึงพิจารณาขนาดตัวอย่างในสองกรณีย่อย ได้แก่ $n=300$ และ $n=1,200$

กำหนดขนาดตัวอย่าง $n = 300$ แบ่งออกเป็น 3 กรณี ดังตารางที่ 3.1.1.3

ตารางที่ 3.1.1.3 จำนวนข้อมูลแต่ละกลุ่มสำหรับขนาดตัวอย่างเท่ากับ 300 เมื่อจำนวนกลุ่มเท่ากับ 5

กรณี	จำนวนข้อมูลแต่ละกลุ่ม				
	n_1	n_2	n_3	n_4	n_5
1	60	60	60	60	60
2	40	50	60	70	80
3	80	70	60	50	40

และสำหรับ $n = 1,200$ แบ่งออกเป็น 3 กรณี ดังตารางที่ 3.1.1.4

ตารางที่ 3.1.1.4 จำนวนข้อมูลแต่ละกลุ่มสำหรับขนาดตัวอย่างเท่ากับ 1,200 เมื่อจำนวนกลุ่มเท่ากับ 5

กรณี	จำนวนข้อมูลแต่ละกลุ่ม				
	n_1	n_2	n_3	n_4	n_5
1	240	240	240	240	240
2	160	200	240	280	320
3	320	280	240	200	160

3.1.2 กำหนดขนาดตัวอย่าง

กำหนดขนาดตัวอย่าง (n) เท่ากับ 300 และ 1,200 แทน ข้อมูลขนาดเล็กและขนาดใหญ่ตามลำดับ สำหรับข้อมูลที่มีการแจกแจงปกติและข้อมูลมีลักษณะเป็น نرم และกำหนดจำนวนกลุ่ม (K) เท่ากับ 3 กลุ่ม และ 5 กลุ่ม

3.1.3 วิธีการระบุค่าพารามิเตอร์ Eps

วิธีการระบุค่าพารามิเตอร์ Eps มีดังนี้

3.1.3.1 วิธี Daszykowski วิธีที่ใช้ระยะทางจากทุกจุดในข้อมูลไปยังจุดที่ห่างออกไปเป็นลำดับที่ k มาประกอบ โดยสุ่มให้มีการกระจายแบบสม่ำเสมอทั่วทั้งบริเวณ

3.1.3.2 วิธี Xia วิธีที่ได้จากการ fit กราฟฮิสโทแกรมระยะห่างด้วยการแจกแจงเกาส์เซียนแบบผกผัน ด้วยค่าเฉลี่ย μ และพารามิเตอร์รูปร่าง (shape parameter) λ

$$Eps_k = \frac{\mu_k \sqrt{9\mu_k^2 + 4\lambda_k^2} - 3\mu_k^2}{2\lambda_k}$$

3.1.3.3 วิธีที่ประยุกต์จากวิธีของ Xia ปรับปรุงหาค่า Eps นอกเหนือจากสูตรผ่านกระบวนการ fit ด้วยการแจกแจงเกาส์เซียนแบบผกผัน โดยใช้เป็นค่าเฉลี่ย ค่ามัธยฐาน ค่า quantile ที่ 75 หรือ 95

3.1.3.4 วิธี Karami วิธีที่ใช้การคำนวณจากสูตร

$$Eps = \left(\frac{\prod_{d=1}^D (\max_{j \in \{1,2,\dots,n\}} x_{jd} - \min_{j \in \{1,2,\dots,n\}} x_{jd}) \times k \times \gamma(0.5 \times D + 1)}{n \times \sqrt{\pi^D}} \right)^{1/D}$$

เมื่อ X คือ เมทริกซ์ข้อมูล $\begin{bmatrix} x_{11} & \cdots & x_{1D} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nD} \end{bmatrix}$

D คือ จำนวนมิติของเซตข้อมูลหรือจำนวนตัวแปร

n คือ จำนวนข้อมูล

k คือ MinPts

γ คือ ฟังก์ชันแกมมา

3.1.3.5 วิธีที่คำนวณจากระยะทางของแต่ละจุดไปยังทุกจุดมาคำนวณเป็นค่าเฉลี่ย ค่ามัธยฐาน ค่า quantile ที่ 75 หรือ 95 เพื่อนำค่าดังกล่าวเป็นค่า Eps ในที่นี้เรียกว่าวิธี mean, median, P75 และ P95 ตามลำดับ

3.1.4 จำลองข้อมูลเชิงตัวเลข และประมวลผลโดยใช้โปรแกรม R เวอร์ชัน 3.1.3

3.1.5 กำหนดการจำลองข้อมูล 500 ชุดต่อ 1 กรณี

3.2 ขั้นตอนในการดำเนินการศึกษา

3.2.1 กำหนดข้อมูลไม่ปกติใน 4 ลักษณะได้แก่ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน ข้อมูลวงแหวนรูปแบบที่ 1 ข้อมูลวงแหวนรูปแบบที่ 2 ข้อมูลมีลักษณะเป็นมุม และข้อมูลที่มีการแจกแจงแบบปกติ 2 ตัวแปร 24 ลักษณะ

3.2.2 นำข้อมูลแบบต่าง ๆ มาคำนวณหาค่า Eps ด้วยวิธีต่อไปนี้

3.2.2.1 วิธี Daszykowski วิธีที่ใช้ระยะทางจากทุกจุดในข้อมูลไปยังจุดที่ห่างออกไปเป็นลำดับที่ k มาประกอบ โดยสุ่มให้มีการกระจายแบบสม่ำเสมอทั่วทั้งบริเวณ

3.2.2.2 วิธี Xia วิธีที่ได้จากการ fit กราฟฮิสโทแกรมระยะห่างด้วยการแจกแจงเกาส์เซียนแบบผกผัน ด้วยค่าเฉลี่ย μ และพารามิเตอร์รูปร่าง (shape parameter) λ ฐานนิยมที่ได้ นั่นคือ Eps

$$Eps_k = \frac{\mu_k \sqrt{9\mu_k^2 + 4\lambda_k^2 - 3\mu_k^2}}{2\lambda_k}$$

3.2.2.3 วิธีที่ประยุกต์จากวิธีของ Xia ปรับหาค่า Eps นอกเหนือจากสูตรผ่านกระบวนการ fit ด้วยการแจกแจงเกาส์เซียนแบบผกผัน โดยใช้เป็นค่าเฉลี่ย ค่ามัธยฐาน ค่า quantile ที่ 75 หรือ 95 ในที่นี้เรียกว่าวิธี Xia mean, Xia median, Xia P75 และ Xia P95 ตามลำดับ

3.2.2.4 วิธี Karami วิธีที่ใช้การคำนวณจากสูตร

$$Eps = \left(\frac{\prod_{d=1}^D \left(\max_{j \in \{1,2,\dots,n\}} x_{jd} - \min_{j \in \{1,2,\dots,n\}} x_{jd} \right) \times k \times \gamma(0.5 \times D + 1)}{n \times \sqrt{\pi^D}} \right)^{1/D}$$

เมื่อ X คือ เมตริกซ์ข้อมูล $\begin{bmatrix} x_{11} & \cdots & x_{1D} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nD} \end{bmatrix}$

D คือ จำนวนมิติของเซตข้อมูลหรือจำนวนตัวแปร

n คือ จำนวนข้อมูล

k คือ MinPts

γ คือ ฟังก์ชันแกมมา

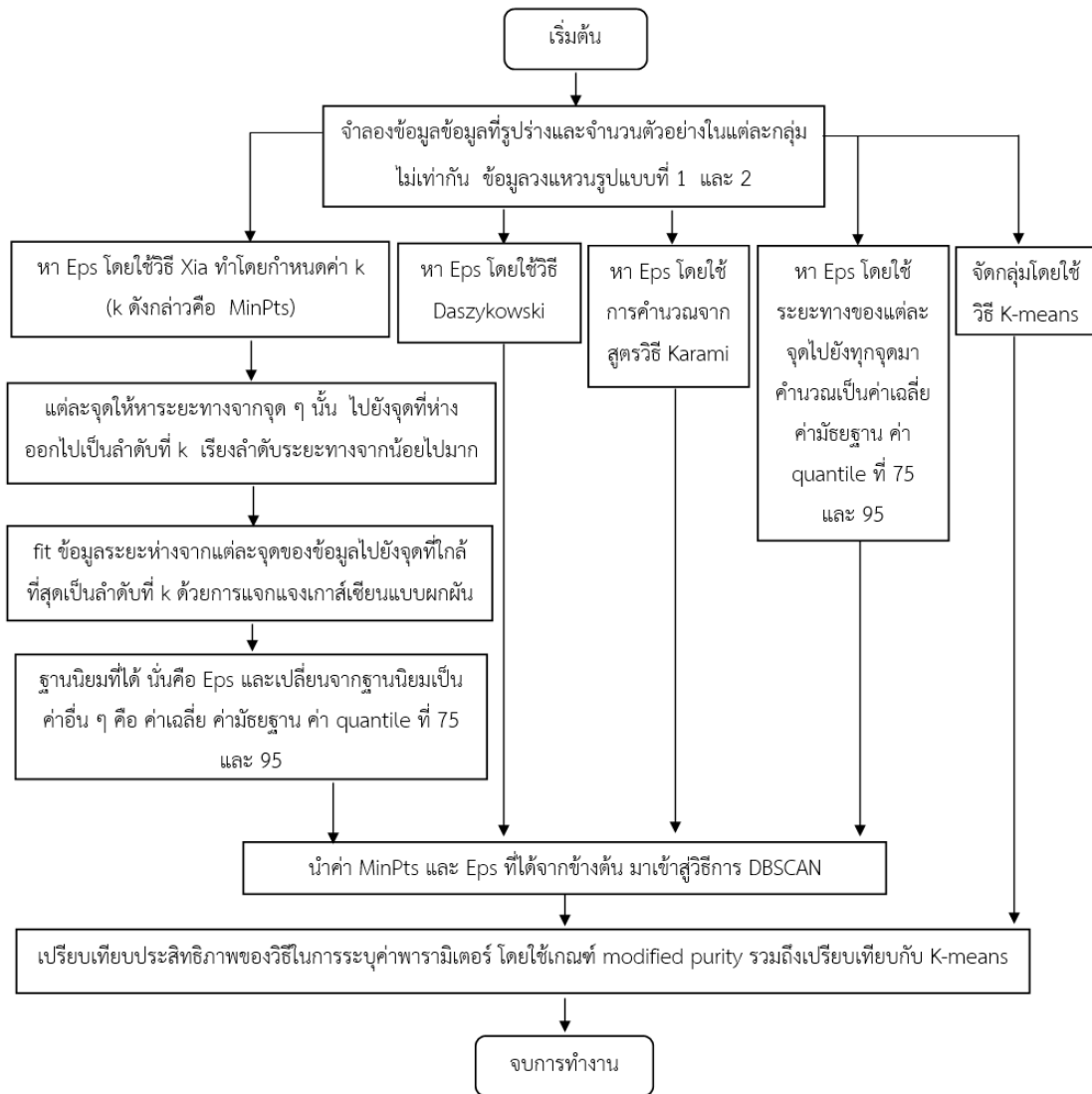
3.2.2.5 วิธีที่คำนวณจากระยะทางของแต่ละจุดไปยังทุกจุดมาคำนวณเป็นค่าเฉลี่ยค่ามัธยฐาน ค่า quantile ที่ 75 หรือ 95 เพื่อนำค่าดังกล่าวเป็นค่า Eps ในที่นี้เรียกว่าวิธี mean, median, P75 และ P95 ตามลำดับ

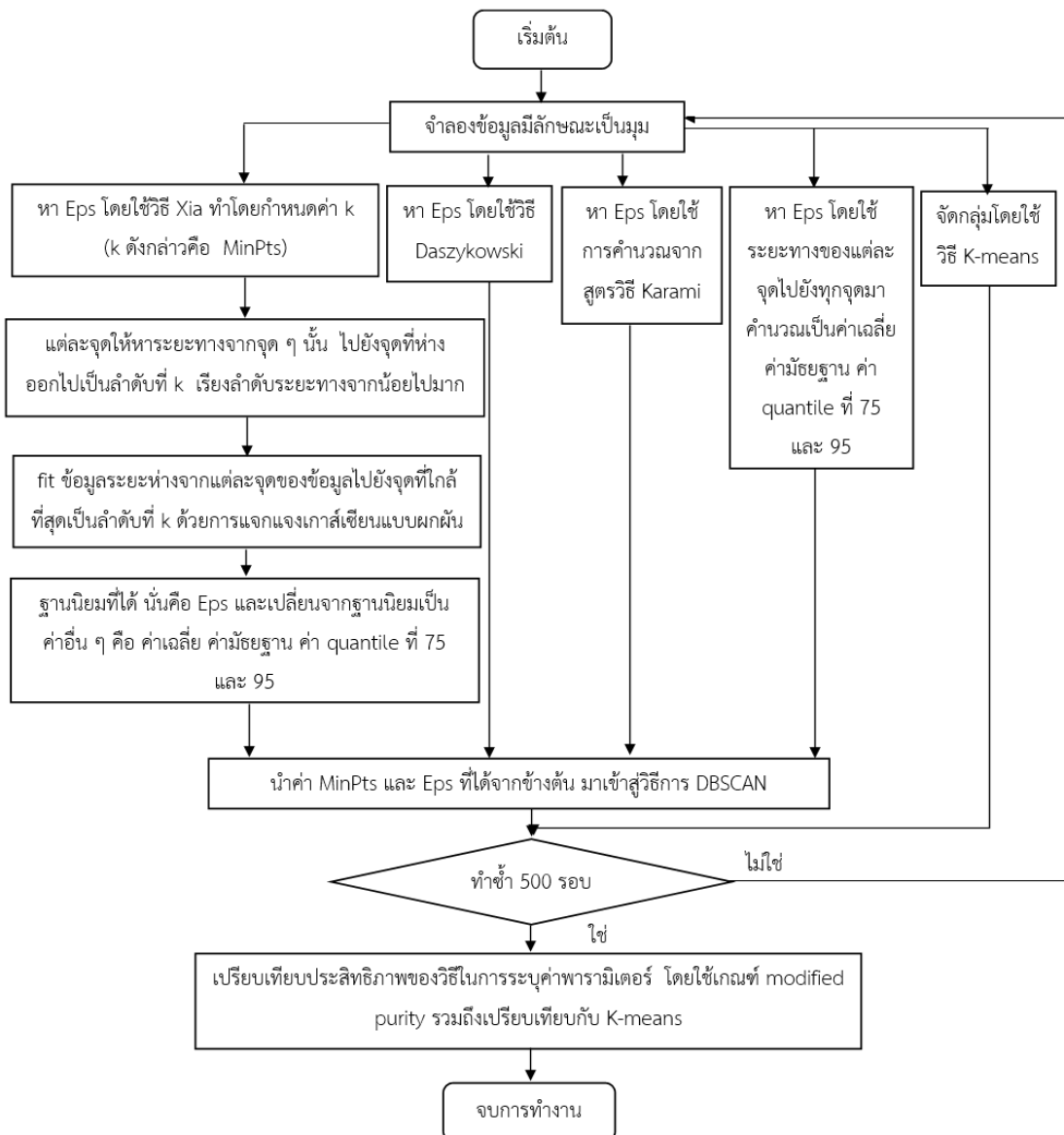
3.2.3 นำค่า MinPts และ Eps ที่ได้จากข้างต้น มาเข้าสู่วิธีการ DBSCAN

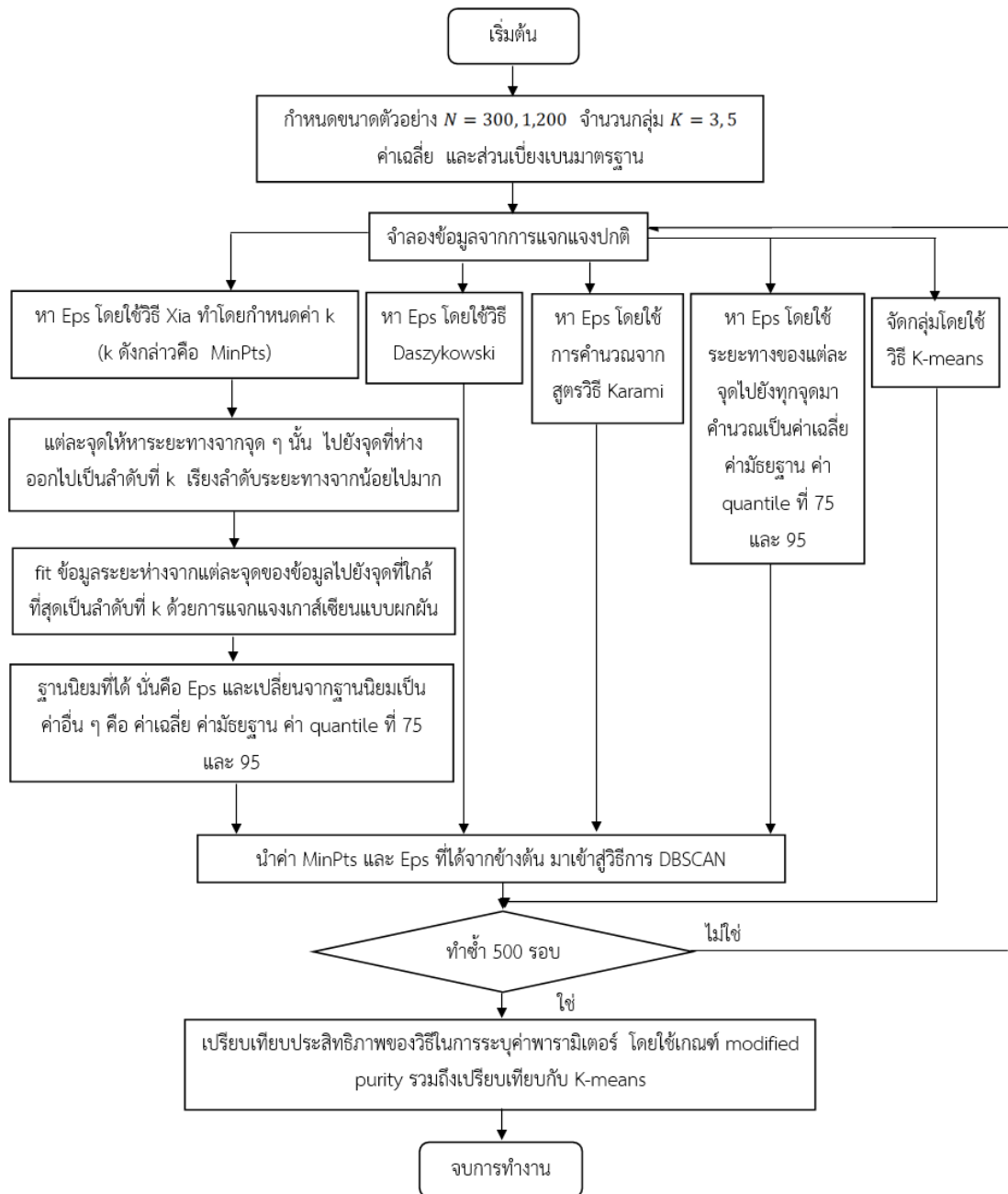
3.2.4 เปรียบเทียบประสิทธิภาพของวิธีในการระบุค่าพารามิเตอร์ด้วยค่า modified purity โดยนำเสนอในรูปแบบของกราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จากวิธีต่าง ๆ ด้วยค่า MinPts ที่ต่างกัน และกราฟแสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ด้วยค่า MinPts ที่ต่างกัน รวมถึงเปรียบเทียบกับ K-means ด้วย



3.3 ขั้นตอนการทำงานของโปรแกรม







บทที่ 4

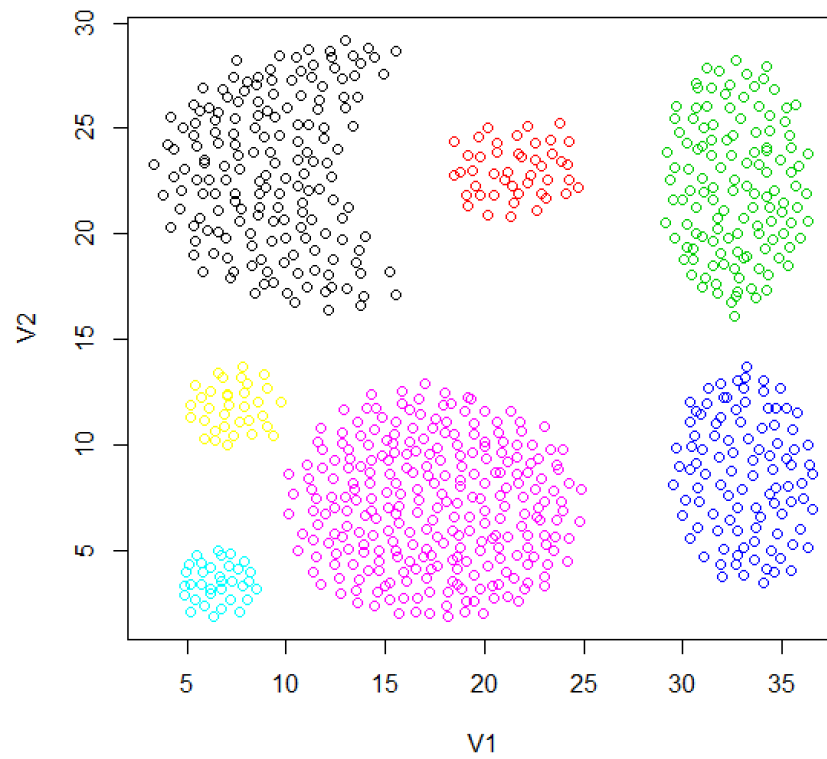
ผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีที่ใช้ในการระบุค่าพารามิเตอร์ Eps นั้นคือวิธีของ Daszykowski วิธี mean วิธี median วิธี P75 วิธี P95 วิธีของ Xia วิธีที่ประยุกต์จากวิธีของ Xia (Xia mean Xia median Xia P75 และ Xia P95) และวิธีของ Karami โดยการเปรียบเทียบประสิทธิภาพนี้จะพิจารณาจากค่า modified purity ทั้งนี้กำหนดขอบเขตข้อมูลคือข้อมูล 2 มิติ ซึ่งเป็นข้อมูลที่มีการแจกแจงไม่ปกติ 4 รูปแบบ ได้แก่ ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มหลากหลาย ข้อมูลวงแหวนรูปแบบที่ 1 ข้อมูลวงแหวนรูปแบบที่ 2 ข้อมูลมีลักษณะเป็นมุม และข้อมูลที่มีการแจกแจงปกติ 24 รูปแบบ ตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนที่กำหนด

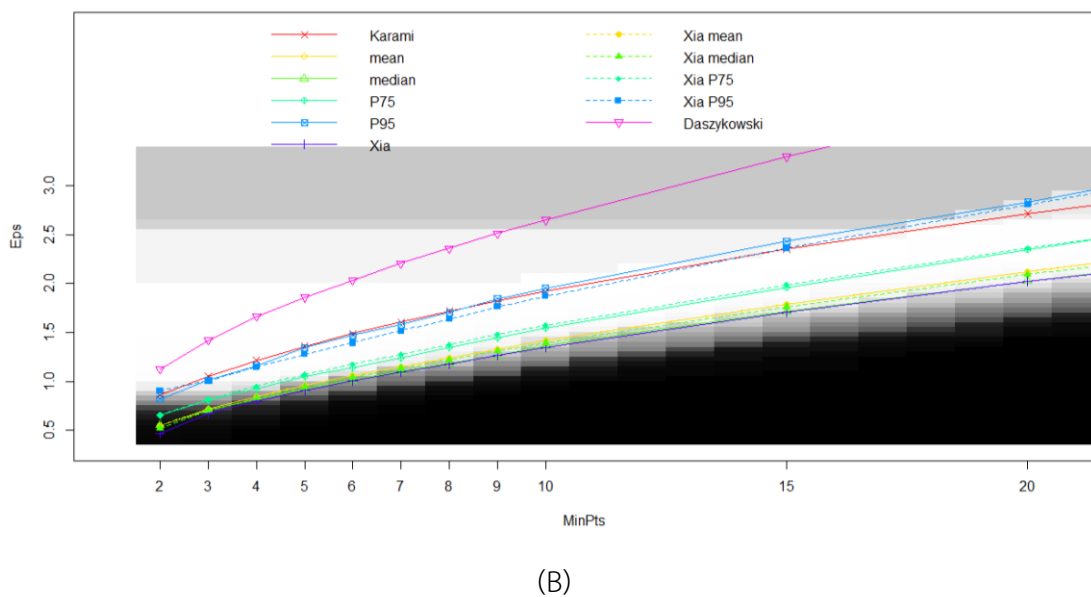
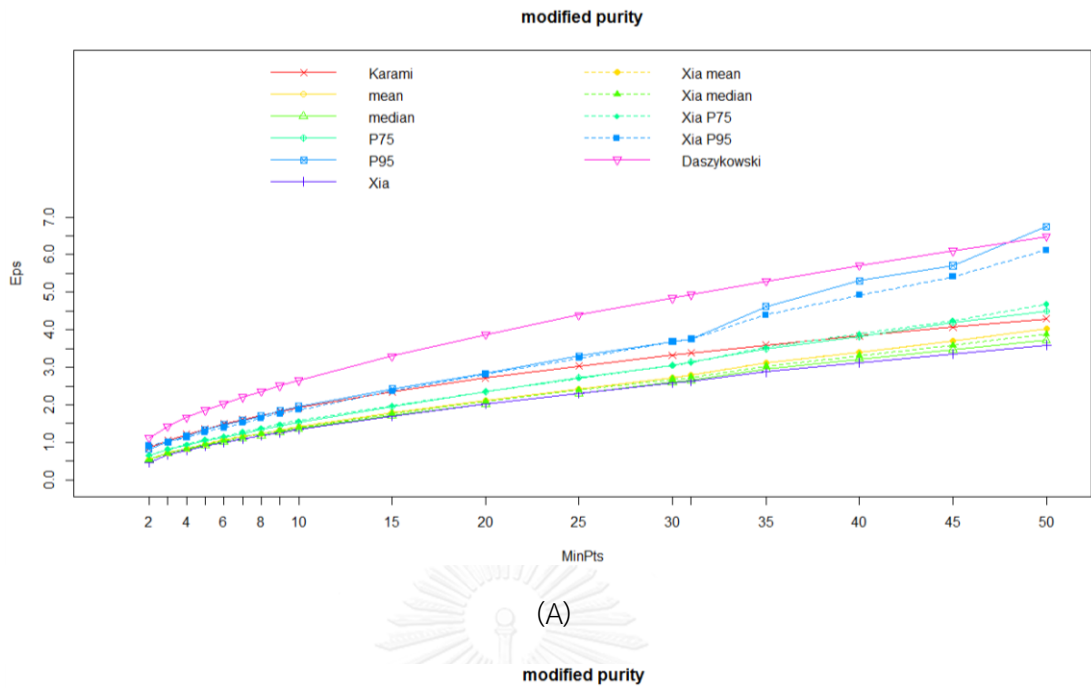
อักษรและสัญลักษณ์ต่าง ๆ ที่ปรากฏในการนำเสนอผลการวิจัยมีดังนี้

N	แทน ขนาดของตัวอย่าง
μ	แทน เวกเตอร์ค่าเฉลี่ย
Σ	แทน เมทริกซ์ความแปรปรวน
MinPts	แทน จำนวนจุดที่น้อยที่สุดที่อยู่ภายใน neighborhood ของจุด ๆ หนึ่ง ในรัศมี Eps ซึ่ง ใช้บ่งชี้ว่าจุด ๆ นั้นเป็นจุดชนิดใด
Eps	แทน รัศมีที่ใช้กำหนด neighborhood ของจุด ๆ หนึ่ง

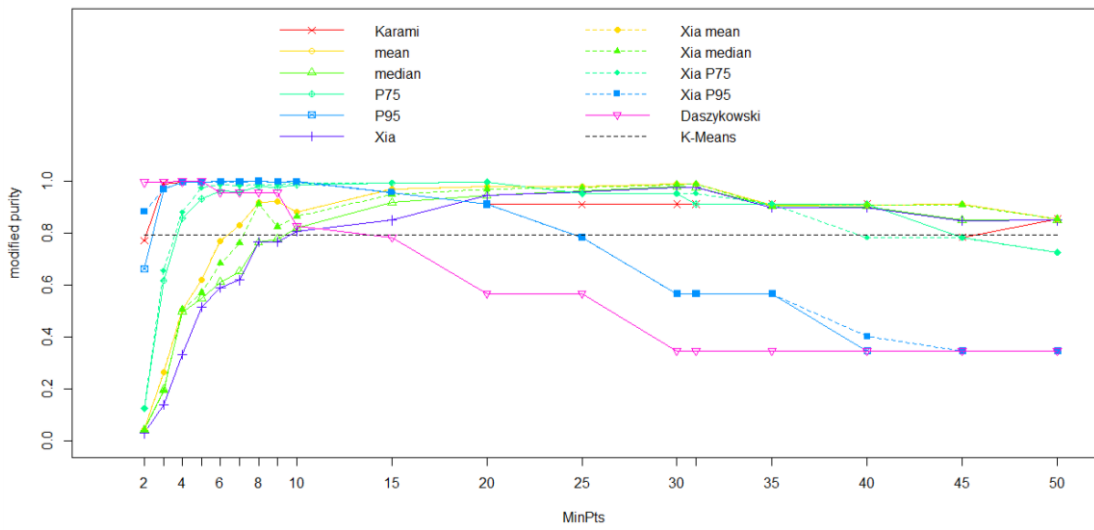
4.1 ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน



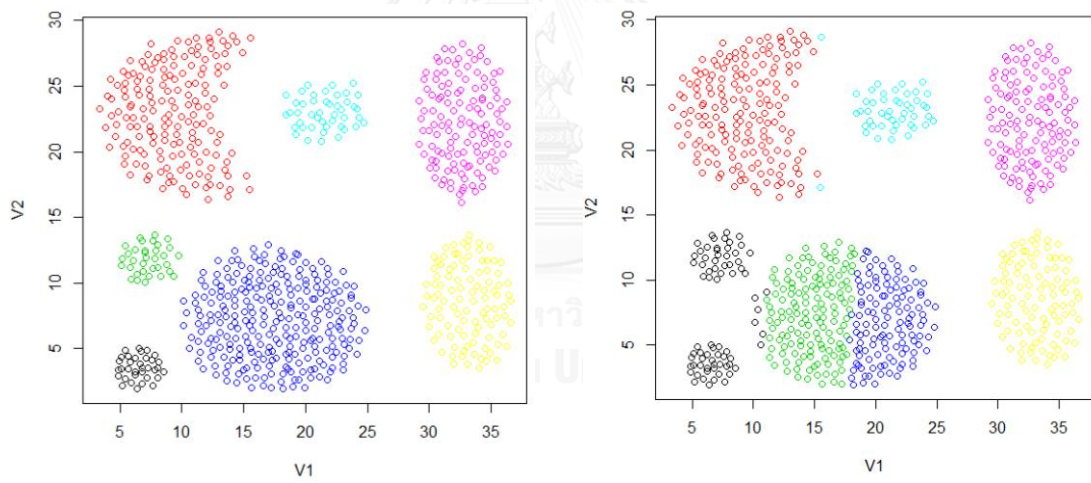
รูปที่ 4.1.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.1



รูปที่ 4.1.2 (A) กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.1 และ (B) กราฟขยายส่วนต้นของรูป (A) โดยแสดงค่า modified purity ด้วยความสว่างของพื้นหลังของข้อมูลกรณี 4.1



รูปที่ 4.1.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.1



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.1.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.1

สำหรับข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน พบว่าจากรูปที่ 4.1.2 (A) และ (B) ซึ่งรูป 4.1.2 (B) มีสีพื้นหลังเป็นค่า modified purity ไล่ตามระดับสีจาก 0 ถึง 1 (ดำไล่ไปถึงขาว) ถ้าเป็นสีเข้มแสดงว่าการใช้ค่าพารามิเตอร์นั้น ๆ ในการทำ DBSCAN ไม่เหมาะสม แต่ถ้าเป็นสีขาวแสดงว่าการใช้พารามิเตอร์นั้น ๆ ในการทำ DBSCAN เหมาะสม เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95 มีค่า Eps สูงกว่าวิธีที่เหลือ แต่ในช่วง MinPts ค่าน้อย ๆ วิธี P95 และ Xia P95 ให้ค่า Eps ใกล้เคียงกับวิธี Karami

จากรูปที่ 4.1.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าในช่วง MinPts 2-5 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าอยู่ระหว่าง 0.996-1 ซึ่งสูงกว่ากลุ่มอื่น แต่เมื่อค่า MinPts สูงกว่านั้นแล้ว ค่า Eps จะสูงเกินไป ทำให้ค่า modified purity ค่อย ๆ ลดลง จึงทำให้พารามิเตอร์นั้น ๆ แบ่งกลุ่มข้อมูลไม่เหมาะสมเมื่อ MinPts มีค่ามากกว่า 9 จนกระทั่งที่ MinPts 30 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.346 (270/781) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 3-10 ค่า modified purity มีค่าใกล้เคียง 1 จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 40 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.346 (270/781) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าในช่วง MinPts 3-5 ค่า Eps อยู่ในระดับที่แบ่งกลุ่มได้ค่อนข้างดี อีกทั้งในช่วง MinPts 6-10 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าอยู่ระหว่าง 0.997-1 ซึ่งสูงกว่ากลุ่มอื่น แต่เมื่อค่า MinPts สูงกว่านั้นแล้ว ค่า Eps จะสูงเกินไป

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 จนกระทั่งสูงสุดที่ MinPts 20 จากนั้น modified purity มีค่าลดลง

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-8 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ และลดลงช่วง MinPts ค่าปลายตั้งแต่ MinPts 15-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง จนกระทั่งสูงสุดที่ MinPts 30-31 จากนั้น modified purity มีค่าลดลง

จากรูปที่ 4.1.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มแตกต่างกับแผนภาพกระจายของข้อมูลต้นแบบ กล่าวคือจากรูปต้นแบบเมื่อวิเคราะห์กลุ่มจุดสีน้ำเงินด้วย K-means พบว่าแบ่งย่อยออกเป็น 2 กลุ่ม ทั้งที่ความเป็นจริงควรอยู่เป็นกลุ่มเดียวกัน นอกจากนี้กลุ่มจุดสีเขียวและกลุ่มจุดสีดำจากรูปต้นแบบอยู่คนละกลุ่มกัน แต่เมื่อวิเคราะห์ด้วย K-means พบว่าอยู่กลุ่มเดียวกัน ซึ่งเมื่อพิจารณาจากรูปที่ 4.1.3 พบว่าค่า modified purity ของ K-means เป็น 0.7912932

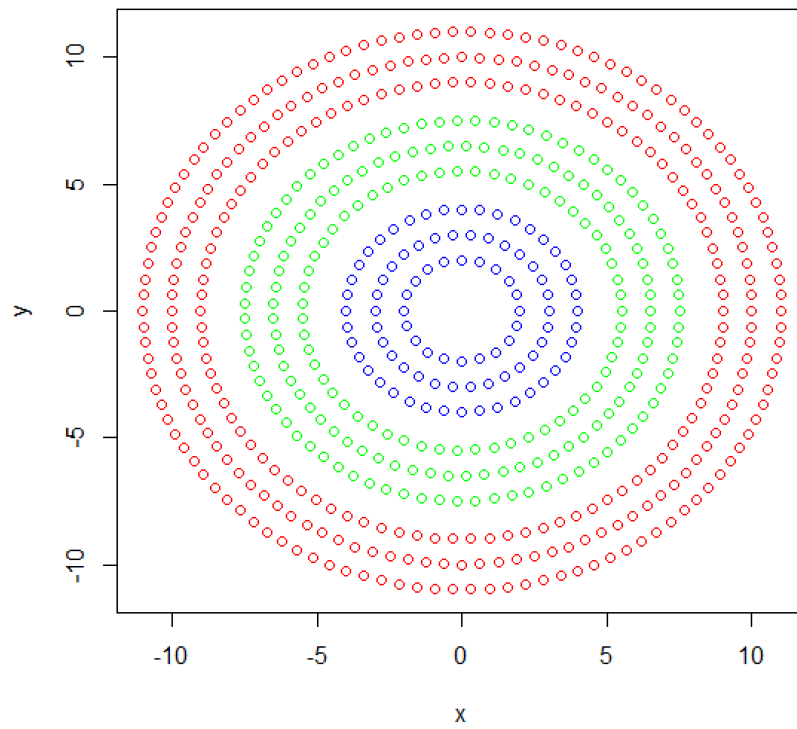
ตารางที่ 4.1.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีที่ 4.1

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.8580	0.5502	0.5500	0.6500	0.8201	0.4618	0.5501	0.5199	0.6528	0.9021	1.1226
3	1.0508	0.7202	0.7071	0.8078	1.0112	0.6691	0.7202	0.7030	0.8156	1.0086	1.4215
4	1.2133	0.8446	0.8201	0.9220	1.1597	0.7959	0.8446	0.8282	0.9465	1.1457	1.6643
5	1.3566	0.9582	0.9220	1.0500	1.3435	0.9096	0.9582	0.9419	1.0674	1.2767	1.8585
6	1.4860	1.0549	1.0124	1.1402	1.4705	1.0052	1.0549	1.0382	1.1712	1.3918	2.0307
7	1.6051	1.1480	1.1011	1.2379	1.5788	1.0939	1.1480	1.1298	1.2745	1.5147	2.2062
8	1.7159	1.2381	1.1800	1.3463	1.7029	1.1785	1.2381	1.2181	1.3759	1.6382	2.3590
9	1.8200	1.3290	1.2659	1.4422	1.8446	1.2640	1.3290	1.3072	1.4780	1.7623	2.5110
10	1.9185	1.4150	1.3647	1.5443	1.9506	1.3476	1.4150	1.3924	1.5717	1.8695	2.6499
15	2.3496	1.7882	1.7117	1.9602	2.4300	1.7033	1.7882	1.7597	1.9859	2.3616	3.2978
20	2.7131	2.1229	2.0248	2.3479	2.8293	2.0213	2.1229	2.0888	2.3585	2.8067	3.8748
25	3.0334	2.4268	2.3049	2.7295	3.2973	2.3001	2.4268	2.3843	2.7071	3.2469	4.3908
30	3.3229	2.7248	2.6043	3.0602	3.6831	2.5759	2.7248	2.6748	3.0460	3.6690	4.8430
31	3.3778	2.7846	2.6575	3.1484	3.7590	2.6302	2.7846	2.7328	3.1151	3.7572	4.9363
35	3.5891	3.1091	2.9517	3.4828	4.6157	2.8750	3.1091	3.0303	3.5317	4.3958	5.2988
40	3.8369	3.4060	3.2191	3.8138	5.3052	3.1140	3.4060	3.3075	3.8958	4.9221	5.7162
45	4.0697	3.6944	3.4713	4.2012	5.7064	3.3537	3.6944	3.5794	4.2427	5.4083	6.1032
50	4.2898	4.0327	3.7216	4.4972	6.7502	3.5836	4.0327	3.8805	4.6810	6.1159	6.4822

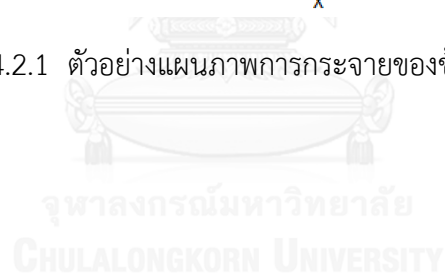
ตารางที่ 4.1.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.1

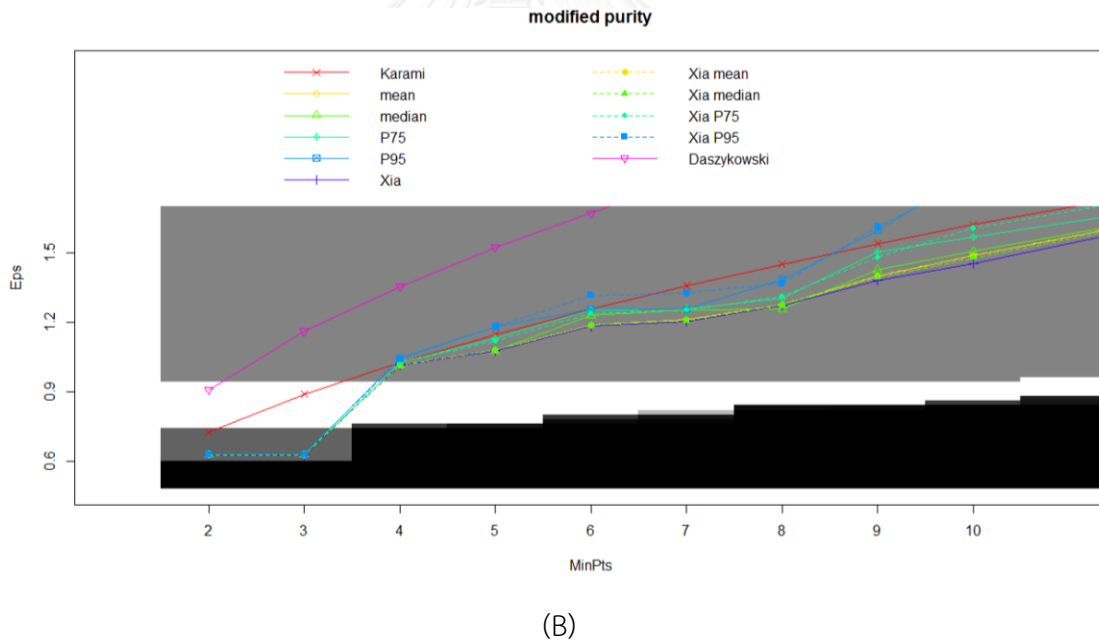
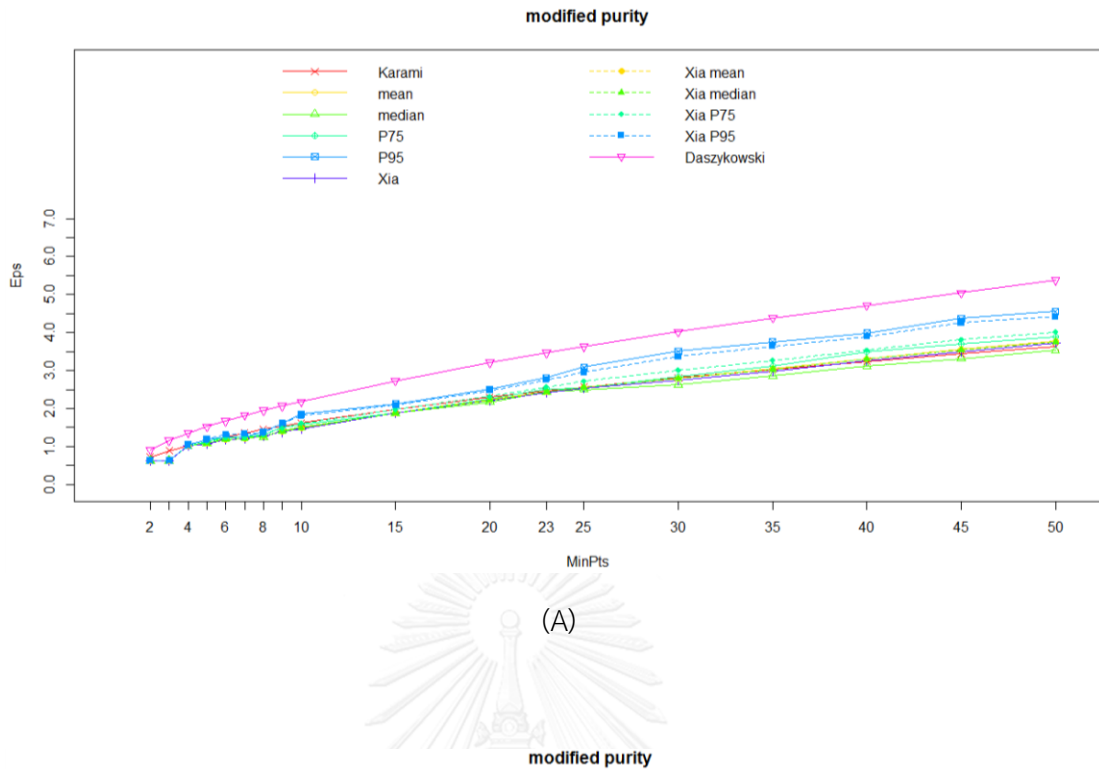
MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7721	0.0423	0.0423	0.1229	0.6633	0.0282	0.0423	0.0397	0.1255	0.8822	0.9962
3	0.9910	0.2650	0.1959	0.6172	0.9706	0.1383	0.2650	0.1921	0.6556	0.9706	0.9974
4	0.9974	0.5045	0.4955	0.8566	0.9974	0.3303	0.5045	0.5045	0.8796	0.9974	1.0000
5	0.9974	0.6184	0.5493	0.9321	0.9974	0.5147	0.6184	0.5698	0.9744	0.9962	1.0000
6	0.9987	0.7708	0.6108	0.9641	0.9987	0.5890	0.7708	0.6825	0.9872	0.9974	0.9565
7	0.9987	0.8310	0.6517	0.9590	0.9987	0.6184	0.8310	0.7606	0.9834	0.9974	0.9565
8	1.0000	0.9181	0.7644	0.9808	1.0000	0.7644	0.9181	0.9117	0.9872	0.9987	0.9565
9	0.9974	0.9219	0.7772	0.9744	0.9974	0.7644	0.9219	0.8233	0.9846	0.9974	0.9565
10	0.9987	0.8796	0.8195	0.9859	0.9987	0.8079	0.8796	0.8630	0.9923	0.9974	0.8271
15	0.9552	0.9706	0.9168	0.9923	0.9565	0.8489	0.9706	0.9475	0.9936	0.9565	0.7836
20	0.9129	0.9782	0.9462	0.9962	0.9117	0.9462	0.9782	0.9693	0.9962	0.9117	0.5659
25	0.9104	0.9795	0.9641	0.9526	0.7836	0.9577	0.9795	0.9757	0.9526	0.7836	0.5659
30	0.9117	0.9885	0.9795	0.9539	0.5659	0.9744	0.9885	0.9859	0.9539	0.5659	0.3457
31	0.9117	0.9885	0.9757	0.9117	0.5659	0.9744	0.9885	0.9885	0.9539	0.5659	0.3457
35	0.9117	0.9091	0.9052	0.9117	0.5659	0.8976	0.9091	0.9052	0.9117	0.5659	0.3457
40	0.9117	0.9078	0.9027	0.9117	0.3457	0.8988	0.9078	0.9078	0.7836	0.4020	0.3457
45	0.7836	0.9104	0.8502	0.7836	0.3457	0.8464	0.9104	0.9091	0.7836	0.3457	0.3457
50	0.8540	0.8528	0.8515	0.7260	0.3457	0.8489	0.8528	0.8528	0.7260	0.3457	0.3457

4.2 ข้อมูลวงแหวนรูปแบบที่ 1

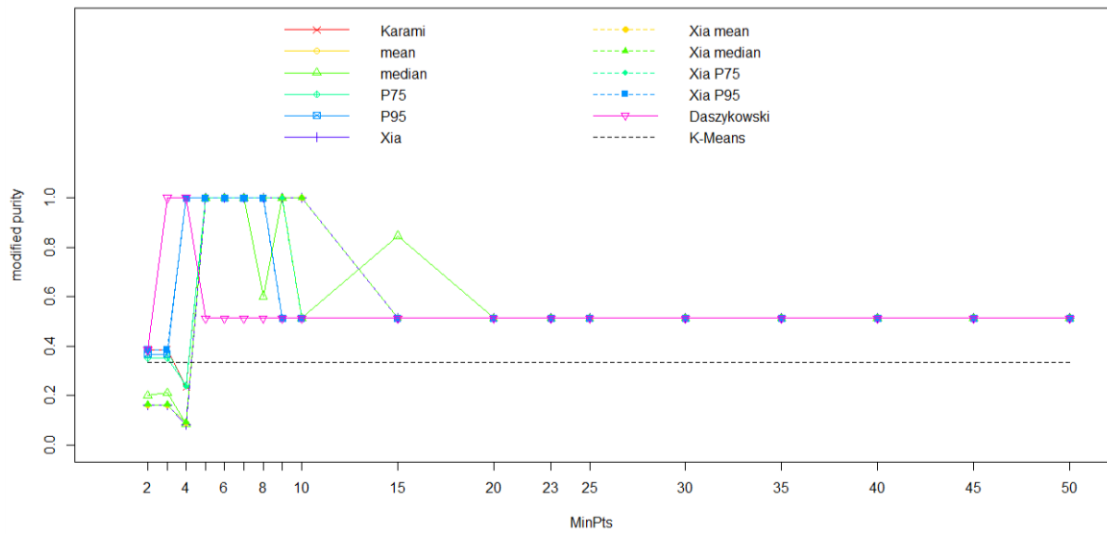


รูปที่ 4.2.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.2

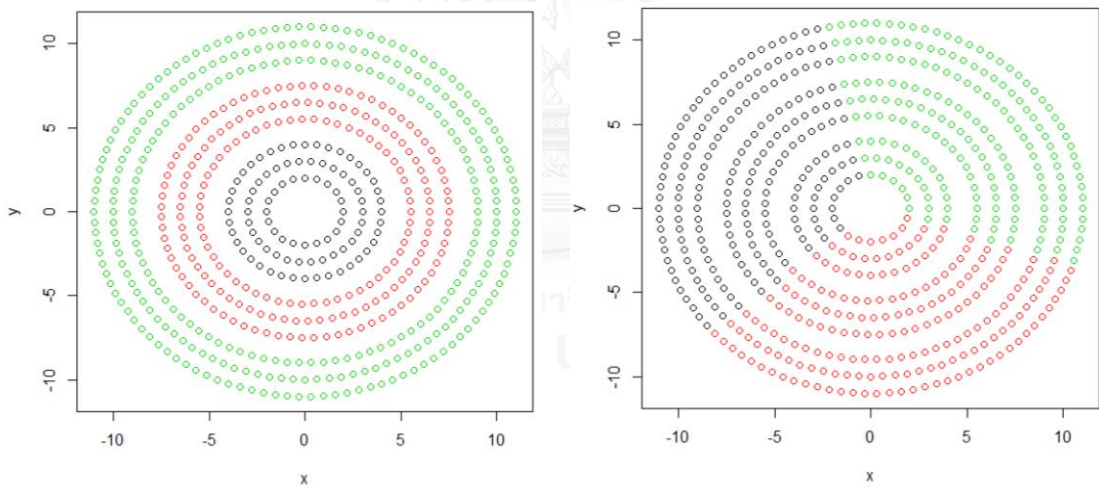




รูปที่ 4.2.2 (A) กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.2 และ (B) กราฟขยายส่วนต้นของรูป (A) โดยแสดงค่า modified purity ด้วยความสว่างของพื้นหลังของข้อมูลกรณี 4.2



รูปที่ 4.2.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีศึกษาที่ 4.2



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.2.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีศึกษาที่ 4.2

สำหรับข้อมูลวงแหวนรูปแบบที่ 1 พบว่าจากรูปที่ 4.2.2 (A) และ (B) ซึ่งรูป 4.2.2 (B) มีสีพื้นหลังเป็นค่า modified purity ไล่ตามระดับสีจาก 0 ถึง 1 (ดำไปถึงขาว) ถ้าเป็นสีเข้มแสดงว่าการใช้ค่าพารามิเตอร์นั้น ๆ ในการทำ DBSCAN ไม่เหมาะสม แต่ถ้าเป็นสีขาวแสดงว่าการใช้พารามิเตอร์นั้น ๆ ในการทำ DBSCAN เหมาะสม เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps ดังนี้ วิธี Daszykowski สูงกว่าวิธีที่เหลือ และทุกวิธียกเว้นวิธี Daszykowski ให้ค่า Eps ใกล้เคียงกัน

จากรูปที่ 4.2.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าในช่วง MinPts 3-4 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 ซึ่งสูงกว่ากลุ่มอื่น แต่เมื่อค่า MinPts สูงกว่านั้นแล้ว ค่า Eps จะสูงเกินไป ทำให้ค่า modified purity ค่อย ๆ ลดลง จึงทำให้พารามิเตอร์นั้น ๆ แบ่งกลุ่มข้อมูลไม่เหมาะสมเมื่อ MinPts มีค่ามากกว่า 4 ค่า modified purity ลู่เข้าสู่ 0.513 (300/585) ซึ่งเป็นกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 4-8 ค่า modified purity มีค่าเป็น 1 จากนั้นลดลง จนกระทั่งที่ MinPts 9 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.513 (300/585) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าในช่วง MinPts 2-4 ค่า Eps มีค่าต่ำเกินไป จากนั้นในช่วง MinPts 5-8 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 แต่เมื่อค่า MinPts ค่อย ๆ เพิ่มขึ้น ค่า Eps ก็จะมีค่าสูงเกินไป ทำให้ค่า modified purity ค่อย ๆ ลดลง จนกระทั่งที่ MinPts 9 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.513 (300/585) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 พบว่าที่ค่า MinPts 5-8 ค่า modified purity มีค่าเป็น 1 จากนั้นลดลง จนกระทั่งที่ MinPts 9 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.513 (300/585) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Xia P75 พบว่าช่วง MinPts 5-9 ค่า modified purity มีค่าเป็น 1 จากนั้นลดลง จนกระทั่งที่ MinPts 10 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.513 (300/585) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี mean และ Xia mean พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำมาก และที่ MinPts 5-10 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.513 (300/585) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี median พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก และที่ MinPts 5-7, 9 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 จนกระทั่งที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.513 (300/585) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก และช่วง MinPts 5-10 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.513 (300/585) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก และที่ MinPts 5-10 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.513 (300/585) ด้วยเหตุผลเดียวกับวิธี Daszykowski

จากรูปที่ 4.2.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มแตกต่างกับแผนภาพกระจายของข้อมูลต้นแบบ กล่าวคือจากรูปต้นแบบเมื่อวิเคราะห์กลุ่มด้วย K-means พบว่าแบ่งย่อยออกเป็น 3 ส่วนคล้ายใบพัด ทั้งที่ความเป็นจริงควรแบ่งชั้น ๆ ที่อยู่ใกล้กันเป็นกลุ่มเดียวกัน ซึ่งเมื่อพิจารณาจากรูปที่ 4.2.3 พบว่าค่า modified purity ของ K-means เป็น 0.3350427

ตารางที่ 4.2.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.2

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7256	0.6280	0.6282	0.6282	0.6282	0.6280	0.6280	0.6280	0.6283	0.6288	0.9079
3	0.8887	0.6280	0.6282	0.6282	0.6282	0.6280	0.6280	0.6280	0.6283	0.6288	1.1603
4	1.0262	1.0161	1.0160	1.0283	1.0439	1.0158	1.0161	1.0161	1.0262	1.0410	1.3549
5	1.1473	1.0811	1.0761	1.1317	1.1818	1.0764	1.0811	1.0795	1.1194	1.1795	1.5243
6	1.2568	1.1895	1.2315	1.2552	1.2560	1.1825	1.1895	1.1872	1.2387	1.3167	1.6712
7	1.3575	1.2095	1.2539	1.2556	1.2560	1.2038	1.2095	1.2076	1.2545	1.3252	1.8152
8	1.4512	1.2759	1.2558	1.3050	1.3850	1.2725	1.2759	1.2759	1.3129	1.3679	1.9473
9	1.5392	1.3989	1.4270	1.5044	1.5972	1.3823	1.3989	1.3933	1.4796	1.6131	2.0634
10	1.6225	1.4898	1.5060	1.5686	1.8498	1.4554	1.4898	1.4783	1.6078	1.8138	2.1784
15	1.9871	1.8911	1.8764	1.8815	2.1151	1.8804	1.8911	1.8875	1.9674	2.0882	2.7338
20	2.2945	2.2217	2.1724	2.2662	2.5078	2.2070	2.2217	2.2168	2.3184	2.4727	3.2078
23	2.4606	2.4476	2.4404	2.5067	2.8095	2.4281	2.4476	2.4411	2.5642	2.7520	3.4666
25	2.5654	2.5674	2.5003	2.5262	3.0921	2.5379	2.5674	2.5576	2.7133	2.9539	3.6288
30	2.8102	2.7996	2.6314	2.8284	3.5058	2.7429	2.7996	2.7807	3.0081	3.3676	4.0178
35	3.0354	3.0445	2.8609	3.1287	3.7515	2.9871	3.0445	3.0253	3.2638	3.6394	4.3830
40	3.2450	3.3151	3.1111	3.5035	3.9872	3.2651	3.3151	3.2984	3.5299	3.8912	4.7118
45	3.4418	3.5552	3.3155	3.6933	4.3690	3.4861	3.5552	3.5321	3.8148	4.2607	5.0453
50	3.6280	3.7729	3.5343	3.8848	4.5586	3.7174	3.7729	3.7544	4.0144	4.4199	5.3771

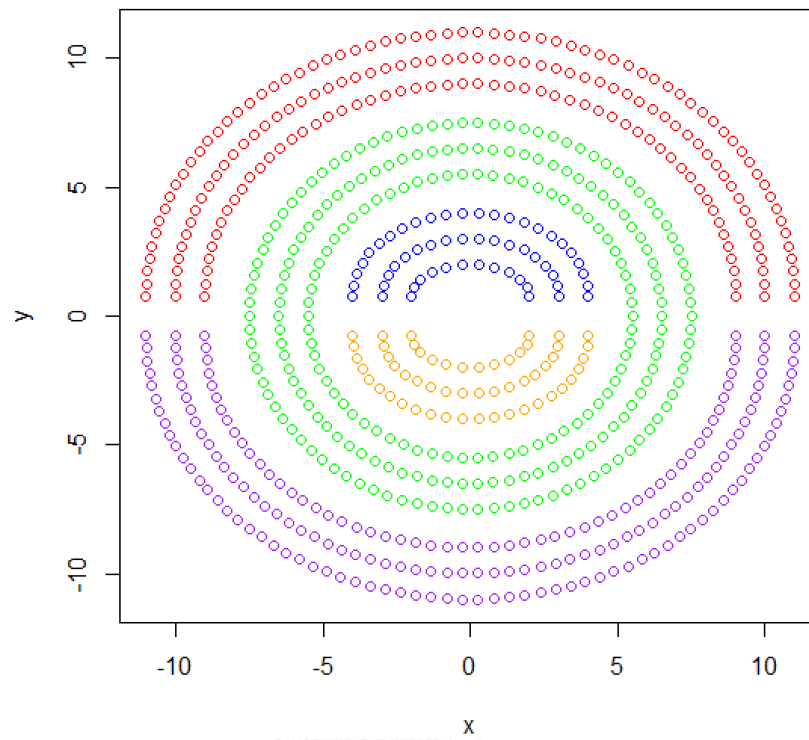


ตารางที่ 4.2.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.2

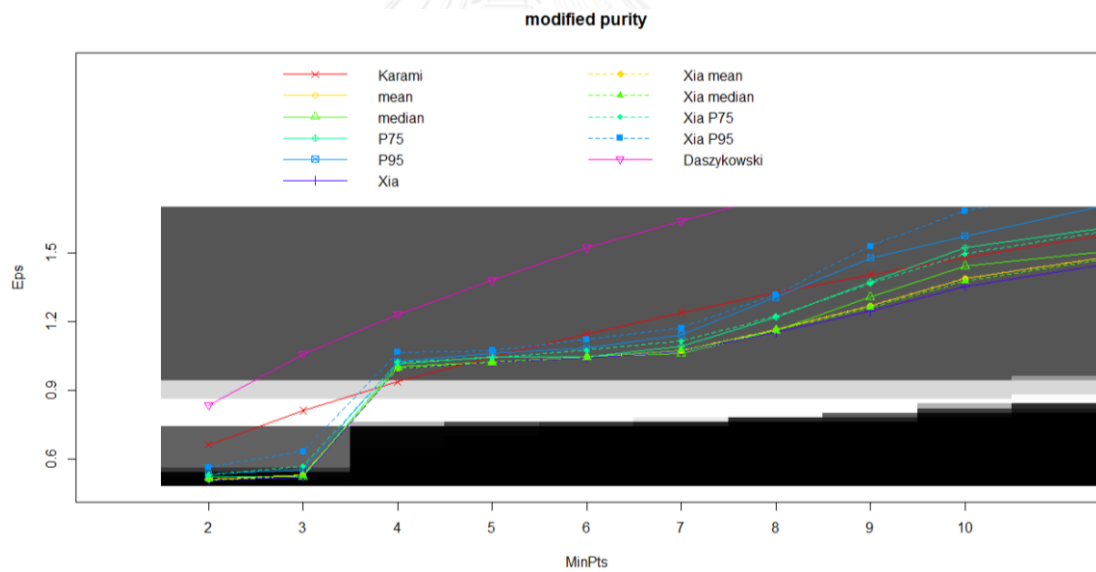
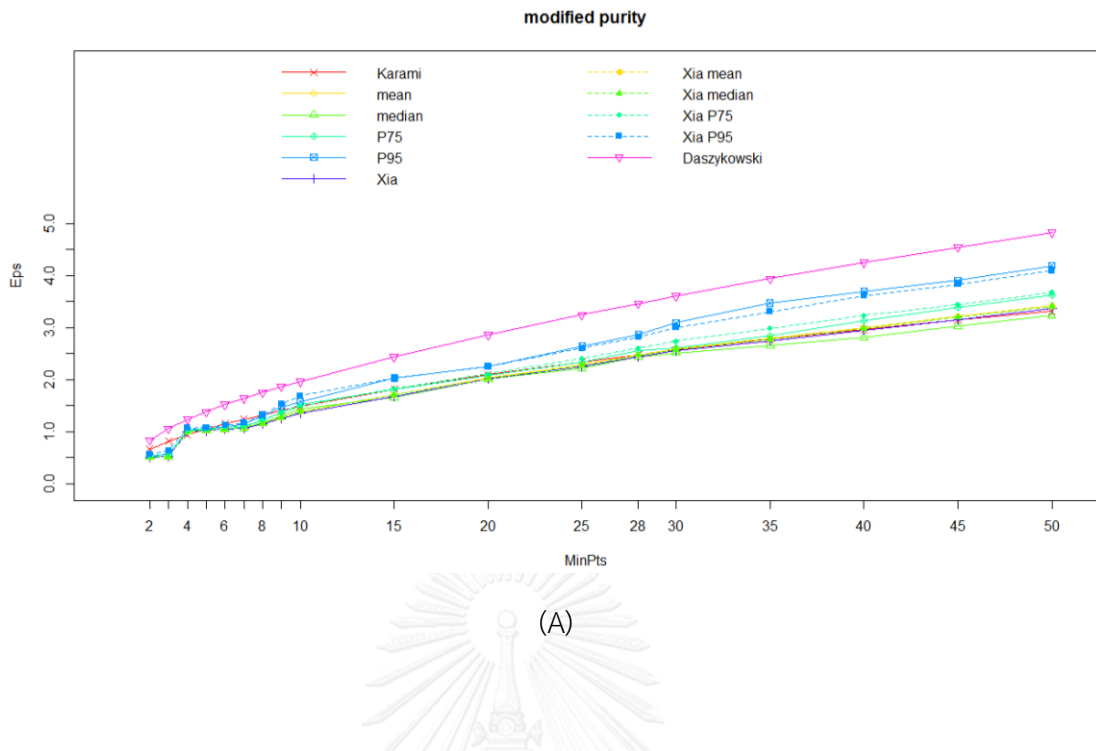
MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.3846	0.1624	0.2000	0.3504	0.3675	0.1624	0.1624	0.1624	0.3846	0.3846	0.3846
3	0.3846	0.1624	0.2103	0.3504	0.3675	0.1624	0.1624	0.1624	0.3846	0.3846	1.0000
4	0.2359	0.0906	0.0855	0.2393	1.0000	0.0821	0.0906	0.0906	0.2359	1.0000	1.0000
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5128
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5128
7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5128
8	1.0000	1.0000	0.6017	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5128
9	0.5128	1.0000	1.0000	0.5128	0.5128	1.0000	1.0000	1.0000	1.0000	0.5128	0.5128
10	0.5128	1.0000	0.5128	0.5128	0.5128	1.0000	1.0000	1.0000	0.5128	0.5128	0.5128
15	0.5128	0.5128	0.8462	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128
20	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128
23	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128
25	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128
30	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128
35	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128
40	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128
45	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128
50	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128	0.5128



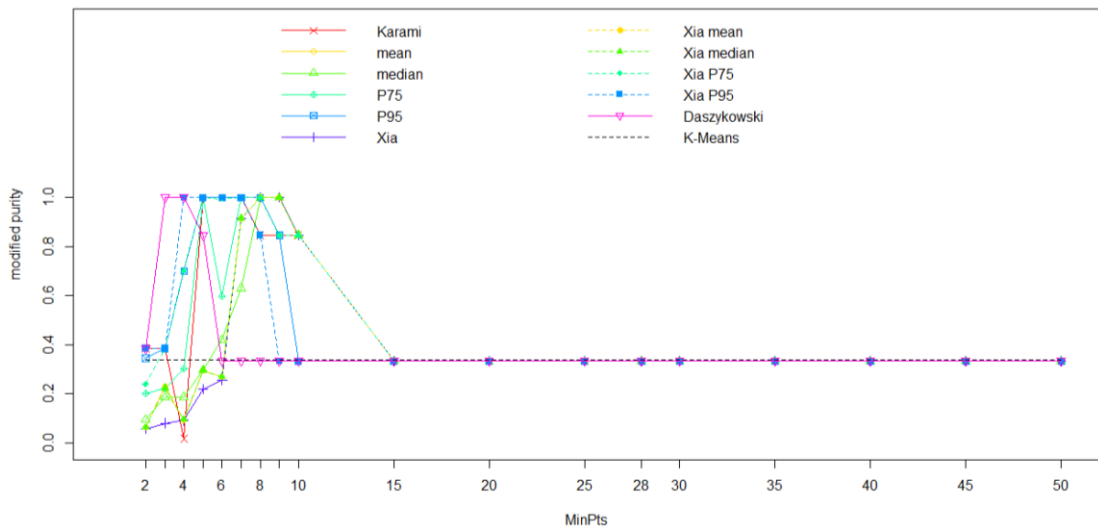
4.3 ข้อมูลวงแหวนรูปแบบที่ 2



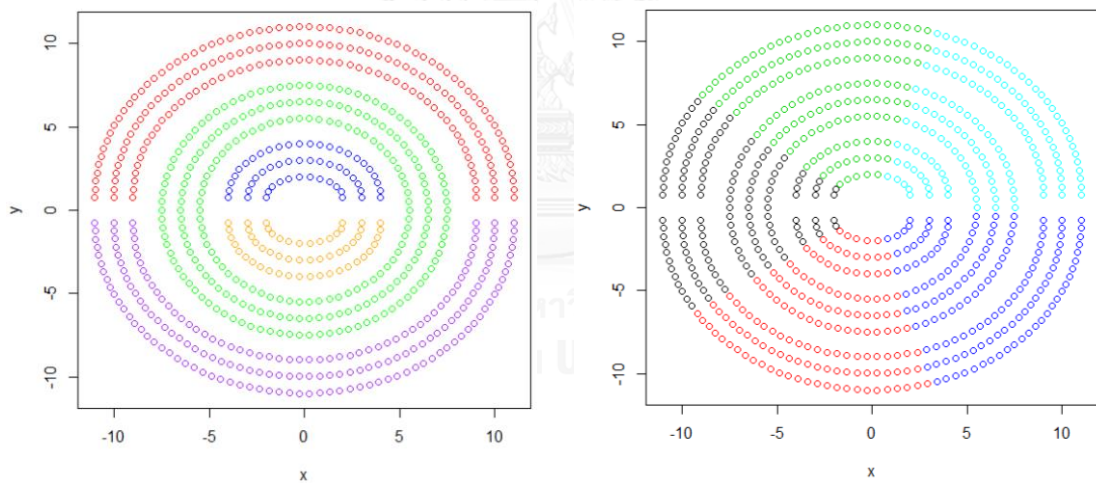
รูปที่ 4.3.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.3



รูปที่ 4.3.2 (A) กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.3 และ (B) กราฟขยายส่วนต้นของรูป (A) โดยแสดงค่า modified purity ด้วยความสว่างของ พื้นหลังของข้อมูลกรณี 4.3



รูปที่ 4.3.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณี 4.3



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.3.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.3

สำหรับข้อมูลวงแหวนรูปแบบที่ 2 พบว่าจากรูปที่ 4.3.2 (A) และ (B) ซึ่งรูป 4.3.2 (B) มีสีพื้นหลังเป็นค่า modified purity ไล่ตามระดับสีจาก 0 ถึง 1 (ดำไปถึงขาว) ถ้าเป็นสีเข้มแสดงว่าการใช้ค่าพารามิเตอร์นั้น ๆ ในการทำ DBSCAN ไม่เหมาะสม แต่ถ้าเป็นสีขาวแสดงว่าการใช้พารามิเตอร์นั้น ๆ ในการทำ DBSCAN เหมาะสม เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุก

กรณี ซึ่งมีระดับค่า Eps ดังนี้ วิธี Daszykowski สูงกว่าวิธีที่เหลือ และทุกวิธียกเว้นวิธี Daszykowski ให้ค่า Eps ในช่วงค่า MinPts น้อย ๆ ใกล้เคียงกัน

จากรูปที่ 4.3.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าในช่วง MinPts 3-4 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 ซึ่งสูงกว่ากลุ่มอื่น แต่เมื่อค่า MinPts สูงกว่านั้นแล้ว ค่า Eps จะสูงเกินไป ทำให้ค่า modified purity ค่อย ๆ ลดลง จึงทำให้พารามิเตอร์นั้น ๆ แบ่งกลุ่มข้อมูลไม่เหมาะสมเมื่อ MinPts มีค่ามากกว่า 5 ค่า modified purity ลู่เข้าสู่ 0.333 (234/702) ซึ่งเป็นกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 พบว่าช่วง MinPts 5-8 ค่า modified purity มีค่าเป็น 1 จากนั้นลดลงจนกระทั่งที่ MinPts 10 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (234/702) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Xia P95 พบว่าช่วง MinPts 4-7 ค่า modified purity มีค่าเป็น 1 จากนั้นลดลงจนกระทั่งที่ MinPts 9 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (234/702) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าในช่วง MinPts 2-4 ค่า Eps มีค่าต่ำเกินไป จากนั้นในช่วง MinPts 5-7 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 แต่เมื่อค่า MinPts ค่อย ๆ เพิ่มขึ้น ค่า Eps ก็จะมีค่าสูงเกินไป ทำให้ค่า modified purity ค่อย ๆ ลดลง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (234/702) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 พบว่าที่ค่า MinPts 5, 7-8 ค่า modified purity มีค่าเป็น 1 จากนั้นลดลงจนกระทั่งที่ MinPts 10 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (234/702) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Xia P75 พบว่าช่วง MinPts 5, 7-8 ค่า modified purity มีค่าเป็น 1 จากนั้นลดลงจนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (234/702) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก และที่ MinPts 8-9 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (234/702) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง และที่ MinPts 8-9 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าเป็น 1 จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (234/702) ด้วยเหตุผลเดียวกับวิธี Daszykowski

จากรูปที่ 4.3.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มแตกต่างกับแผนภาพกระจายของข้อมูลต้นแบบ กล่าวคือจากรูปต้นแบบเมื่อวิเคราะห์กลุ่มด้วย K-means พบว่าแบ่งย่อยออกเป็น 5 ส่วนคล้ายใบพัด ทั้งที่ความเป็นจริงควรแบ่งชั้น ๆ ที่อยู่ใกล้กันเป็นกลุ่มเดียวกัน ซึ่งเมื่อพิจารณาจากรูปที่ 4.3.3 พบว่าค่า modified purity ของ K-means เป็น 0.3390313

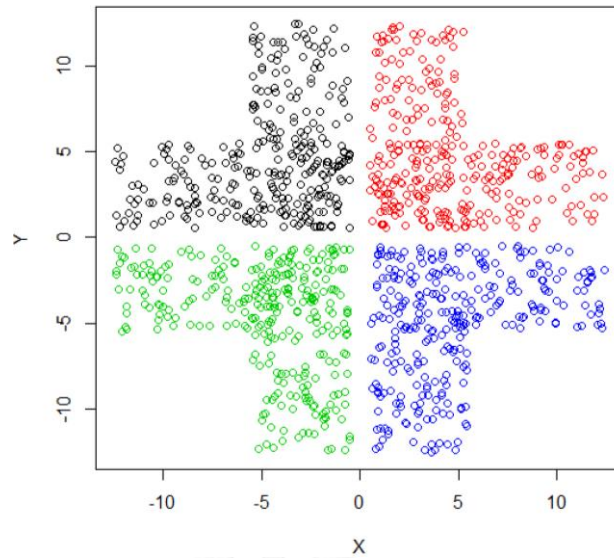
ตารางที่ 4.3.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.3

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.6624	0.5098	0.5234	0.5235	0.5321	0.5070	0.5098	0.5089	0.5304	0.5630	0.8346
3	0.8113	0.5286	0.5234	0.5241	0.5534	0.5186	0.5286	0.5253	0.5668	0.6322	1.0585
4	0.9368	0.9967	1.0028	1.0134	1.0250	0.9942	0.9967	0.9967	1.0246	1.0661	1.2322
5	1.0474	1.0228	1.0237	1.0456	1.0627	1.0215	1.0228	1.0228	1.0430	1.0729	1.3809
6	1.1473	1.0445	1.0456	1.0463	1.0837	1.0416	1.0445	1.0445	1.0753	1.1210	1.5228
7	1.2393	1.0746	1.0596	1.0915	1.1429	1.0701	1.0746	1.0746	1.1136	1.1722	1.6384
8	1.3248	1.1651	1.1623	1.2202	1.3057	1.1547	1.1651	1.1616	1.2236	1.3184	1.7523
9	1.4052	1.2700	1.3082	1.3732	1.4782	1.2440	1.2700	1.2613	1.3651	1.5293	1.8638
10	1.4812	1.3864	1.4431	1.5213	1.5746	1.3551	1.3864	1.3759	1.4950	1.6842	1.9617
15	1.8141	1.7105	1.6569	1.8165	2.0306	1.6807	1.7105	1.7006	1.8292	2.0310	2.4365
20	2.0948	2.0257	2.0066	2.0750	2.2556	2.0133	2.0257	2.0216	2.1106	2.2456	2.8623
25	2.3420	2.2819	2.2187	2.3312	2.6416	2.2591	2.2819	2.2743	2.4034	2.6019	3.2462
28	2.4786	2.4698	2.4349	2.5555	2.8675	2.4449	2.4698	2.4615	2.6017	2.8172	3.4581
30	2.5656	2.5841	2.5130	2.6135	3.0952	2.5499	2.5841	2.5727	2.7413	3.0031	3.6058
35	2.7711	2.7939	2.6571	2.8402	3.4724	2.7466	2.7939	2.7781	2.9850	3.3095	3.9395
40	2.9624	3.0044	2.8144	3.1308	3.6953	2.9432	3.0044	2.9839	3.2289	3.6162	4.2517
45	3.1421	3.2152	3.0252	3.3842	3.9116	3.1567	3.2152	3.1957	3.4428	3.8316	4.5435
50	3.3121	3.4278	3.2314	3.6171	4.1908	3.3625	3.4278	3.4060	3.6758	4.1011	4.8245

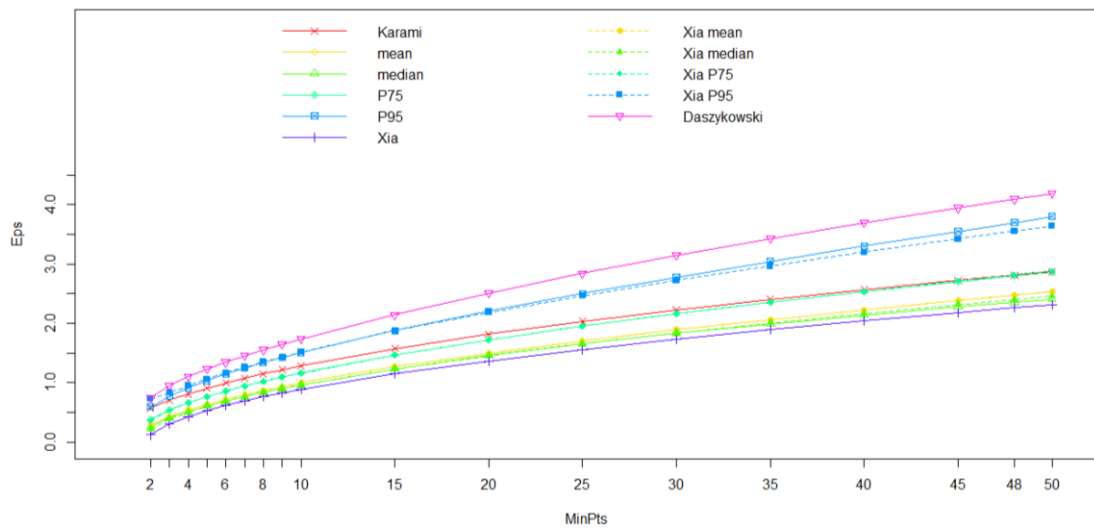
ตารางที่ 4.3.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.3

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.3846	0.0655	0.0954	0.2023	0.3447	0.0570	0.0655	0.0627	0.2393	0.3846	0.3846
3	0.3846	0.2308	0.1852	0.2222	0.3846	0.0798	0.2308	0.2222	0.3846	0.3846	1.0000
4	0.0171	0.0940	0.1866	0.3034	0.7009	0.0912	0.0940	0.0940	0.7009	1.0000	1.0000
5	1.0000	0.2949	0.2977	1.0000	1.0000	0.2179	0.2949	0.2949	1.0000	1.0000	0.8462
6	1.0000	0.2678	0.4202	0.5983	1.0000	0.2564	0.2678	0.2678	0.9943	1.0000	0.3333
7	1.0000	0.9145	0.6296	1.0000	1.0000	0.9145	0.9145	0.9145	1.0000	1.0000	0.3333
8	0.8462	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.8462	0.3333
9	0.8462	1.0000	1.0000	0.8462	0.8462	1.0000	1.0000	1.0000	0.8462	0.3333	0.3333
10	0.8462	0.8462	0.8462	0.3333	0.3333	0.8462	0.8462	0.8462	0.8462	0.3333	0.3333
15	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
20	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
25	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
28	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
30	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
35	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
40	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
45	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
50	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333

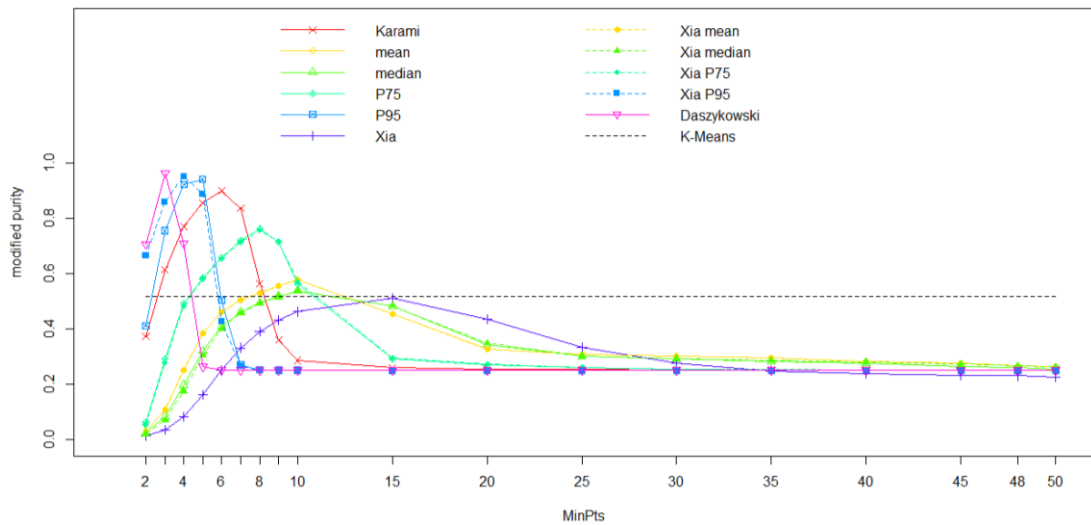
4.4 ข้อมูลมีลักษณะเป็นมุม



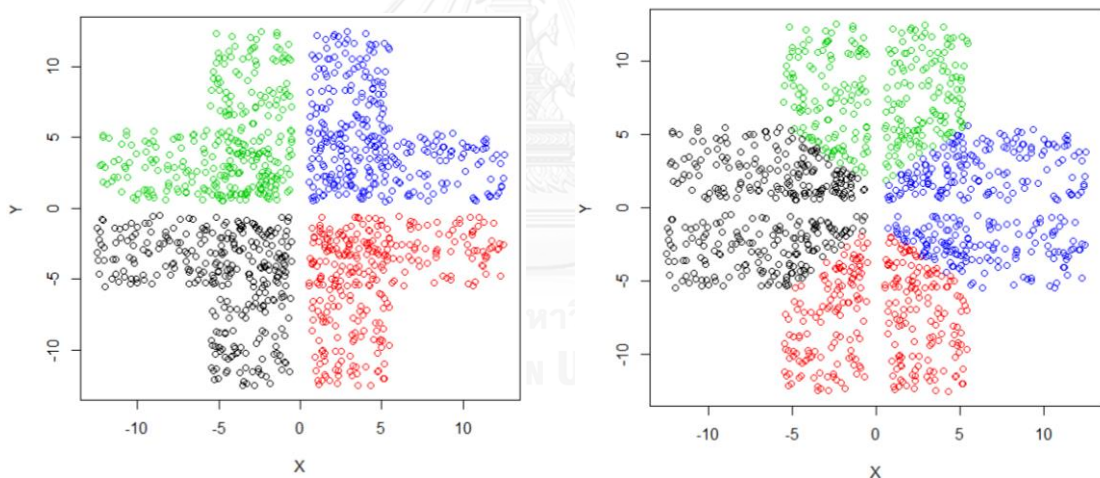
รูปที่ 4.4.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.4



รูปที่ 4.4.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.4



รูปที่ 4.4.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของ ข้อมูลกรณีที่ 4.4



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.4.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.4

สำหรับข้อมูลมีลักษณะเป็นมุม พบว่าจากรูปที่ 4.4.2 (A) และ (B) ซึ่งรูป 4.4.2 (B) มีสีพื้นหลังเป็นค่า modified purity ไต่ตามระดับสีจาก 0 ถึง 1 (ดำไปถึงขาว) ถ้าเป็นสีเข้มแสดงว่าการใช้ค่าพารามิเตอร์นั้น ๆ ในการทำ DBSCAN ไม่เหมาะสม แต่ถ้าเป็นสีขาวแสดงว่าการใช้พารามิเตอร์นั้น ๆ ในการทำ DBSCAN เหมาะสม เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมี

ระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski กลุ่มที่ประกอบด้วยวิธี P95 และ วิธี Xia P95 กลุ่มที่ประกอบด้วยวิธี Karami วิธี P75 และวิธี Xia P75 กลุ่มที่ประกอบด้วยวิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.4.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าในช่วง MinPts 2-4 ค่า Eps อยู่ในระดับที่ค่า modified purity มีค่าอยู่ระหว่าง 0.706–0.964 ซึ่งสูงกว่ากลุ่มอื่น โดยเฉพาะที่ MinPts 3 ให้ค่า modified purity เป็น 0.964 ซึ่งเป็นค่าที่แบ่งกลุ่มข้อมูลได้ดีที่สุด แต่เมื่อค่า MinPts สูงกว่านั้นแล้ว ค่า Eps จะสูงเกินไป ทำให้ค่า modified purity ค่อย ๆ ลดลง จึงทำให้พารามิเตอร์นั้น ๆ แบ่งกลุ่มข้อมูลไม่เหมาะสมเมื่อ MinPts มีค่ามากกว่า 5 ค่า modified purity ลู่เข้าสู่ 0.25 (150/600) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 4-5 ค่า modified purity มีค่าเป็นสูง จากนั้นลดลง จนกระทั่งที่ MinPts 8 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.25 (150/600) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 จนกระทั่งสูงสุดที่ MinPts 6 จากนั้น modified purity มีค่าลดลง จนกระทั่งที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.25 (150/600) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และวิธี Xia P75 ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 จนกระทั่งสูงสุดที่ MinPts 8 จากนั้น modified purity มีค่าลดลง จนกระทั่งที่ MinPts 25 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.25 (150/600)

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-6 แล้วค่อย ๆ เพิ่มสูงขึ้น จนกระทั่งสูงสุดที่ MinPts 10 จากนั้น modified purity มีค่าลดลง จนกระทั่งที่ MinPts 48 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.25 (150/600)

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง จนกระทั่งสูงสุดที่ MinPts 15 จากนั้น modified purity มีค่าลดลง

จากรูปที่ 4.4.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มแตกต่างกับแผนภาพกระจายของข้อมูลต้นแบบ กล่าวคือจากรูปต้นแบบเมื่อวิเคราะห์กลุ่มด้วย K-means พบว่าแบ่งย่อยออกเป็น 4 ส่วน ทั้งที่ความเป็นจริงควรแบ่งส่วนมุมเป็นกลุ่มเดียวกัน ซึ่งเมื่อพิจารณาจากรูปที่ 4.4.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.5165433

ตารางที่ 4.4.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีที่ 4.4

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.5739	0.2858	0.2591	0.3777	0.5928	0.1227	0.2858	0.2182	0.3606	0.7304	0.7473
3	0.7029	0.4329	0.4074	0.5387	0.7719	0.3079	0.4329	0.3886	0.5327	0.8319	0.9490
4	0.8117	0.5452	0.5181	0.6614	0.9129	0.4335	0.5452	0.5066	0.6568	0.9494	1.1016
5	0.9075	0.6403	0.6111	0.7648	1.0360	0.5335	0.6403	0.6037	0.7615	1.0593	1.2312
6	0.9941	0.7246	0.6935	0.8576	1.1461	0.6190	0.7246	0.6886	0.8546	1.1621	1.3468
7	1.0738	0.8012	0.7681	0.9430	1.2471	0.6954	0.8012	0.7652	0.9393	1.2578	1.4530
8	1.1479	0.8719	0.8375	1.0222	1.3402	0.7647	0.8719	0.8355	1.0177	1.3482	1.5518
9	1.2175	0.9380	0.9024	1.0958	1.4264	0.8293	0.9380	0.9011	1.0911	1.4332	1.6457
10	1.2834	1.0001	0.9641	1.1652	1.5097	0.8896	1.0001	0.9626	1.1601	1.5140	1.7354
15	1.5718	1.2718	1.2338	1.4676	1.8779	1.1500	1.2718	1.2307	1.4634	1.8740	2.1431
20	1.8150	1.5014	1.4592	1.7227	2.2001	1.3674	1.5014	1.4562	1.7207	2.1842	2.5086
25	2.0292	1.7050	1.6554	1.9503	2.4980	1.5587	1.7050	1.6557	1.9499	2.4635	2.8404
30	2.2229	1.8904	1.8306	2.1577	2.7762	1.7316	1.8904	1.8369	2.1595	2.7213	3.1455
35	2.4010	2.0626	1.9901	2.3495	3.0421	1.8910	2.0626	2.0048	2.3549	2.9642	3.4287
40	2.5668	2.2247	2.1381	2.5293	3.2990	2.0400	2.2247	2.1624	2.5396	3.1959	3.6944
45	2.7225	2.3789	2.2766	2.6986	3.5486	2.1806	2.3789	2.3120	2.7163	3.4199	3.9450
48	2.8117	2.4683	2.3546	2.7962	3.6975	2.2613	2.4683	2.3985	2.8193	3.5520	4.0898
50	2.8697	2.5269	2.4056	2.8599	3.7974	2.3140	2.5269	2.4551	2.8868	3.6390	4.1842

ตารางที่ 4.4.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.4

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.3727	0.0293	0.0244	0.0622	0.4111	0.0106	0.0293	0.0189	0.0531	0.6654	0.7063
3	0.6138	0.1058	0.0820	0.2911	0.7562	0.0349	0.1058	0.0679	0.2785	0.8606	0.9639
4	0.7714	0.2516	0.1964	0.4906	0.9239	0.0830	0.2516	0.1749	0.4826	0.9518	0.7097
5	0.8570	0.3845	0.3198	0.5856	0.9419	0.1601	0.3845	0.3036	0.5812	0.8889	0.2624
6	0.9001	0.4610	0.4096	0.6589	0.5023	0.2500	0.4610	0.3996	0.6542	0.4271	0.2502
7	0.8362	0.5034	0.4611	0.7201	0.2709	0.3312	0.5034	0.4559	0.7131	0.2654	0.2500
8	0.5639	0.5316	0.4942	0.7631	0.2512	0.3907	0.5316	0.4916	0.7574	0.2506	0.2500
9	0.3591	0.5550	0.5184	0.7155	0.2503	0.4324	0.5550	0.5170	0.7193	0.2503	0.2500
10	0.2852	0.5765	0.5381	0.5608	0.2500	0.4619	0.5765	0.5365	0.5708	0.2501	0.2500
15	0.2590	0.4536	0.4820	0.2916	0.2500	0.5098	0.4536	0.4836	0.2944	0.2500	0.2500
20	0.2544	0.3268	0.3436	0.2699	0.2500	0.4345	0.3268	0.3480	0.2721	0.2500	0.2500
25	0.2525	0.3079	0.3011	0.2591	0.2500	0.3336	0.3079	0.3012	0.2603	0.2500	0.2500
30	0.2517	0.3014	0.2906	0.2539	0.2500	0.2764	0.3014	0.2914	0.2549	0.2500	0.2500
35	0.2507	0.2937	0.2824	0.2515	0.2500	0.2485	0.2937	0.2849	0.2520	0.2500	0.2500
40	0.2505	0.2834	0.2742	0.2507	0.2500	0.2388	0.2834	0.2787	0.2506	0.2500	0.2500
45	0.2502	0.2744	0.2646	0.2502	0.2500	0.2327	0.2744	0.2720	0.2502	0.2500	0.2500
48	0.2501	0.2676	0.2591	0.2501	0.2500	0.2297	0.2676	0.2659	0.2501	0.2500	0.2500
50	0.2501	0.2646	0.2544	0.2501	0.2500	0.2258	0.2646	0.2627	0.2501	0.2500	0.2500

4.5 ข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร

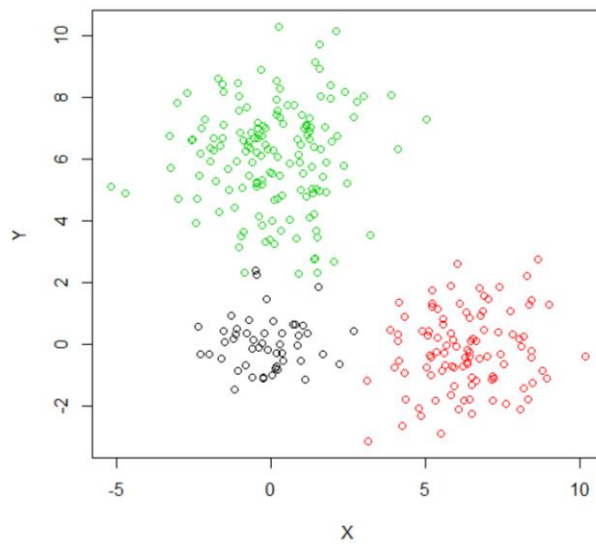
ข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร ทำซ้ำ 500 รอบ

4.5.1 กรณีจำนวนกลุ่มเท่ากับ 3

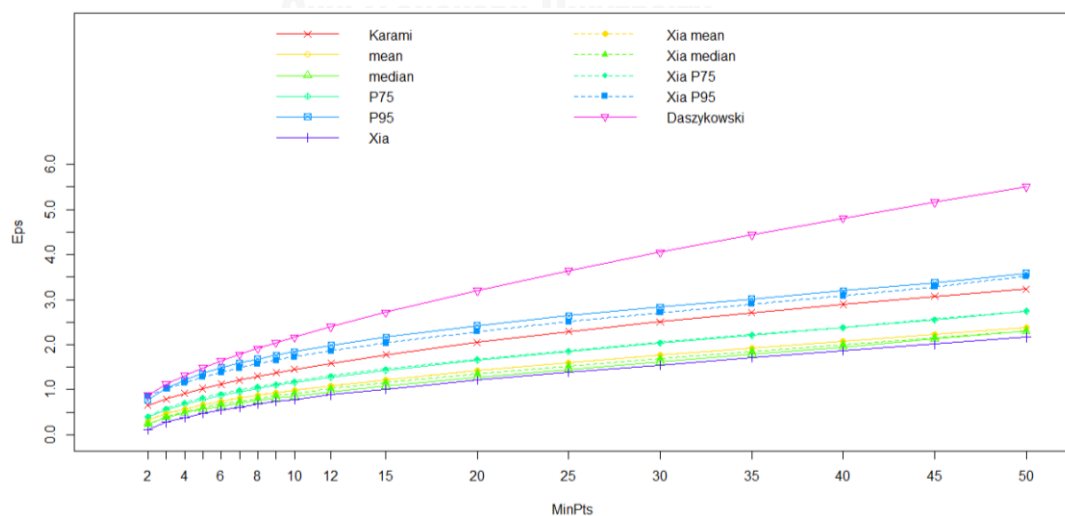
4.5.1.1 ขนาดตัวอย่าง $N=300$ กลุ่มละ 50, 100, 150 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$, $\mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

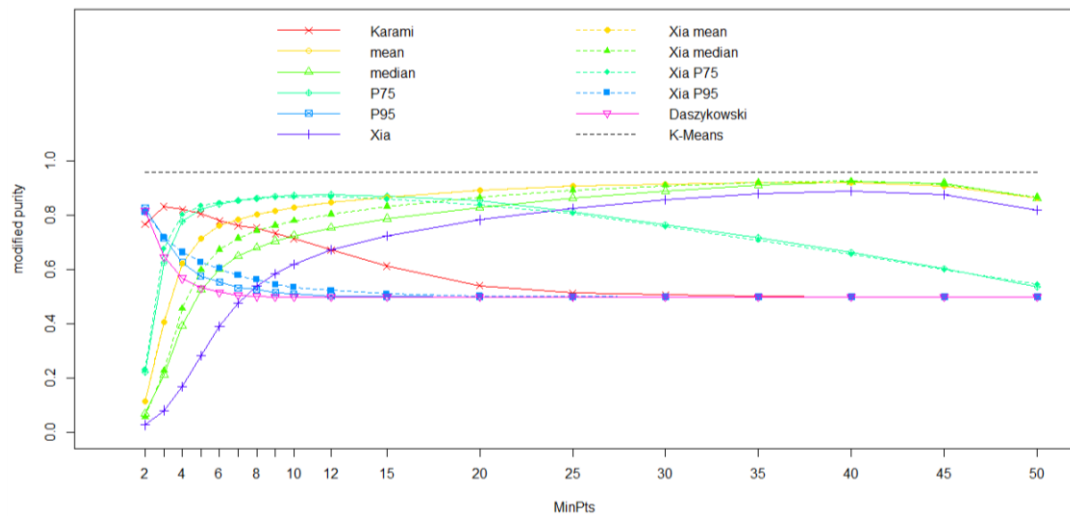
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



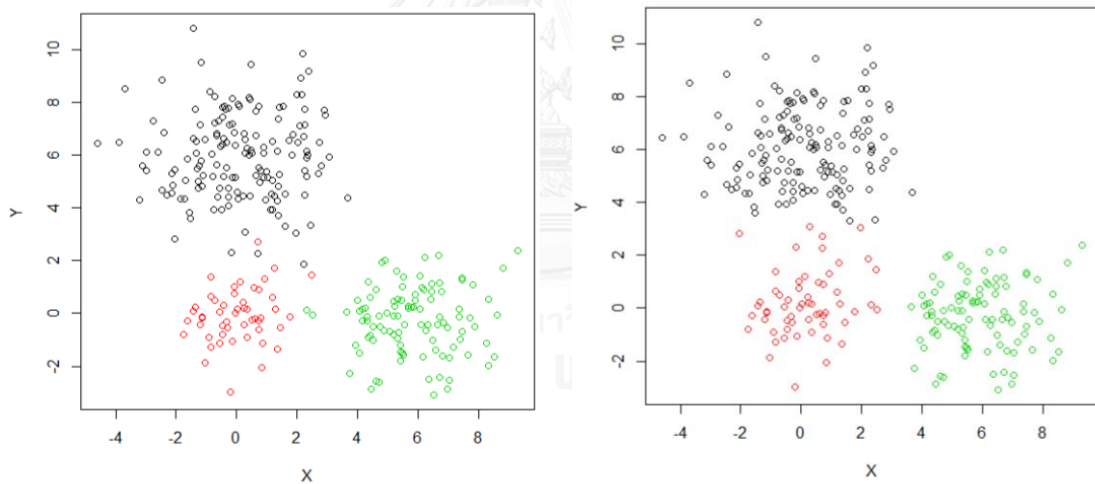
รูปที่ 4.5.1.1.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.1.1



รูปที่ 4.5.1.1.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.1



รูปที่ 4.5.1.1.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณี 4.5.1.1



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.1.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.5.1.1

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.1.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี

Xia ตามลำดับ ซึ่งวิธี Daszykowski ตั้งแต่ MinPts 20 เป็นต้นไปให้ค่า Eps สูงกว่าวิธีอื่น ๆ เป็นอย่างมาก

จากรูปที่ 4.5.1.1.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.5 (150/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (150/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 3 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 30 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (150/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 12 จากนั้น modified purity มีค่าลดลงและมีแนวโน้มที่จะลู่เข้าสู่ 0.5

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-6 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 10-45 จากนั้นลดลงเล็กน้อย

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.1.4 การวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.1.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9610267

ตารางที่ 4.5.1.1.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.1

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.6471	0.3143	0.2516	0.3894	0.7698	0.1125	0.3143	0.2245	0.3934	0.8680	0.8788
3	0.7925	0.4639	0.3858	0.5546	1.0303	0.2721	0.4639	0.3921	0.5820	1.0144	1.1276
4	0.9151	0.5732	0.4825	0.6764	1.2123	0.3818	0.5732	0.5041	0.7112	1.1560	1.3195
5	1.0231	0.6640	0.5641	0.7806	1.3590	0.4705	0.6640	0.5952	0.8175	1.2801	1.4863
6	1.1208	0.7422	0.6334	0.8711	1.4846	0.5456	0.7422	0.6729	0.9083	1.3887	1.6378
7	1.2106	0.8117	0.6966	0.9516	1.5882	0.6129	0.8117	0.7421	0.9886	1.4842	1.7788
8	1.2942	0.8753	0.7542	1.0250	1.6815	0.6742	0.8753	0.8052	1.0619	1.5718	1.9130
9	1.3727	0.9340	0.8082	1.0935	1.7665	0.7310	0.9340	0.8635	1.1290	1.6516	2.0403
10	1.4469	0.9885	0.8580	1.1563	1.8436	0.7839	0.9885	0.9176	1.1913	1.7254	2.1632
12	1.5850	1.0893	0.9533	1.2729	1.9872	0.8824	1.0893	1.0180	1.3058	1.8598	2.3944
15	1.7721	1.2237	1.0817	1.4275	2.1676	1.0142	1.2237	1.1518	1.4573	2.0363	2.7162
20	2.0463	1.4214	1.2753	1.6536	2.4213	1.2099	1.4214	1.3491	1.6781	2.2895	3.2009
25	2.2878	1.5973	1.4530	1.8517	2.6387	1.3850	1.5973	1.5250	1.8729	2.5101	3.6413
30	2.5061	1.7605	1.6209	2.0340	2.8345	1.5486	1.7605	1.6886	2.0521	2.7102	4.0512
35	2.7069	1.9160	1.7840	2.2064	3.0151	1.7048	1.9160	1.8445	2.2221	2.8987	4.4391
40	2.8938	2.0691	1.9498	2.3750	3.1934	1.8572	2.0691	1.9974	2.3893	3.0860	4.8100
45	3.0694	2.2248	2.1228	2.5469	3.3732	2.0091	2.2248	2.1519	2.5612	3.2845	5.1654
50	3.2354	2.3894	2.2997	2.7433	3.5810	2.1616	2.3894	2.3124	2.7481	3.5169	5.5078

ตารางที่ 4.5.1.1.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

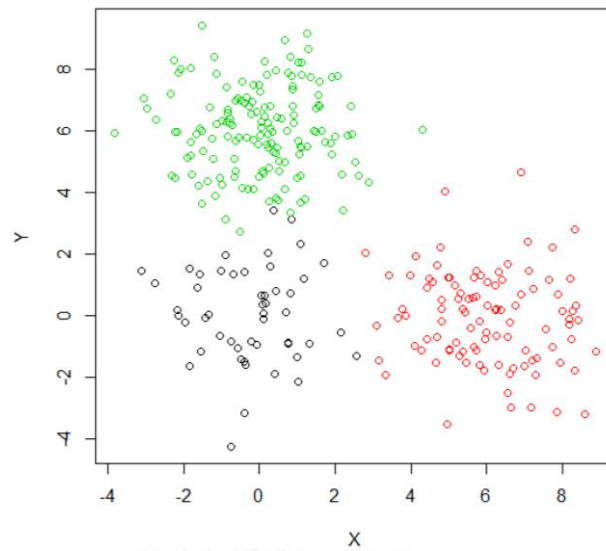
4.5.1.1

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7682	0.1141	0.0684	0.2212	0.8271	0.0262	0.1141	0.0556	0.2311	0.8108	0.8168
3	0.8321	0.4053	0.2123	0.6269	0.7165	0.0800	0.4053	0.2260	0.6793	0.7206	0.6469
4	0.8221	0.6231	0.3928	0.7770	0.6267	0.1677	0.6231	0.4549	0.8047	0.6638	0.5687
5	0.8059	0.7163	0.5252	0.8237	0.5762	0.2820	0.7163	0.5985	0.8368	0.6273	0.5319
6	0.7800	0.7612	0.6017	0.8418	0.5551	0.3903	0.7612	0.6731	0.8474	0.6045	0.5160
7	0.7622	0.7849	0.6499	0.8553	0.5324	0.4766	0.7849	0.7149	0.8559	0.5786	0.5062
8	0.7541	0.8023	0.6808	0.8629	0.5262	0.5406	0.8023	0.7431	0.8612	0.5631	0.5013
9	0.7340	0.8165	0.7049	0.8702	0.5152	0.5836	0.8165	0.7631	0.8666	0.5457	0.4994
10	0.7146	0.8276	0.7254	0.8746	0.5106	0.6184	0.8276	0.7795	0.8684	0.5349	0.4992
12	0.6728	0.8471	0.7543	0.8779	0.5023	0.6724	0.8471	0.8042	0.8696	0.5233	0.4991
15	0.6127	0.8668	0.7876	0.8714	0.5005	0.7234	0.8668	0.8333	0.8613	0.5097	0.4996
20	0.5411	0.8915	0.8300	0.8535	0.4993	0.7831	0.8915	0.8668	0.8398	0.5032	0.4999
25	0.5137	0.9078	0.8640	0.8123	0.4991	0.8245	0.9078	0.8928	0.8085	0.5000	0.5000
30	0.5072	0.9164	0.8905	0.7642	0.4994	0.8566	0.9164	0.9102	0.7587	0.4991	0.5000
35	0.5001	0.9220	0.9116	0.7177	0.4996	0.8790	0.9219	0.9213	0.7098	0.4994	0.5000
40	0.4992	0.9223	0.9231	0.6629	0.4997	0.8904	0.9223	0.9271	0.6563	0.4996	0.5000
45	0.4995	0.9083	0.9184	0.6028	0.4998	0.8766	0.9083	0.9153	0.6011	0.4997	0.5000
50	0.4997	0.8646	0.8646	0.5353	0.4999	0.8204	0.8643	0.8677	0.5470	0.4999	0.5000

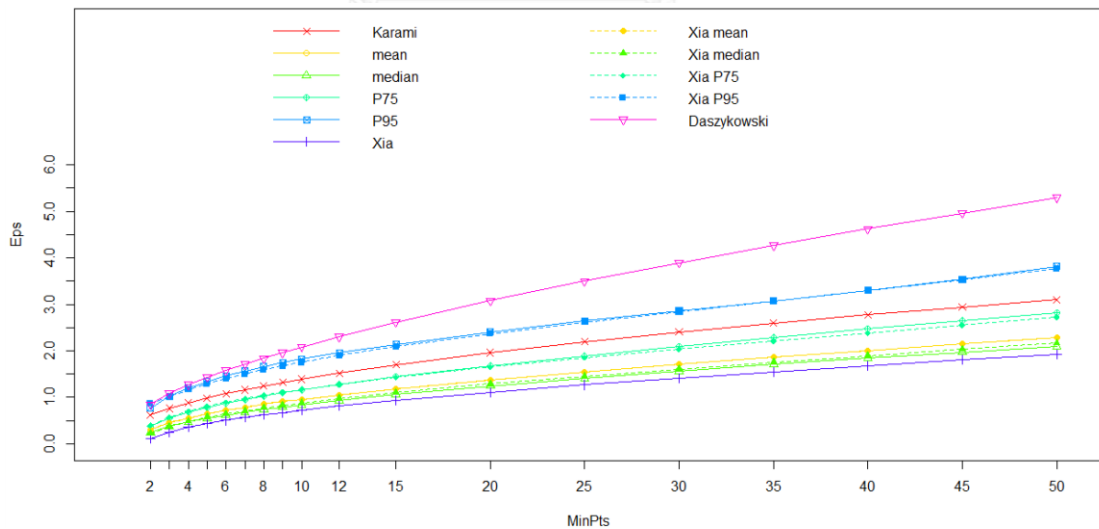
4.5.1.2 ขนาดตัวอย่าง N=300 กลุ่มละ 50, 100, 150 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

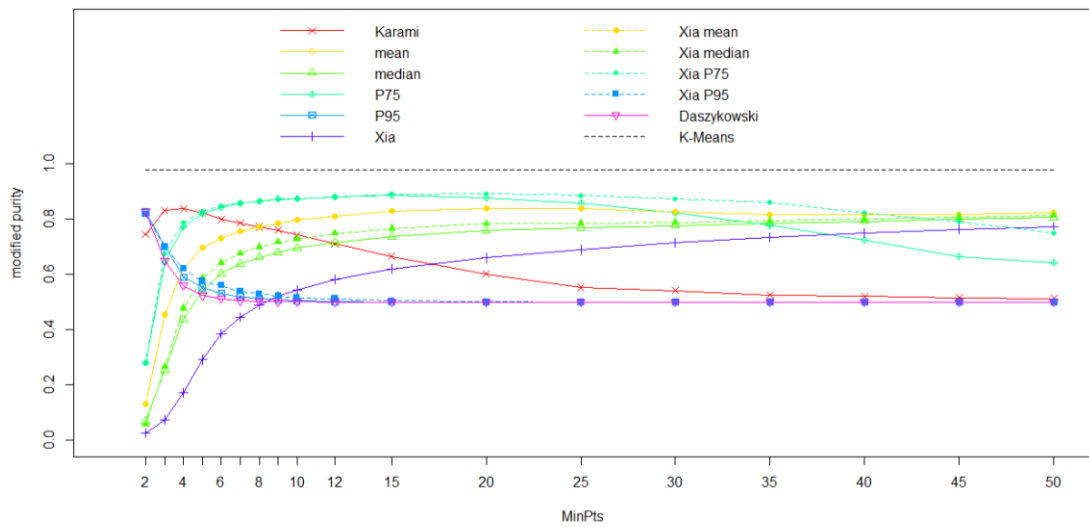
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



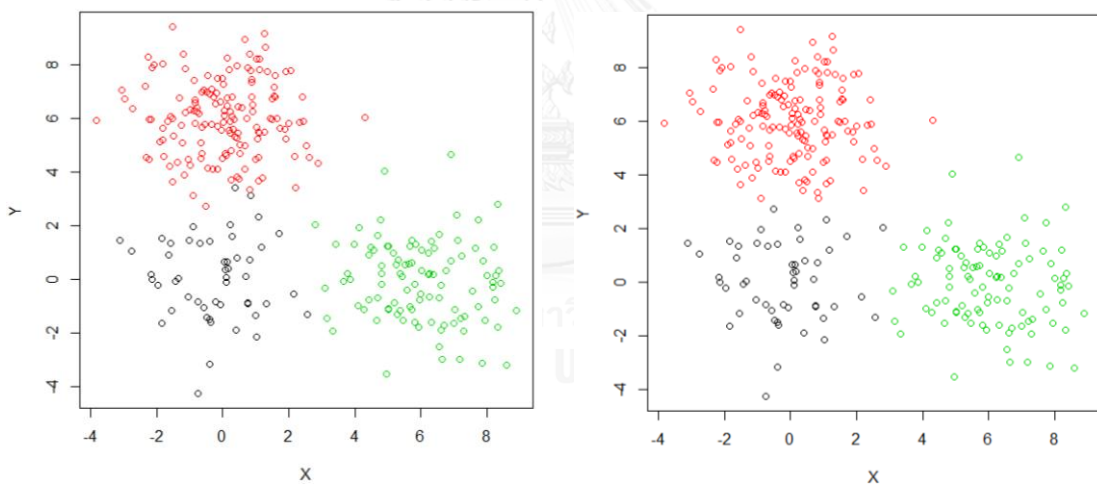
รูปที่ 4.5.1.2.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.1.2



รูปที่ 4.5.1.2.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.2



รูปที่ 4.5.1.2.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณี 4.5.1.2



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.2.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.5.1.2

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.2.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี P75, วิธี Xia P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.1.2.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.5 (150/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (150/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 3 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 40 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (150/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 15 จากนั้น modified purity มีค่าลดลง

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 10-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.2.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.2.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9774867

ตารางที่ 4.5.1.2.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.2

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.6201	0.3039	0.2384	0.3794	0.7605	0.1017	0.3039	0.2119	0.3791	0.8607	0.8438
3	0.7594	0.4494	0.3695	0.5468	1.0159	0.2493	0.4494	0.3732	0.5654	1.0140	1.0836
4	0.8769	0.5554	0.4635	0.6727	1.1967	0.3500	0.5554	0.4801	0.6930	1.1624	1.2683
5	0.9804	0.6419	0.5425	0.7772	1.3372	0.4311	0.6419	0.5659	0.7958	1.2876	1.4284
6	1.0740	0.7174	0.6116	0.8681	1.4563	0.5011	0.7174	0.6402	0.8849	1.3980	1.5746
7	1.1600	0.7848	0.6742	0.9496	1.5618	0.5629	0.7848	0.7062	0.9644	1.4984	1.7102
8	1.2401	0.8464	0.7310	1.0247	1.6574	0.6193	0.8464	0.7663	1.0368	1.5901	1.8389
9	1.3153	0.9033	0.7855	1.0937	1.7442	0.6718	0.9033	0.8220	1.1033	1.6737	1.9617
10	1.3865	0.9561	0.8354	1.1574	1.8204	0.7204	0.9561	0.8735	1.1650	1.7515	2.0793
12	1.5188	1.0535	0.9287	1.2772	1.9561	0.8104	1.0535	0.9687	1.2783	1.8940	2.3020
15	1.6981	1.1844	1.0573	1.4380	2.1411	0.9310	1.1844	1.0965	1.4304	2.0858	2.6114
20	1.9608	1.3758	1.2433	1.6750	2.4080	1.1079	1.3758	1.2833	1.6516	2.3634	3.0771
25	2.1922	1.5467	1.4097	1.8880	2.6403	1.2652	1.5467	1.4498	1.8490	2.6124	3.4996
30	2.4015	1.7054	1.5607	2.0889	2.8577	1.4100	1.7053	1.6039	2.0323	2.8455	3.8949
35	2.5939	1.8566	1.7038	2.2845	3.0730	1.5461	1.8566	1.7502	2.2078	3.0722	4.2673
40	2.7730	2.0037	1.8381	2.4720	3.2980	1.6762	2.0037	1.8915	2.3792	3.2976	4.6245
45	2.9412	2.1488	1.9677	2.6525	3.5413	1.8011	2.1488	2.0297	2.5499	3.5279	4.9656
50	3.1003	2.2936	2.0958	2.8122	3.8147	1.9226	2.2936	2.1665	2.7218	3.7656	5.2948

ตารางที่ 4.5.1.2.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

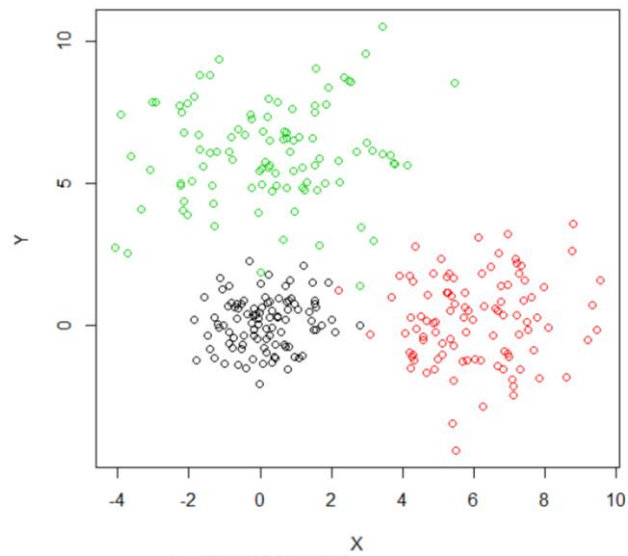
4.5.1.2

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7446	0.1291	0.0678	0.2798	0.8259	0.0243	0.1291	0.0551	0.2785	0.8169	0.8267
3	0.8316	0.4524	0.2530	0.6456	0.7009	0.0734	0.4524	0.2645	0.6759	0.7000	0.6492
4	0.8374	0.6234	0.4372	0.7716	0.5899	0.1713	0.6234	0.4759	0.7871	0.6199	0.5604
5	0.8214	0.6956	0.5397	0.8183	0.5526	0.2903	0.6956	0.5856	0.8271	0.5763	0.5237
6	0.7992	0.7321	0.6018	0.8424	0.5308	0.3831	0.7321	0.6423	0.8474	0.5592	0.5123
7	0.7857	0.7550	0.6373	0.8566	0.5182	0.4452	0.7550	0.6752	0.8569	0.5377	0.5048
8	0.7719	0.7734	0.6620	0.8654	0.5128	0.4881	0.7734	0.6990	0.8649	0.5295	0.5014
9	0.7605	0.7846	0.6791	0.8736	0.5081	0.5216	0.7846	0.7161	0.8691	0.5217	0.5005
10	0.7428	0.7961	0.6948	0.8743	0.5044	0.5438	0.7961	0.7282	0.8737	0.5158	0.5000
12	0.7104	0.8107	0.7151	0.8804	0.5015	0.5810	0.8107	0.7479	0.8798	0.5111	0.4997
15	0.6644	0.8298	0.7371	0.8868	0.4993	0.6186	0.8298	0.7660	0.8909	0.5044	0.4998
20	0.6016	0.8394	0.7584	0.8769	0.4996	0.6591	0.8394	0.7831	0.8917	0.5002	0.5000
25	0.5525	0.8374	0.7696	0.8589	0.4998	0.6892	0.8372	0.7851	0.8858	0.4998	0.5000
30	0.5407	0.8263	0.7764	0.8213	0.4999	0.7139	0.8263	0.7887	0.8728	0.4999	0.5000
35	0.5252	0.8169	0.7851	0.7784	0.5000	0.7334	0.8169	0.7931	0.8609	0.5000	0.5000
40	0.5194	0.8154	0.7922	0.7227	0.5000	0.7490	0.8154	0.7998	0.8235	0.5000	0.5000
45	0.5159	0.8177	0.7999	0.6649	0.5000	0.7623	0.8177	0.8073	0.7911	0.5000	0.5000
50	0.5114	0.8213	0.8065	0.6419	0.5000	0.7729	0.8213	0.8130	0.7499	0.5000	0.5000

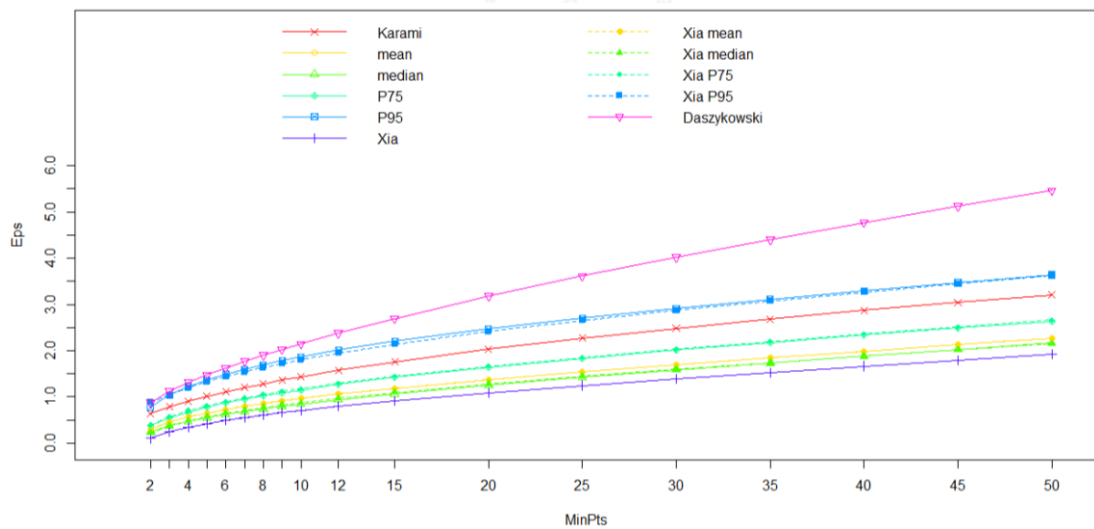
4.5.1.3 ขนาดตัวอย่าง N=300 กลุ่มละ 100, 100, 100 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$, $\mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

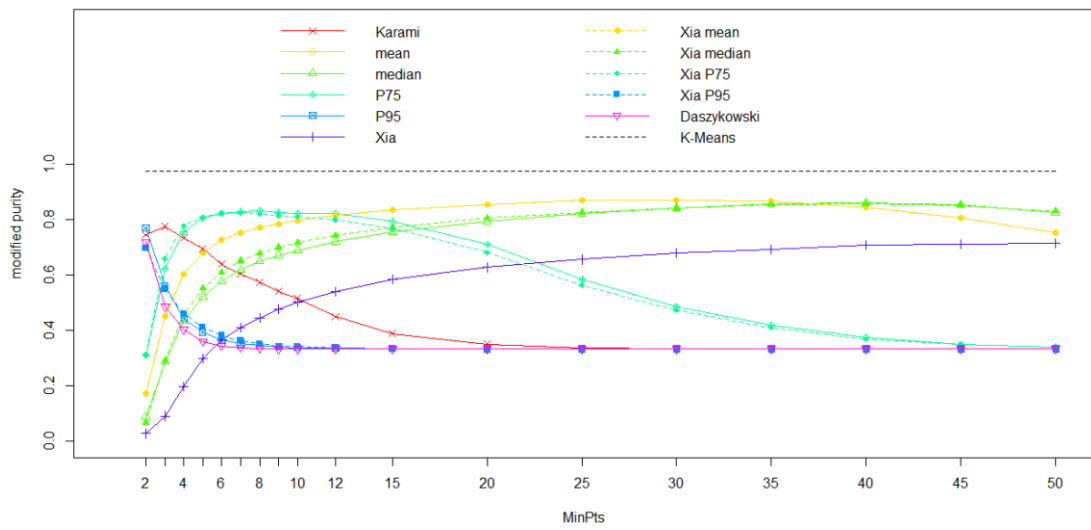
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



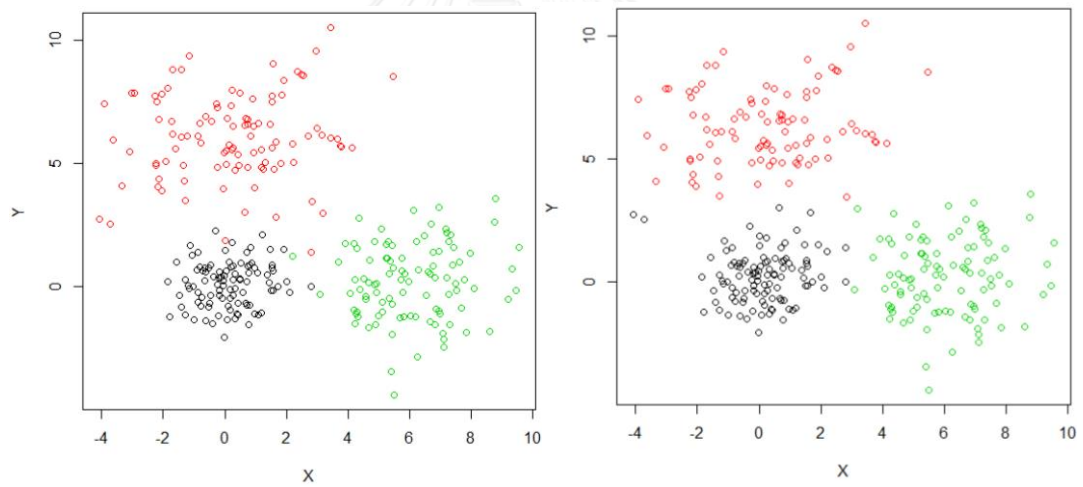
รูปที่ 4.5.1.3.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.3



รูปที่ 4.5.1.3.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.3



รูปที่ 4.5.1.3.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณี 4.5.1.3



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.3.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.5.1.3

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.3.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ ซึ่งวิธี Daszykowski ตั้งแต่ MinPts 20 เป็นต้นไปให้ค่า Eps สูงกว่าวิธีอื่น ๆ เป็นอย่างมาก

จากรูปที่ 4.5.1.3.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-5 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่อู่เข้าสู่ 0.333 (100/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-5 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 10 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.333 (100/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 3 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 25 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.333 (100/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 7 จากนั้น modified purity มีค่าลดลง และมีแนวโน้มที่จะลู่อู่เข้าสู่ 0.333

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-7 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 20-50 ยกเว้นวิธี mean และวิธี Xia mean ที่ MinPts ตั้งแต่ 35 เป็นต้นไปมีค่าลดลง

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.3.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.3.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9749267

ตารางที่ 4.5.1.3.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีที่ 4.5.1.3

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.6413	0.3071	0.2397	0.3827	0.7756	0.1000	0.3071	0.2120	0.3824	0.8783	0.8713
3	0.7854	0.4526	0.3708	0.5494	1.0357	0.2427	0.4526	0.3719	0.5702	1.0396	1.1181
4	0.9069	0.5589	0.4676	0.6718	1.2190	0.3420	0.5589	0.4787	0.6991	1.1913	1.3088
5	1.0140	0.6460	0.5482	0.7757	1.3601	0.4217	0.6460	0.5646	0.8034	1.3212	1.4743
6	1.1108	0.7215	0.6179	0.8637	1.4848	0.4892	0.7215	0.6380	0.8934	1.4370	1.6249
7	1.1998	0.7888	0.6809	0.9444	1.5910	0.5494	0.7888	0.7033	0.9734	1.5404	1.7642
8	1.2826	0.8500	0.7385	1.0181	1.6899	0.6041	0.8500	0.7627	1.0460	1.6347	1.8972
9	1.3604	0.9064	0.7926	1.0858	1.7797	0.6552	0.9064	0.8176	1.1125	1.7197	2.0238
10	1.4340	0.9591	0.8431	1.1483	1.8604	0.7030	0.9591	0.8688	1.1746	1.7994	2.1451
12	1.5709	1.0563	0.9364	1.2642	2.0104	0.7909	1.0563	0.9632	1.2886	1.9458	2.3753
15	1.7563	1.1861	1.0638	1.4175	2.2001	0.9103	1.1861	1.0899	1.4400	2.1371	2.6946
20	2.0280	1.3758	1.2535	1.6417	2.4700	1.0861	1.3758	1.2753	1.6598	2.4122	3.1758
25	2.2673	1.5441	1.4253	1.8356	2.6989	1.2434	1.5441	1.4402	1.8537	2.6527	3.6110
30	2.4838	1.6989	1.5842	2.0120	2.9083	1.3900	1.6989	1.5925	2.0307	2.8687	4.0178
35	2.6828	1.8449	1.7343	2.1766	3.1032	1.5288	1.8449	1.7364	2.1970	3.0703	4.4036
40	2.8680	1.9852	1.8800	2.3333	3.2869	1.6631	1.9852	1.8748	2.3562	3.2615	4.7716
45	3.0420	2.1223	2.0235	2.4855	3.4663	1.7951	2.1223	2.0104	2.5110	3.4460	5.1237
50	3.2065	2.2583	2.1651	2.6332	3.6406	1.9266	2.2583	2.1450	2.6638	3.6266	5.4629

ตารางที่ 4.5.1.3.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

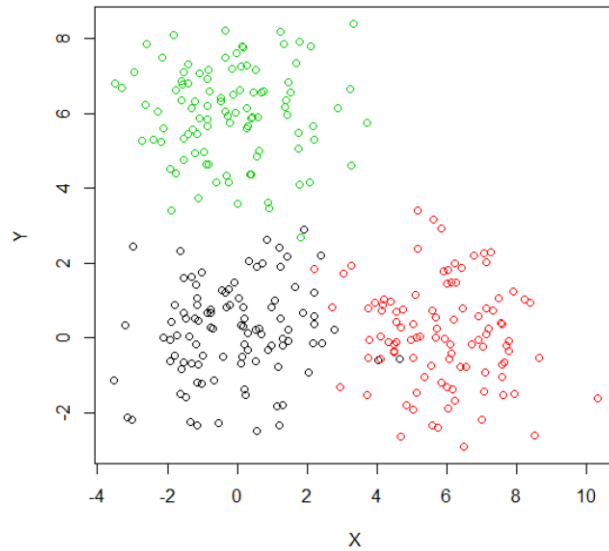
4.5.1.3

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7461	0.1694	0.0861	0.3096	0.7696	0.0264	0.1694	0.0664	0.3096	0.6975	0.7182
3	0.7750	0.4495	0.2866	0.6240	0.5609	0.0897	0.4495	0.2912	0.6597	0.5494	0.4850
4	0.7343	0.6026	0.4355	0.7574	0.4427	0.1975	0.6026	0.4584	0.7767	0.4599	0.4021
5	0.6943	0.6800	0.5215	0.8056	0.3927	0.2984	0.6800	0.5506	0.8091	0.4122	0.3590
6	0.6399	0.7266	0.5773	0.8234	0.3641	0.3644	0.7266	0.6090	0.8221	0.3805	0.3430
7	0.6039	0.7525	0.6188	0.8302	0.3521	0.4086	0.7525	0.6511	0.8257	0.3628	0.3366
8	0.5737	0.7708	0.6495	0.8334	0.3443	0.4444	0.7708	0.6788	0.8220	0.3507	0.3340
9	0.5402	0.7846	0.6710	0.8264	0.3405	0.4756	0.7846	0.6987	0.8145	0.3418	0.3333
10	0.5153	0.7966	0.6901	0.8231	0.3372	0.5008	0.7966	0.7168	0.8105	0.3417	0.3333
12	0.4494	0.8165	0.7207	0.8223	0.3346	0.5388	0.8165	0.7423	0.8003	0.3365	0.3333
15	0.3882	0.8347	0.7566	0.7941	0.3340	0.5840	0.8347	0.7726	0.7685	0.3340	0.3333
20	0.3491	0.8552	0.7942	0.7106	0.3333	0.6274	0.8551	0.8058	0.6824	0.3333	0.3333
25	0.3353	0.8690	0.8221	0.5843	0.3333	0.6560	0.8691	0.8259	0.5625	0.3333	0.3333
30	0.3333	0.8705	0.8421	0.4852	0.3333	0.6797	0.8704	0.8421	0.4734	0.3333	0.3333
35	0.3333	0.8674	0.8568	0.4185	0.3333	0.6930	0.8673	0.8531	0.4103	0.3333	0.3333
40	0.3333	0.8445	0.8609	0.3730	0.3333	0.7077	0.8446	0.8572	0.3685	0.3333	0.3333
45	0.3333	0.8081	0.8537	0.3476	0.3333	0.7122	0.8081	0.8520	0.3488	0.3333	0.3333
50	0.3333	0.7529	0.8261	0.3393	0.3333	0.7158	0.7529	0.8305	0.3404	0.3333	0.3333

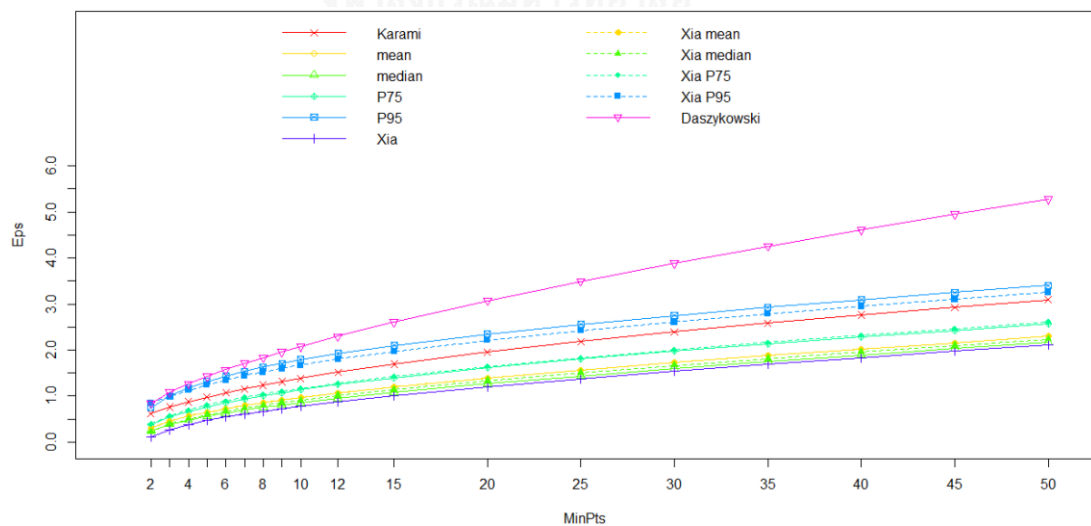
4.5.1.4 ขนาดตัวอย่าง N=300 กลุ่มละ 100, 100, 100 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

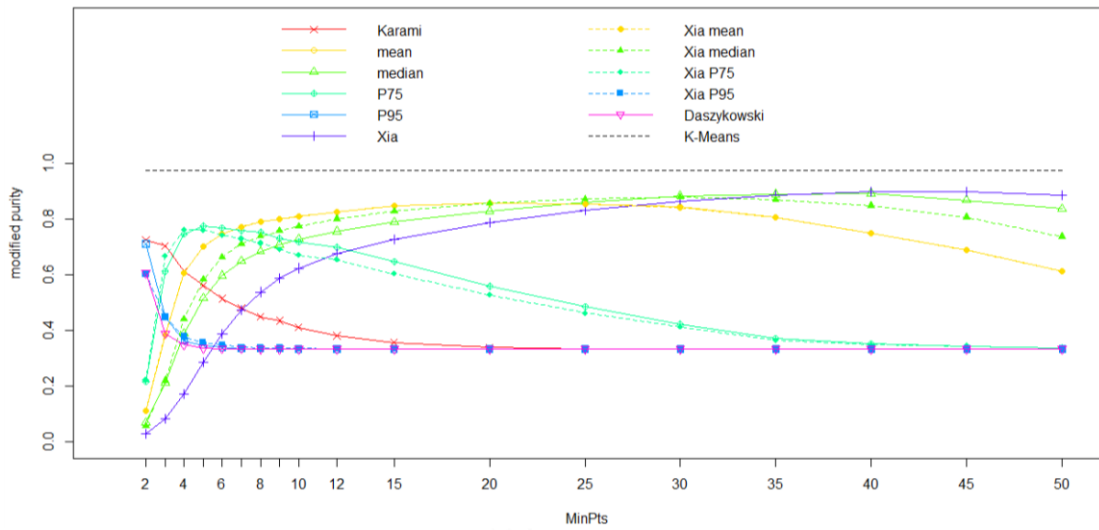
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



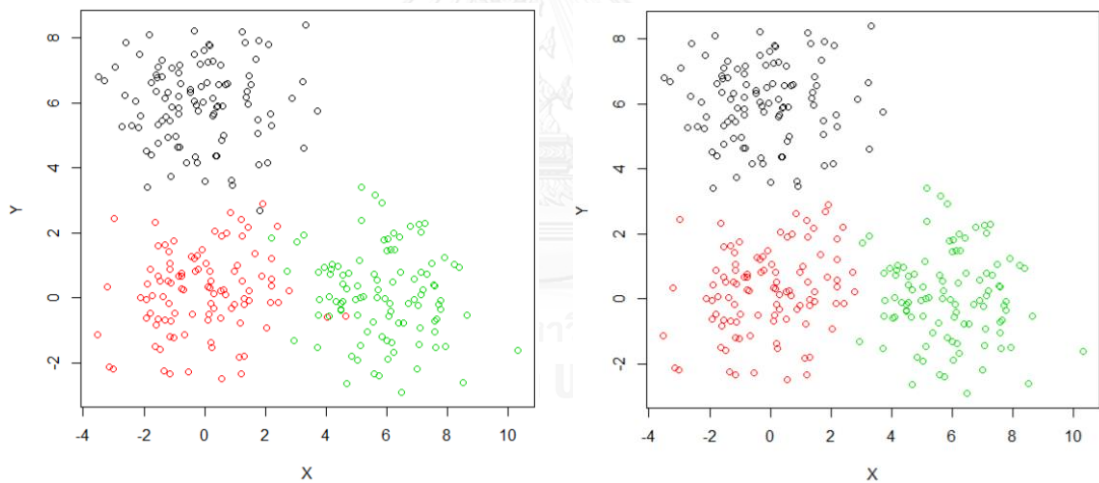
รูปที่ 4.5.1.4.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.1.4



รูปที่ 4.5.1.4.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.4



รูปที่ 4.5.1.4.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณี 4.5.1.4



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.4.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.5.1.4

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.4.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ ซึ่งวิธี Daszykowski ตั้งแต่ MinPts 20 เป็นต้นไปให้ค่า Eps สูงกว่าวิธีอื่น ๆ เป็นอย่างมาก

จากรูปที่ 4.5.1.4.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-5 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่อู่เข้าสู่ 0.333 (100/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 7 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.333 (100/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 3-6 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.333 (100/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 5 จากนั้น modified purity มีค่าลดลง และมีแนวโน้มที่จะลู่อู่เข้าสู่ 0.333

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-7 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ และเมื่อถึงค่า MinPts 25 ค่า modified purity มีค่าลดลง

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.4.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.4.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9768067

ตารางที่ 4.5.1.4.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.4

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.6178	0.3085	0.2492	0.3838	0.7479	0.1113	0.3085	0.2211	0.3865	0.8494	0.8429
3	0.7566	0.4557	0.3824	0.5450	0.9981	0.2695	0.4557	0.3862	0.5713	0.9916	1.0809
4	0.8736	0.5635	0.4794	0.6645	1.1722	0.3801	0.5635	0.4975	0.6982	1.1265	1.2653
5	0.9768	0.6518	0.5594	0.7655	1.3135	0.4677	0.6518	0.5865	0.8009	1.2440	1.4253
6	1.0700	0.7280	0.6290	0.8532	1.4298	0.5426	0.7280	0.6629	0.8889	1.3465	1.5712
7	1.1557	0.7966	0.6920	0.9320	1.5337	0.6102	0.7966	0.7316	0.9674	1.4377	1.7068
8	1.2355	0.8589	0.7490	1.0034	1.6270	0.6715	0.8589	0.7938	1.0385	1.5204	1.8347
9	1.3105	0.9170	0.8030	1.0710	1.7091	0.7285	0.9170	0.8517	1.1046	1.5977	1.9573
10	1.3813	0.9710	0.8534	1.1337	1.7870	0.7812	0.9710	0.9054	1.1659	1.6696	2.0747
12	1.5132	1.0703	0.9488	1.2464	1.9224	0.8793	1.0703	1.0046	1.2779	1.7989	2.2973
15	1.6918	1.2029	1.0753	1.3961	2.1007	1.0107	1.2029	1.1371	1.4263	1.9692	2.6058
20	1.9535	1.3959	1.2637	1.6131	2.3457	1.2033	1.3959	1.3303	1.6403	2.2109	3.0695
25	2.1841	1.5677	1.4349	1.8042	2.5571	1.3763	1.5677	1.5027	1.8289	2.4204	3.4923
30	2.3926	1.7261	1.5950	1.9754	2.7466	1.5365	1.7261	1.6619	2.0013	2.6092	3.8856
35	2.5843	1.8751	1.7456	2.1327	2.9243	1.6882	1.8751	1.8119	2.1621	2.7828	4.2571
40	2.7627	2.0186	1.8918	2.2808	3.0891	1.8350	2.0186	1.9566	2.3157	2.9461	4.6134
45	2.9303	2.1582	2.0326	2.4251	3.2503	1.9783	2.1582	2.0975	2.4641	3.1023	4.9552
50	3.0888	2.2958	2.1726	2.5636	3.4058	2.1202	2.2958	2.2367	2.6095	3.2532	5.2822

ตารางที่ 4.5.1.4.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

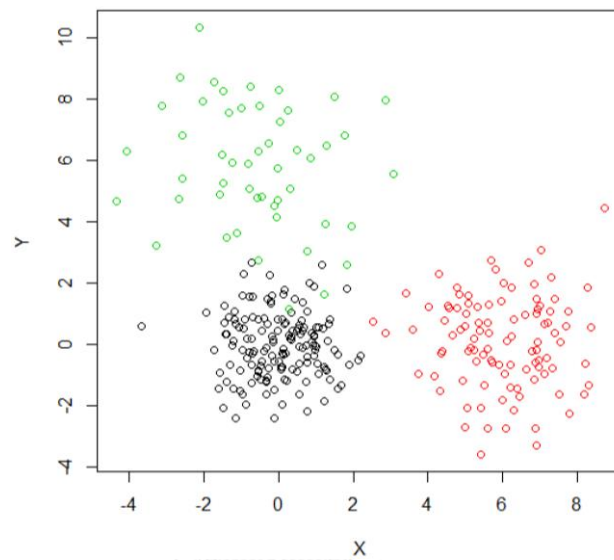
4.5.1.4

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7251	0.1107	0.0686	0.2168	0.7114	0.0270	0.1107	0.0559	0.2218	0.6031	0.6059
3	0.7044	0.3830	0.2119	0.6125	0.4486	0.0811	0.3830	0.2189	0.6680	0.4477	0.3884
4	0.6139	0.6050	0.3884	0.7480	0.3630	0.1715	0.6050	0.4397	0.7632	0.3777	0.3477
5	0.5606	0.7018	0.5157	0.7760	0.3471	0.2843	0.7018	0.5826	0.7612	0.3551	0.3362
6	0.5145	0.7478	0.5977	0.7675	0.3380	0.3866	0.7478	0.6633	0.7423	0.3480	0.3342
7	0.4799	0.7720	0.6494	0.7587	0.3345	0.4725	0.7720	0.7100	0.7304	0.3390	0.3344
8	0.4490	0.7907	0.6842	0.7541	0.3341	0.5381	0.7907	0.7379	0.7146	0.3362	0.3333
9	0.4349	0.8008	0.7083	0.7305	0.3343	0.5886	0.8008	0.7575	0.6915	0.3364	0.3333
10	0.4100	0.8114	0.7284	0.7193	0.3331	0.6240	0.8114	0.7747	0.6718	0.3347	0.3333
12	0.3807	0.8256	0.7563	0.6999	0.3332	0.6764	0.8256	0.8006	0.6557	0.3343	0.3333
15	0.3562	0.8483	0.7898	0.6472	0.3333	0.7276	0.8483	0.8279	0.6048	0.3338	0.3333
20	0.3381	0.8584	0.8296	0.5586	0.3333	0.7890	0.8584	0.8563	0.5281	0.3333	0.3333
25	0.3345	0.8549	0.8620	0.4843	0.3333	0.8317	0.8549	0.8725	0.4628	0.3333	0.3333
30	0.3333	0.8440	0.8846	0.4236	0.3333	0.8646	0.8434	0.8794	0.4134	0.3333	0.3333
35	0.3333	0.8076	0.8896	0.3708	0.3333	0.8855	0.8076	0.8698	0.3649	0.3333	0.3333
40	0.3333	0.7508	0.8930	0.3517	0.3333	0.8993	0.7502	0.8486	0.3495	0.3333	0.3333
45	0.3333	0.6900	0.8689	0.3424	0.3333	0.8989	0.6907	0.8067	0.3414	0.3333	0.3333
50	0.3333	0.6134	0.8380	0.3363	0.3333	0.8858	0.6128	0.7376	0.3371	0.3333	0.3333

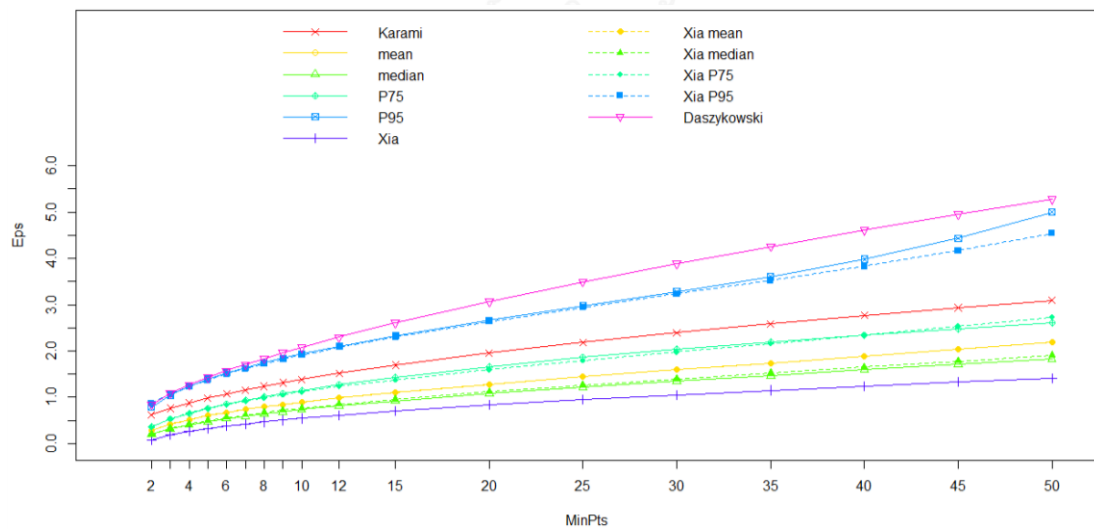
4.5.1.5 ขนาดตัวอย่าง N=300 กลุ่มละ 150, 100, 50 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$, $\mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

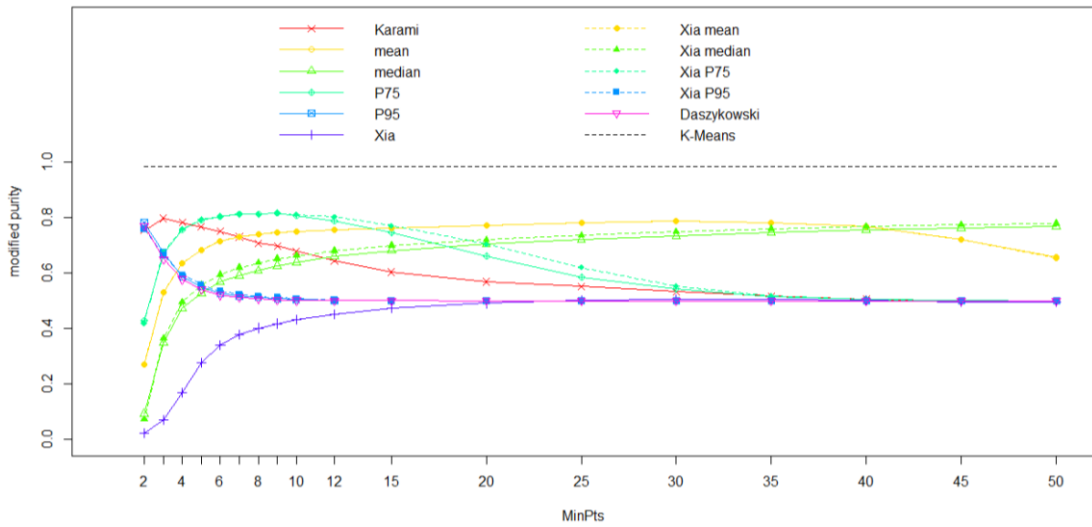
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



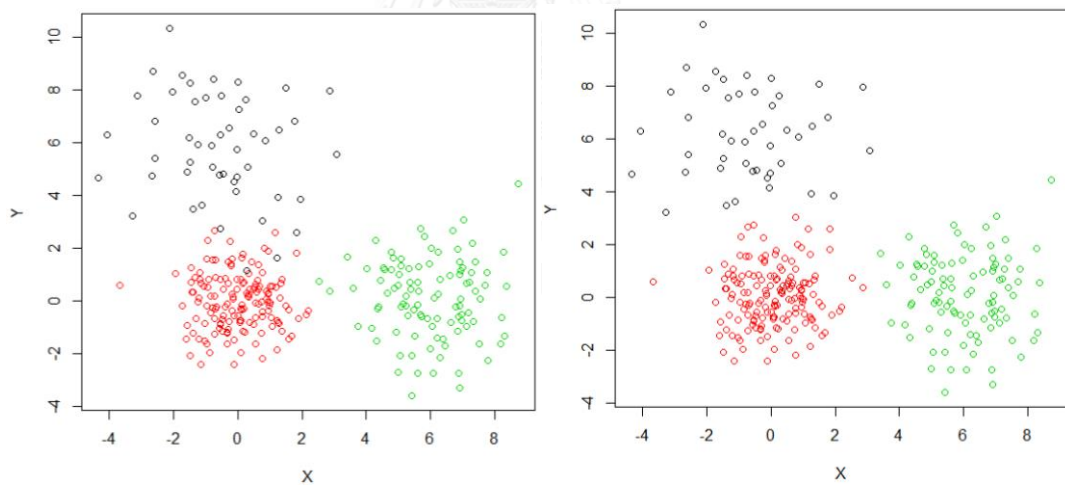
รูปที่ 4.5.1.5.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.5



รูปที่ 4.5.1.5.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.5



รูปที่ 4.5.1.5.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณี 4.5.1.5



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.5.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.5.1.5

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.5.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี P75, วิธี Xia P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.1.5.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.5 (150/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 9 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (150/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 3 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 40 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (150/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 จนกระทั่งสูงสุดที่ MinPts 9 จากนั้น modified purity มีค่าลดลง และมีแนวโน้มที่จะลู่เข้าสู่ 0.5

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 9-50 ยกเว้นวิธี mean และวิธี Xia mean ให้ค่า modified purity ลดลงที่ MinPts 40 เป็นต้นไป

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง ที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (150/300)

จากรูปที่ 4.5.1.5.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.5.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.98354

ตารางที่ 4.5.1.5.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.5

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.6181	0.2847	0.2041	0.3562	0.7795	0.0779	0.2847	0.1835	0.3492	0.8668	0.8425
3	0.7570	0.4190	0.3183	0.5246	1.0441	0.1859	0.4190	0.3237	0.5293	1.0542	1.0805
4	0.8741	0.5177	0.4033	0.6512	1.2407	0.2609	0.5177	0.4171	0.6536	1.2271	1.2654
5	0.9773	0.5984	0.4741	0.7547	1.3912	0.3228	0.5984	0.4928	0.7539	1.3698	1.4250
6	1.0706	0.6685	0.5360	0.8487	1.5256	0.3747	0.6685	0.5572	0.8409	1.4991	1.5703
7	1.1564	0.7308	0.5904	0.9307	1.6400	0.4209	0.7308	0.6143	0.9179	1.6140	1.7058
8	1.2362	0.7877	0.6398	1.0072	1.7460	0.4629	0.7877	0.6664	0.9882	1.7198	1.8345
9	1.3112	0.8406	0.6860	1.0780	1.8431	0.5018	0.8406	0.7146	1.0535	1.8187	1.9562
10	1.3822	0.8901	0.7298	1.1445	1.9348	0.5387	0.8901	0.7599	1.1144	1.9097	2.0735
12	1.5141	0.9808	0.8104	1.2677	2.1022	0.6062	0.9808	0.8429	1.2260	2.0775	2.2963
15	1.6928	1.1026	0.9183	1.4279	2.3318	0.6970	1.1026	0.9542	1.3757	2.3028	2.6048
20	1.9547	1.2817	1.0760	1.6614	2.6640	0.8290	1.2817	1.1170	1.5956	2.6372	3.0684
25	2.1854	1.4424	1.2170	1.8610	2.9770	0.9455	1.4424	1.2623	1.7932	2.9414	3.4905
30	2.3940	1.5929	1.3458	2.0347	3.2861	1.0513	1.5928	1.3970	1.9788	3.2330	3.8838
35	2.5858	1.7385	1.4696	2.1900	3.6094	1.1489	1.7385	1.5253	2.1596	3.5259	4.2553
40	2.7643	1.8841	1.5880	2.3345	3.9814	1.2398	1.8841	1.6509	2.3415	3.8321	4.6106
45	2.9320	2.0345	1.7063	2.4689	4.4353	1.3236	2.0345	1.7764	2.5315	4.1700	4.9520
50	3.0906	2.1920	1.8234	2.6026	4.9978	1.4017	2.1920	1.9038	2.7321	4.5440	5.2804

ตารางที่ 4.5.1.5.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

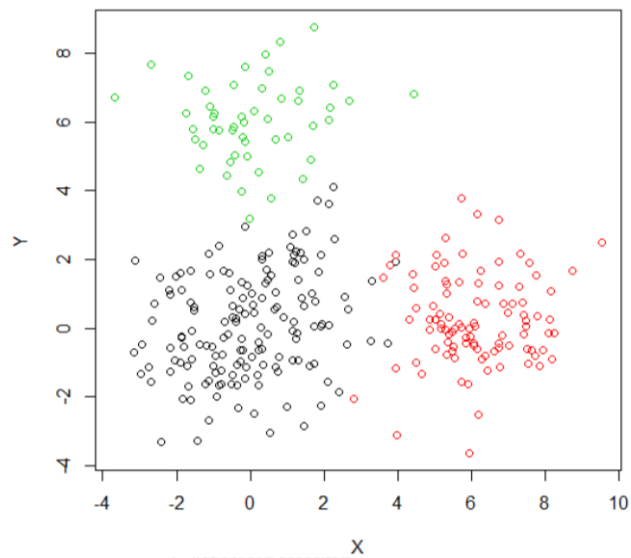
4.5.1.5

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7540	0.2694	0.0917	0.4273	0.7829	0.0216	0.2694	0.0706	0.4188	0.7590	0.7705
3	0.7972	0.5308	0.3478	0.6691	0.6735	0.0682	0.5308	0.3629	0.6755	0.6637	0.6500
4	0.7825	0.6344	0.4717	0.7574	0.5860	0.1678	0.6344	0.4947	0.7597	0.5934	0.5771
5	0.7661	0.6842	0.5275	0.7924	0.5499	0.2748	0.6842	0.5554	0.7930	0.5578	0.5411
6	0.7502	0.7135	0.5675	0.8053	0.5258	0.3389	0.7135	0.5941	0.8040	0.5342	0.5204
7	0.7306	0.7291	0.5916	0.8125	0.5160	0.3767	0.7291	0.6194	0.8127	0.5224	0.5118
8	0.7079	0.7399	0.6093	0.8118	0.5117	0.4013	0.7399	0.6367	0.8123	0.5152	0.5069
9	0.6983	0.7469	0.6257	0.8170	0.5075	0.4173	0.7469	0.6517	0.8159	0.5126	0.5030
10	0.6787	0.7494	0.6391	0.8079	0.5045	0.4314	0.7494	0.6618	0.8114	0.5068	0.5011
12	0.6436	0.7569	0.6601	0.7894	0.5006	0.4512	0.7569	0.6799	0.8025	0.5017	0.5006
15	0.6034	0.7633	0.6808	0.7456	0.5003	0.4718	0.7633	0.6995	0.7705	0.5009	0.5003
20	0.5682	0.7729	0.7041	0.6621	0.5000	0.4905	0.7729	0.7197	0.7044	0.5000	0.5000
25	0.5509	0.7823	0.7206	0.5852	0.5000	0.5018	0.7823	0.7364	0.6203	0.5000	0.5000
30	0.5323	0.7873	0.7336	0.5416	0.5000	0.5037	0.7873	0.7492	0.5538	0.5000	0.5000
35	0.5165	0.7821	0.7464	0.5137	0.5000	0.5047	0.7821	0.7591	0.5183	0.5000	0.5000
40	0.5046	0.7688	0.7553	0.5037	0.5000	0.5006	0.7688	0.7675	0.5040	0.5000	0.5000
45	0.5000	0.7197	0.7626	0.5008	0.5000	0.4956	0.7197	0.7743	0.5000	0.5000	0.5000
50	0.5000	0.6562	0.7674	0.5000	0.5000	0.4943	0.6557	0.7780	0.5000	0.5000	0.5000

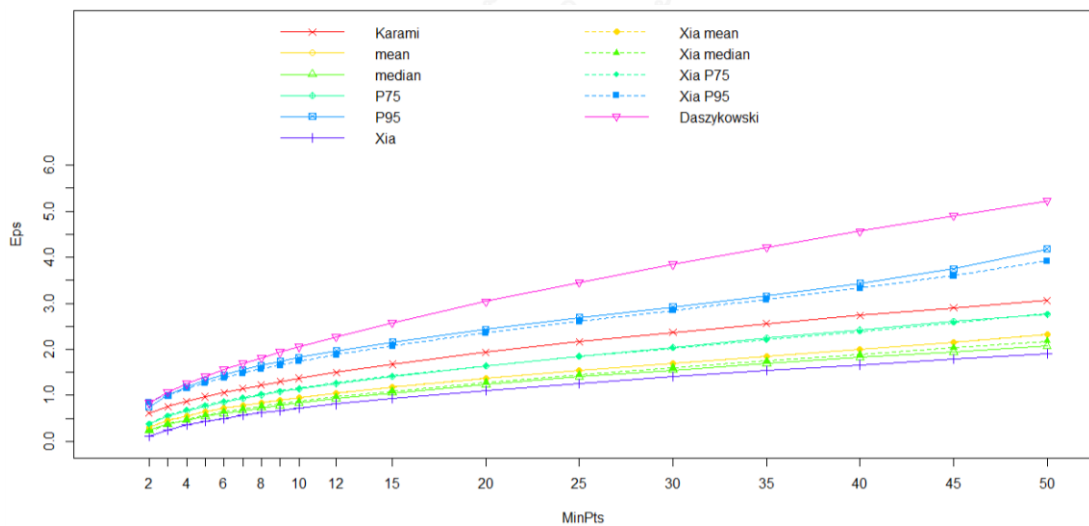
4.5.1.6 ขนาดตัวอย่าง N=300 กลุ่มละ 150, 100, 50 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$, $\mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

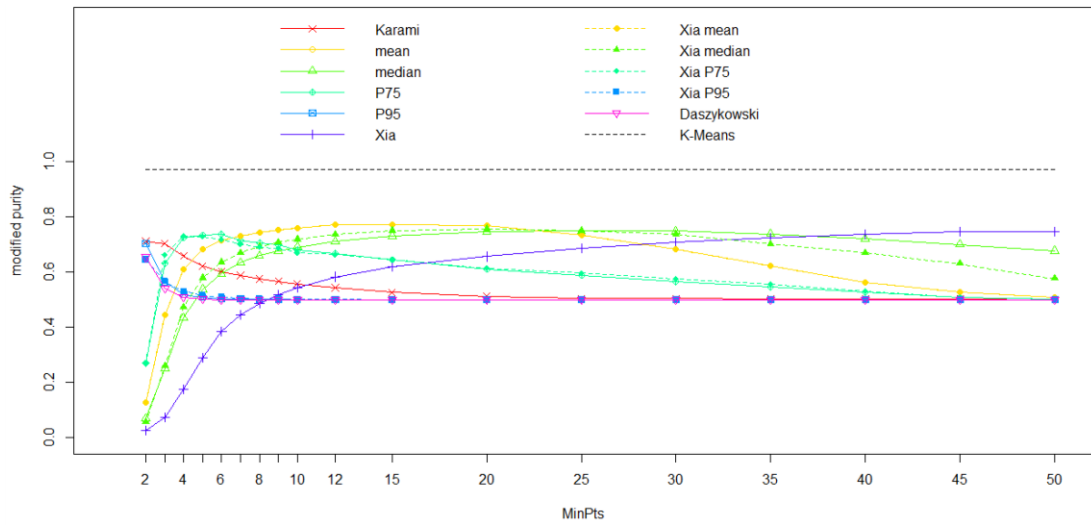
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



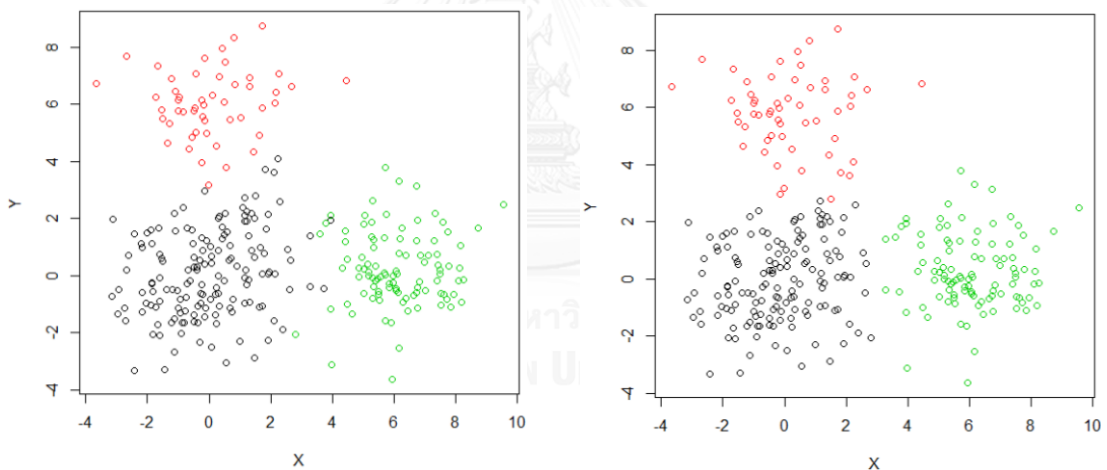
รูปที่ 4.5.1.6.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.1.6



รูปที่ 4.5.1.6.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.6



รูปที่ 4.5.1.6.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.1.6



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.6.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.6

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.6.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี

Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.1.6.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.5 (150/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 6 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (150/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 2 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 25 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (150/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 5 จากนั้น modified purity มีค่าลดลง และมีแนวโน้มที่จะลู่เข้าสู่ 0.5

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ และเมื่อถึงค่า MinPts 20 ค่า modified purity มีค่าลดลง

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.6.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.6.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9724733

ตารางที่ 4.5.1.6.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.6

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.6119	0.3007	0.2379	0.3741	0.7470	0.1019	0.3007	0.2106	0.3753	0.8478	0.8346
3	0.7494	0.4448	0.3671	0.5396	1.0013	0.2481	0.4448	0.3700	0.5595	1.0004	1.0697
4	0.8653	0.5498	0.4613	0.6618	1.1793	0.3500	0.5498	0.4768	0.6855	1.1432	1.2528
5	0.9674	0.6363	0.5401	0.7630	1.3219	0.4306	0.6363	0.5623	0.7881	1.2691	1.4108
6	1.0598	0.7112	0.6078	0.8525	1.4425	0.4996	0.7112	0.6358	0.8766	1.3800	1.5552
7	1.1447	0.7785	0.6709	0.9319	1.5527	0.5616	0.7785	0.7018	0.9559	1.4796	1.6893
8	1.2237	0.8400	0.7287	1.0067	1.6525	0.6174	0.8400	0.7616	1.0281	1.5721	1.8166
9	1.2980	0.8967	0.7810	1.0747	1.7393	0.6694	0.8967	0.8169	1.0946	1.6562	1.9378
10	1.3682	0.9497	0.8324	1.1385	1.8206	0.7179	0.9497	0.8686	1.1565	1.7349	2.0536
12	1.4988	1.0472	0.9253	1.2543	1.9646	0.8069	1.0472	0.9635	1.2702	1.8797	2.2739
15	1.6757	1.1778	1.0515	1.4106	2.1571	0.9270	1.1778	1.0908	1.4220	2.0714	2.5797
20	1.9349	1.3690	1.2369	1.6398	2.4338	1.1024	1.3690	1.2769	1.6434	2.3518	3.0391
25	2.1633	1.5399	1.4005	1.8476	2.6840	1.2585	1.5399	1.4430	1.8414	2.6037	3.4572
30	2.3697	1.6995	1.5495	2.0456	2.9225	1.4020	1.6995	1.5973	2.0267	2.8432	3.8477
35	2.5596	1.8531	1.6881	2.2385	3.1602	1.5364	1.8531	1.7443	2.2065	3.0821	4.2151
40	2.7363	2.0047	1.8205	2.4269	3.4279	1.6640	2.0048	1.8878	2.3863	3.3301	4.5680
45	2.9023	2.1596	1.9466	2.6029	3.7484	1.7866	2.1596	2.0315	2.5732	3.6012	4.9047
50	3.0593	2.3239	2.0694	2.7568	4.1804	1.9026	2.3238	2.1789	2.7774	3.9212	5.2297



ตารางที่ 4.5.1.6.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

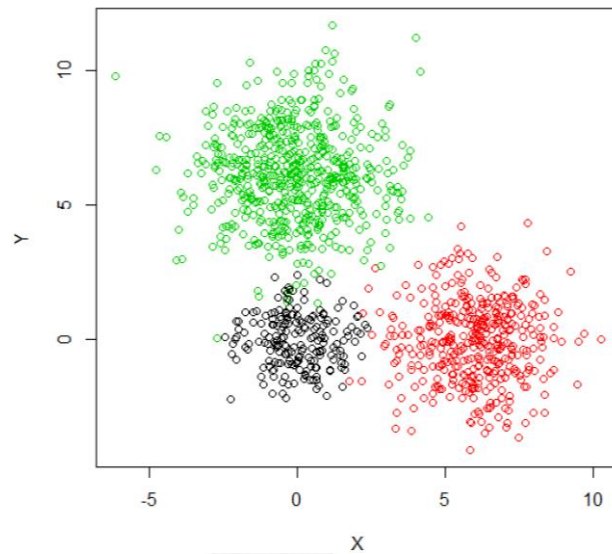
4.5.1.6

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7118	0.1274	0.0682	0.2679	0.7033	0.0245	0.1274	0.0544	0.2695	0.6447	0.6534
3	0.7030	0.4445	0.2494	0.6310	0.5647	0.0735	0.4444	0.2579	0.6627	0.5667	0.5407
4	0.6585	0.6113	0.4340	0.7255	0.5184	0.1731	0.6113	0.4726	0.7319	0.5289	0.5084
5	0.6219	0.6827	0.5368	0.7334	0.5094	0.2894	0.6827	0.5785	0.7280	0.5158	0.5022
6	0.6014	0.7146	0.5947	0.7385	0.5030	0.3835	0.7146	0.6358	0.7180	0.5079	0.5000
7	0.5878	0.7304	0.6337	0.7150	0.5020	0.4439	0.7304	0.6687	0.7018	0.5034	0.4990
8	0.5748	0.7431	0.6592	0.7061	0.5009	0.4848	0.7431	0.6925	0.6905	0.5025	0.4991
9	0.5661	0.7533	0.6766	0.7002	0.5006	0.5171	0.7533	0.7075	0.6874	0.5014	0.4995
10	0.5572	0.7606	0.6902	0.6797	0.4999	0.5432	0.7606	0.7188	0.6701	0.5001	0.4996
12	0.5432	0.7710	0.7112	0.6676	0.4996	0.5824	0.7710	0.7365	0.6633	0.5003	0.4998
15	0.5255	0.7715	0.7310	0.6436	0.4996	0.6187	0.7715	0.7498	0.6450	0.4994	0.4999
20	0.5121	0.7699	0.7468	0.6105	0.4997	0.6559	0.7699	0.7567	0.6142	0.4997	0.5000
25	0.5043	0.7351	0.7485	0.5879	0.4999	0.6857	0.7351	0.7503	0.5958	0.4998	0.5000
30	0.5038	0.6829	0.7483	0.5644	0.5000	0.7076	0.6829	0.7379	0.5754	0.4999	0.5000
35	0.5027	0.6212	0.7381	0.5452	0.5000	0.7250	0.6212	0.7013	0.5563	0.5000	0.5000
40	0.5018	0.5620	0.7204	0.5257	0.5000	0.7377	0.5620	0.6692	0.5300	0.5000	0.5000
45	0.5019	0.5275	0.7003	0.5077	0.5000	0.7459	0.5275	0.6300	0.5084	0.5000	0.5000
50	0.5000	0.5091	0.6772	0.5000	0.5000	0.7477	0.5091	0.5758	0.5003	0.5000	0.5000

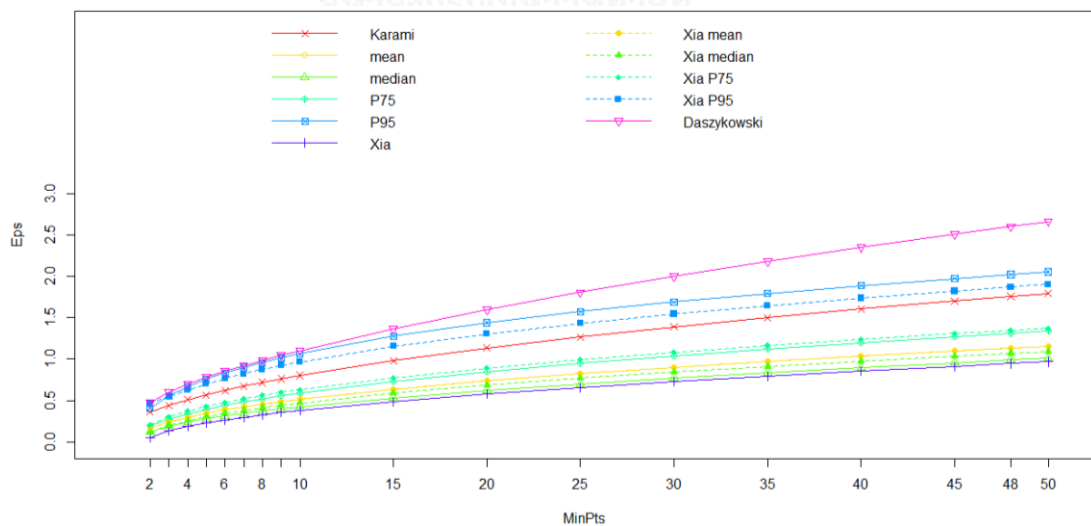
4.5.1.7 ขนาดตัวอย่าง $N=1,200$ กลุ่มละ 200, 400, 600 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

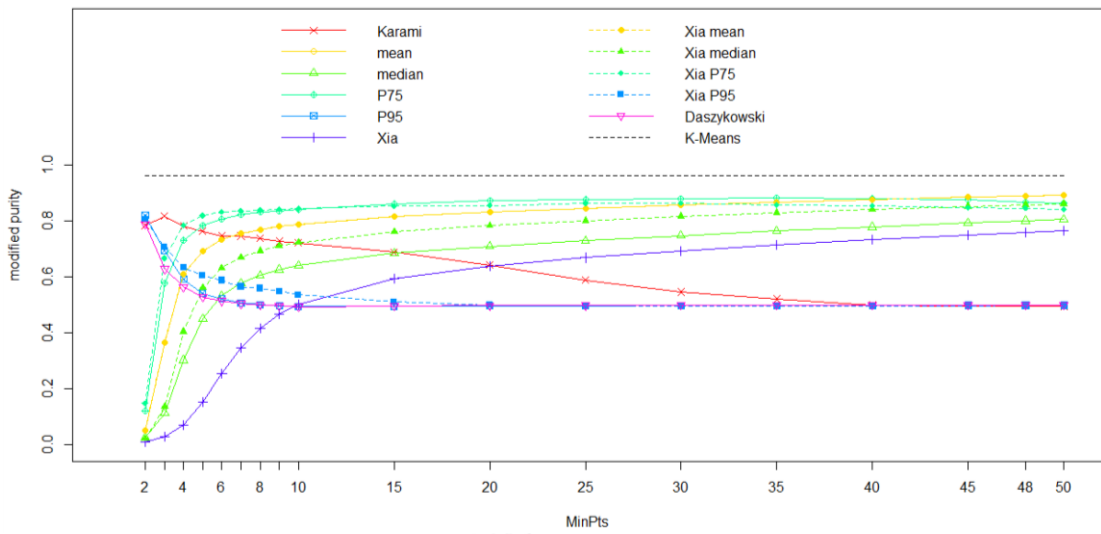
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



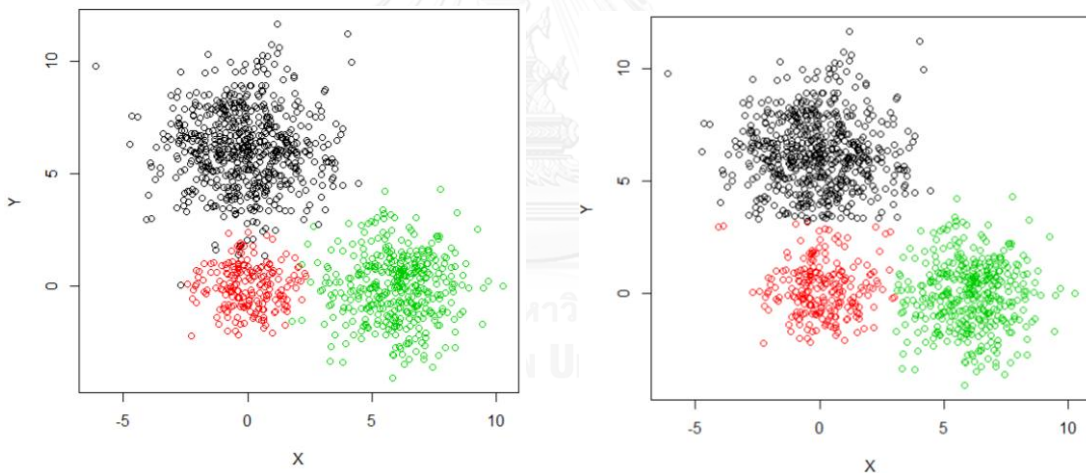
รูปที่ 4.5.1.7.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.1.7



รูปที่ 4.5.1.7.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.7



รูปที่ 4.5.1.7.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.1.7



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.7.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.7

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.7.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.1.7.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และเข้าสู่ 0.5 (600/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 7 เป็นต้นไปของวิธี P95 ค่า modified purity เข้าสู่ 0.5 (600/1200) ส่วนที่ MinPts 20 เป็นต้นไปของวิธี Xia P95 ค่า modified purity เข้าสู่ 0.5 (600/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 3 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 40 เป็นต้นไป ค่า modified purity เข้าสู่ 0.5 (600/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 8 เป็นต้นไป จนกระทั่งที่ค่า MinPts ช่วงปลาย ๆ พบว่าค่า modified purity มีค่าลดลงเล็กน้อย

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-6 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 8-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.7.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.7.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9625217

ตารางที่ 4.5.1.7.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.7

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.3585	0.1615	0.1255	0.1949	0.4039	0.0532	0.1615	0.1124	0.2018	0.4584	0.4751
3	0.4391	0.2398	0.1924	0.2778	0.5540	0.1319	0.2398	0.1988	0.3020	0.5433	0.6033
4	0.5070	0.2975	0.2410	0.3407	0.6644	0.1854	0.2975	0.2564	0.3718	0.6271	0.7003
5	0.5669	0.3450	0.2805	0.3933	0.7551	0.2282	0.3450	0.3028	0.4286	0.6991	0.7827
6	0.6210	0.3861	0.3150	0.4394	0.8319	0.2647	0.3861	0.3427	0.4776	0.7629	0.8562
7	0.6707	0.4228	0.3458	0.4810	0.8992	0.2972	0.4228	0.3781	0.5213	0.8202	0.9236
8	0.7170	0.4563	0.3741	0.5192	0.9592	0.3266	0.4563	0.4104	0.5611	0.8727	0.9867
9	0.7605	0.4872	0.4004	0.5545	1.0151	0.3538	0.4872	0.4401	0.5977	0.9210	1.0464
10	0.8017	0.5159	0.4251	0.5878	1.0665	0.3791	0.5159	0.4678	0.6317	0.9659	1.1033
15	0.9818	0.6384	0.5318	0.7307	1.2765	0.4879	0.6384	0.5859	0.7759	1.1547	1.3628
20	1.1337	0.7381	0.6209	0.8482	1.4383	0.5780	0.7381	0.6826	0.8925	1.3052	1.5947
25	1.2675	0.8245	0.6991	0.9502	1.5744	0.6569	0.8245	0.7666	0.9928	1.4327	1.8060
30	1.3885	0.9016	0.7699	1.0409	1.6914	0.7282	0.9016	0.8418	1.0819	1.5447	2.0000
35	1.4998	0.9719	0.8355	1.1233	1.7938	0.7939	0.9719	0.9107	1.1627	1.6449	2.1799
40	1.6033	1.0370	0.8970	1.1999	1.8886	0.8551	1.0370	0.9746	1.2372	1.7365	2.3485
45	1.7006	1.0980	0.9549	1.2714	1.9741	0.9129	1.0980	1.0346	1.3068	1.8211	2.5080
48	1.7563	1.1330	0.9882	1.3124	2.0219	0.9462	1.1330	1.0690	1.3465	1.8691	2.6003
50	1.7926	1.1557	1.0100	1.3390	2.0535	0.9679	1.1557	1.0914	1.3722	1.9000	2.6602

ตารางที่ 4.5.1.7.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

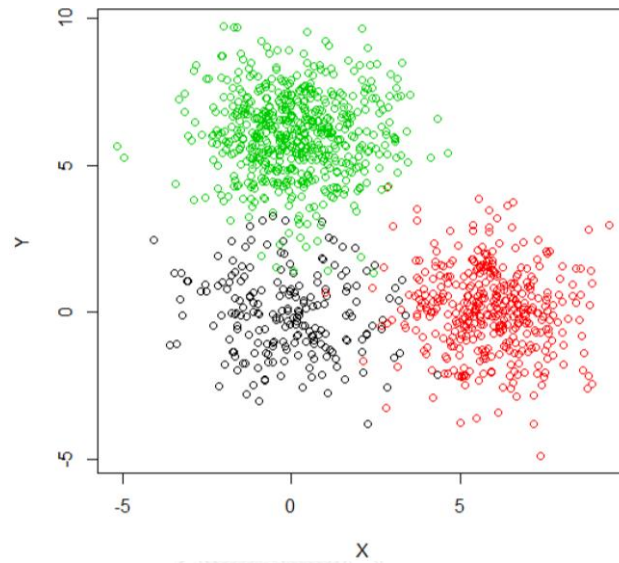
4.5.1.7

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7848	0.0507	0.0244	0.1213	0.8200	0.0080	0.0507	0.0195	0.1482	0.8046	0.7896
3	0.8171	0.3643	0.1118	0.5765	0.6940	0.0278	0.3643	0.1350	0.6672	0.7058	0.6291
4	0.7826	0.6085	0.3014	0.7317	0.5911	0.0702	0.6085	0.4040	0.7846	0.6346	0.5641
5	0.7632	0.6935	0.4500	0.7834	0.5402	0.1511	0.6935	0.5598	0.8196	0.6040	0.5283
6	0.7470	0.7325	0.5319	0.8080	0.5219	0.2530	0.7325	0.6315	0.8329	0.5876	0.5144
7	0.7466	0.7551	0.5773	0.8243	0.5064	0.3455	0.7551	0.6690	0.8370	0.5652	0.5051
8	0.7376	0.7699	0.6055	0.8334	0.5004	0.4167	0.7699	0.6928	0.8390	0.5587	0.4991
9	0.7276	0.7814	0.6264	0.8371	0.4961	0.4658	0.7814	0.7109	0.8428	0.5483	0.4970
10	0.7215	0.7886	0.6400	0.8416	0.4949	0.5023	0.7886	0.7226	0.8445	0.5353	0.4932
15	0.6880	0.8152	0.6850	0.8618	0.4954	0.5949	0.8152	0.7610	0.8531	0.5099	0.4956
20	0.6421	0.8325	0.7098	0.8725	0.4961	0.6379	0.8325	0.7834	0.8556	0.4996	0.4974
25	0.5890	0.8459	0.7297	0.8772	0.4964	0.6687	0.8459	0.8009	0.8643	0.4963	0.4984
30	0.5464	0.8577	0.7476	0.8796	0.4971	0.6930	0.8577	0.8162	0.8649	0.4968	0.4991
35	0.5206	0.8684	0.7641	0.8828	0.4976	0.7139	0.8684	0.8297	0.8577	0.4956	0.4994
40	0.4981	0.8772	0.7787	0.8808	0.4979	0.7327	0.8772	0.8412	0.8551	0.4963	0.4997
45	0.4959	0.8857	0.7926	0.8755	0.4982	0.7505	0.8857	0.8521	0.8469	0.4967	0.4998
48	0.4958	0.8899	0.8003	0.8686	0.4983	0.7599	0.8899	0.8577	0.8463	0.4970	0.4999
50	0.4955	0.8926	0.8054	0.8634	0.4985	0.7659	0.8926	0.8614	0.8413	0.4972	0.4999

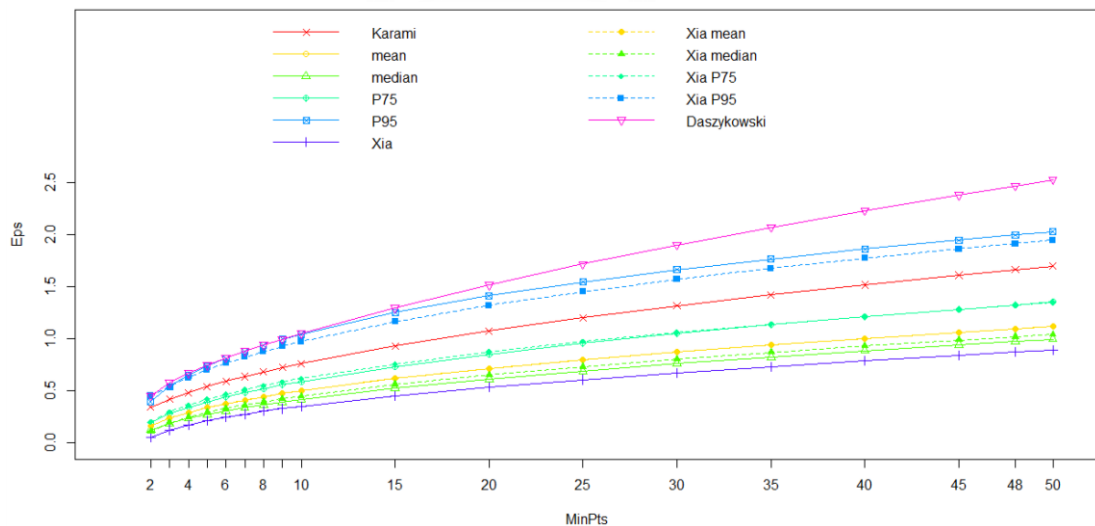
4.5.1.8 ขนาดตัวอย่าง N=1,200 กลุ่มละ 200, 400, 600 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

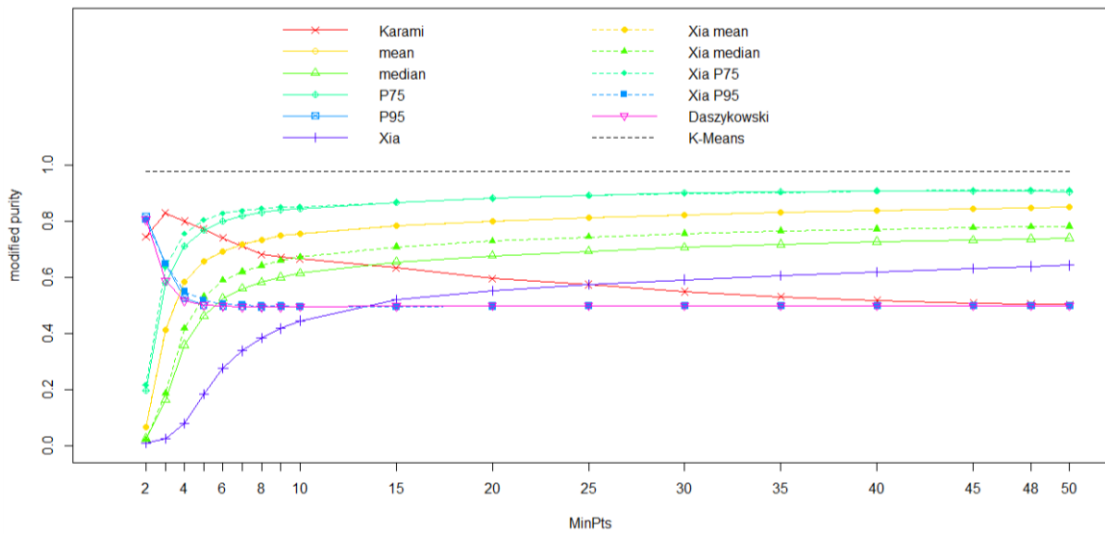
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



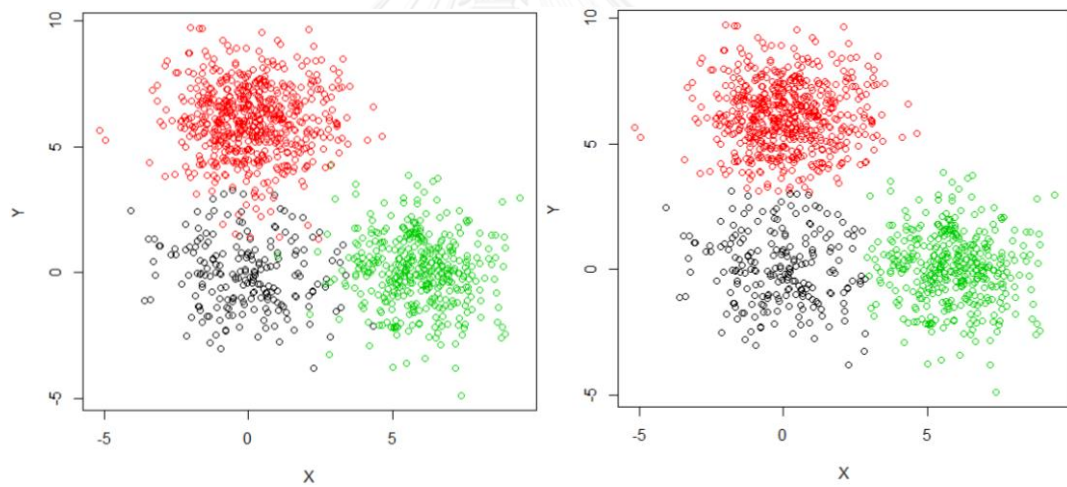
รูปที่ 4.5.1.8.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.1.8



รูปที่ 4.5.1.8.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.8



รูปที่ 4.5.1.8.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณี 4.5.1.8



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.8.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.5.1.8

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.8.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี

ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.1.8.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และเข้าสู่ 0.5 (600/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นที่ MinPts 5 เป็นต้นไป ค่า modified purity อยู่เข้าสู่ 0.5 (600/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 3 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 45 เป็นต้นไป ค่า modified purity อยู่เข้าสู่ 0.5 (600/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นเรื่อย ๆ จนกระทั่งในส่วนปลายของค่า MinPts ค่า modified purity ค่อย ๆ ลดลงเล็กน้อย

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-6 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 8-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.8.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.8.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9789283

ตารางที่ 4.5.1.8.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.8

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.3391	0.1558	0.1191	0.1901	0.3951	0.0487	0.1558	0.1063	0.1939	0.4507	0.4511
3	0.4153	0.2312	0.1837	0.2737	0.5394	0.1204	0.2312	0.1883	0.2917	0.5387	0.5729
4	0.4796	0.2869	0.2308	0.3375	0.6472	0.1696	0.2869	0.2432	0.3599	0.6241	0.6651
5	0.5362	0.3327	0.2699	0.3905	0.7346	0.2090	0.3327	0.2874	0.4154	0.6974	0.7434
6	0.5873	0.3724	0.3040	0.4368	0.8109	0.2427	0.3724	0.3254	0.4634	0.7624	0.8132
7	0.6344	0.4080	0.3347	0.4789	0.8768	0.2725	0.4080	0.3592	0.5063	0.8211	0.8772
8	0.6782	0.4403	0.3630	0.5176	0.9376	0.2997	0.4403	0.3899	0.5452	0.8746	0.9371
9	0.7193	0.4702	0.3894	0.5537	0.9913	0.3248	0.4702	0.4183	0.5811	0.9240	0.9937
10	0.7582	0.4980	0.4139	0.5870	1.0411	0.3482	0.4980	0.4447	0.6145	0.9698	1.0479
15	0.9287	0.6163	0.5200	0.7308	1.2488	0.4483	0.6163	0.5570	0.7559	1.1636	1.2943
20	1.0723	0.7129	0.6085	0.8493	1.4116	0.5311	0.7129	0.6491	0.8708	1.3193	1.5145
25	1.1989	0.7965	0.6863	0.9533	1.5440	0.6036	0.7965	0.7291	0.9697	1.4519	1.7152
30	1.3133	0.8711	0.7571	1.0457	1.6616	0.6690	0.8711	0.8007	1.0578	1.5691	1.8993
35	1.4186	0.9394	0.8219	1.1302	1.7654	0.7291	0.9394	0.8663	1.1380	1.6750	2.0701
40	1.5165	1.0025	0.8825	1.2079	1.8602	0.7852	1.0025	0.9271	1.2121	1.7720	2.2301
45	1.6085	1.0616	0.9396	1.2813	1.9470	0.8380	1.0616	0.9841	1.2812	1.8621	2.3816
48	1.6612	1.0954	0.9720	1.3236	1.9962	0.8684	1.0954	1.0168	1.3207	1.9133	2.4691
50	1.6955	1.1174	0.9934	1.3512	2.0273	0.8881	1.1174	1.0381	1.3464	1.9466	2.5261

ตารางที่ 4.5.1.8.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีที่

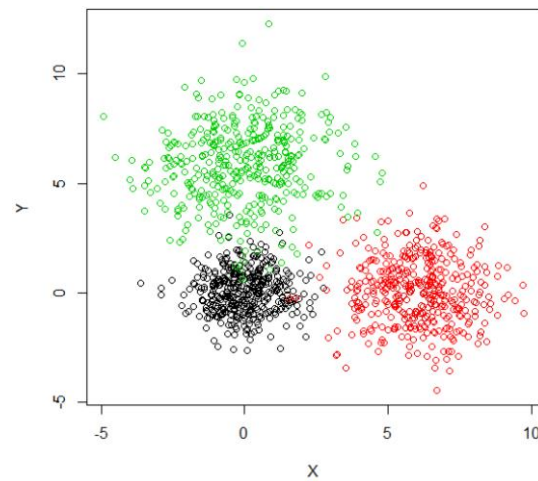
4.5.1.8

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7458	0.0658	0.0249	0.1967	0.8179	0.0076	0.0658	0.0192	0.2160	0.8051	0.8072
3	0.8304	0.4113	0.1623	0.5818	0.6493	0.0256	0.4113	0.1860	0.6379	0.6497	0.5879
4	0.8007	0.5854	0.3584	0.7127	0.5270	0.0778	0.5854	0.4167	0.7564	0.5495	0.5167
5	0.7720	0.6572	0.4622	0.7695	0.5013	0.1829	0.6572	0.5300	0.8047	0.5196	0.5018
6	0.7412	0.6933	0.5249	0.8019	0.4971	0.2754	0.6933	0.5885	0.8284	0.5061	0.4976
7	0.7122	0.7193	0.5602	0.8212	0.4960	0.3380	0.7193	0.6199	0.8383	0.5026	0.4952
8	0.6821	0.7350	0.5837	0.8314	0.4954	0.3843	0.7350	0.6418	0.8454	0.4982	0.4946
9	0.6719	0.7487	0.6008	0.8413	0.4960	0.4175	0.7487	0.6596	0.8515	0.4981	0.4953
10	0.6666	0.7570	0.6148	0.8462	0.4959	0.4458	0.7570	0.6719	0.8515	0.4968	0.4959
15	0.6344	0.7848	0.6558	0.8678	0.4975	0.5195	0.7848	0.7089	0.8682	0.4963	0.4979
20	0.5954	0.8012	0.6777	0.8835	0.4984	0.5527	0.8012	0.7301	0.8818	0.4975	0.4990
25	0.5747	0.8135	0.6935	0.8932	0.4989	0.5739	0.8135	0.7445	0.8919	0.4982	0.4995
30	0.5495	0.8243	0.7071	0.9014	0.4992	0.5921	0.8243	0.7569	0.8983	0.4987	0.4997
35	0.5308	0.8322	0.7184	0.9069	0.4994	0.6065	0.8322	0.7658	0.9039	0.4991	0.4999
40	0.5175	0.8392	0.7272	0.9091	0.4996	0.6197	0.8392	0.7726	0.9087	0.4993	0.4999
45	0.5084	0.8450	0.7334	0.9087	0.4997	0.6327	0.8450	0.7779	0.9128	0.4995	0.5000
48	0.5048	0.8484	0.7378	0.9101	0.4997	0.6397	0.8484	0.7806	0.9127	0.4995	0.5000
50	0.5037	0.8505	0.7402	0.9048	0.4997	0.6434	0.8505	0.7823	0.9120	0.4996	0.5000

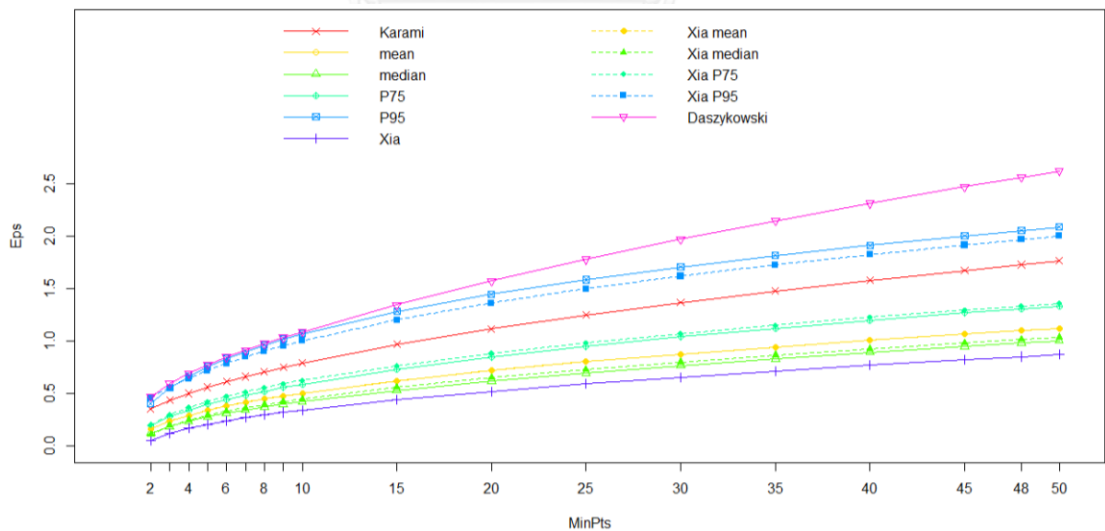
4.5.1.9 ขนาดตัวอย่าง N=1,200 กลุ่มละ 400, 400, 400 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

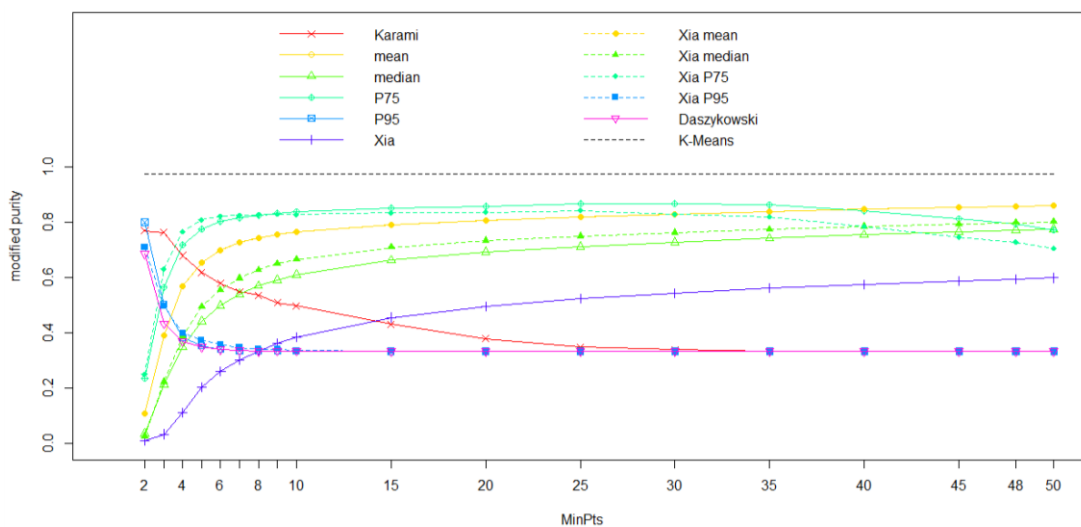
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



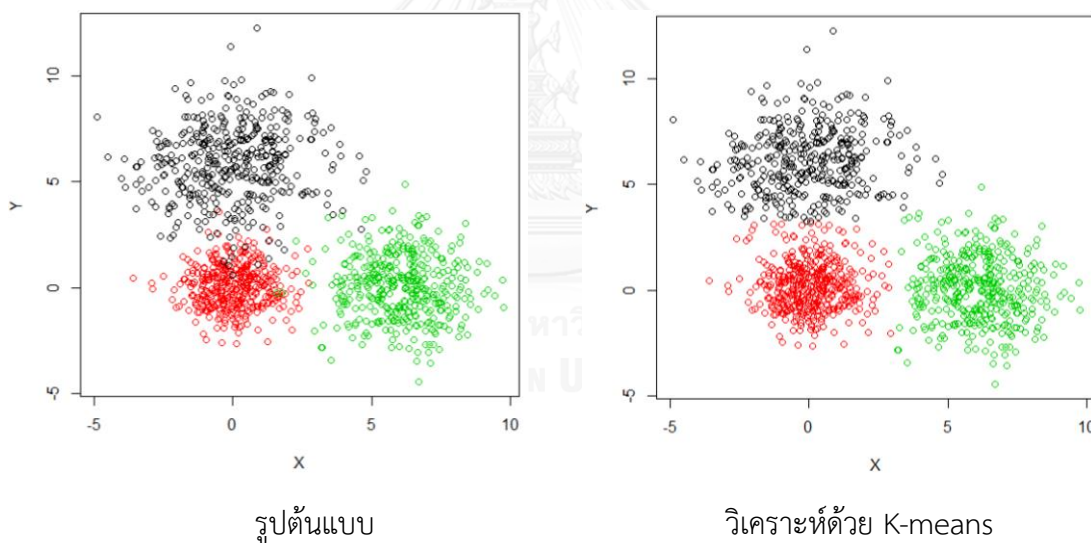
รูปที่ 4.5.1.9.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.1.9



รูปที่ 4.5.1.9.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.9



รูปที่ 4.5.1.9.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณี 4.5.1.9



รูปที่ 4.5.1.9.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.5.1.9

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.9.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี

Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.1.9.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.333 (400/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 7 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (400/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 3 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 35 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (400/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 จนกระทั่งสูงสุดที่ MinPts 25 จากนั้น modified purity มีค่าลดลง

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 15-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.9.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.9.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9756517

ตารางที่ 4.5.1.9.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.9

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.3526	0.1572	0.1193	0.1919	0.4019	0.0481	0.1572	0.1063	0.1953	0.4589	0.4680
3	0.4318	0.2335	0.1853	0.2765	0.5520	0.1174	0.2335	0.1881	0.2949	0.5538	0.5941
4	0.4986	0.2896	0.2336	0.3397	0.6623	0.1655	0.2896	0.2429	0.3641	0.6425	0.6899
5	0.5575	0.3358	0.2735	0.3924	0.7532	0.2040	0.3358	0.2871	0.4205	0.7188	0.7711
6	0.6107	0.3758	0.3084	0.4385	0.8312	0.2366	0.3758	0.3249	0.4691	0.7866	0.8435
7	0.6596	0.4115	0.3398	0.4801	0.8989	0.2656	0.4115	0.3585	0.5125	0.8476	0.9100
8	0.7052	0.4439	0.3684	0.5189	0.9597	0.2920	0.4439	0.3889	0.5518	0.9030	0.9720
9	0.7480	0.4738	0.3949	0.5540	1.0166	0.3163	0.4738	0.4171	0.5880	0.9541	1.0307
10	0.7884	0.5017	0.4198	0.5870	1.0675	0.3391	0.5017	0.4433	0.6217	1.0014	1.0869
15	0.9656	0.6201	0.5267	0.7283	1.2793	0.4368	0.6201	0.5549	0.7643	1.2008	1.3425
20	1.1150	0.7169	0.6160	0.8456	1.4457	0.5178	0.7169	0.6466	0.8803	1.3613	1.5710
25	1.2466	0.8005	0.6940	0.9467	1.5843	0.5888	0.8005	0.7260	0.9800	1.4974	1.7792
30	1.3656	0.8751	0.7648	1.0366	1.7058	0.6530	0.8751	0.7972	1.0686	1.6172	1.9700
35	1.4750	0.9431	0.8299	1.1187	1.8138	0.7121	0.9431	0.8623	1.1491	1.7250	2.1472
40	1.5768	1.0060	0.8907	1.1948	1.9116	0.7672	1.0060	0.9227	1.2234	1.8237	2.3133
45	1.6725	1.0649	0.9478	1.2656	2.0010	0.8192	1.0649	0.9793	1.2926	1.9149	2.4705
48	1.7273	1.0986	0.9808	1.3059	2.0527	0.8492	1.0986	1.0118	1.3322	1.9668	2.5612
50	1.7630	1.1205	1.0020	1.3325	2.0854	0.8687	1.1205	1.0330	1.3579	2.0003	2.6202

ตารางที่ 4.5.1.9.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

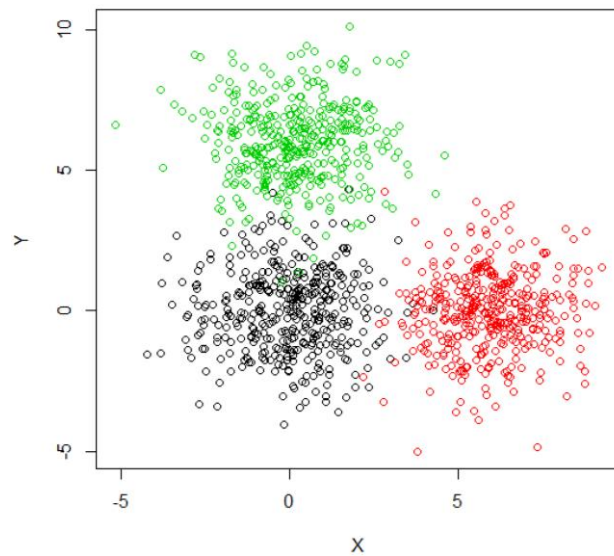
4.5.1.9

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7698	0.1081	0.0339	0.2357	0.8007	0.0081	0.1081	0.0245	0.2489	0.7092	0.6863
3	0.7638	0.3915	0.2118	0.5638	0.5030	0.0324	0.3915	0.2224	0.6316	0.4966	0.4334
4	0.6795	0.5693	0.3484	0.7184	0.3826	0.1095	0.5693	0.3853	0.7667	0.3984	0.3687
5	0.6180	0.6558	0.4409	0.7762	0.3530	0.2037	0.6558	0.4942	0.8093	0.3739	0.3484
6	0.5787	0.6994	0.4993	0.8031	0.3403	0.2595	0.6994	0.5548	0.8215	0.3574	0.3390
7	0.5487	0.7266	0.5397	0.8166	0.3357	0.3004	0.7266	0.5975	0.8247	0.3471	0.3358
8	0.5354	0.7436	0.5702	0.8255	0.3339	0.3346	0.7436	0.6282	0.8268	0.3407	0.3333
9	0.5094	0.7550	0.5911	0.8338	0.3340	0.3613	0.7550	0.6501	0.8306	0.3402	0.3334
10	0.4974	0.7648	0.6095	0.8384	0.3335	0.3826	0.7648	0.6648	0.8281	0.3359	0.3335
15	0.4307	0.7922	0.6631	0.8512	0.3332	0.4552	0.7922	0.7096	0.8351	0.3337	0.3332
20	0.3787	0.8077	0.6921	0.8583	0.3333	0.4940	0.8077	0.7341	0.8370	0.3332	0.3333
25	0.3476	0.8204	0.7115	0.8685	0.3333	0.5229	0.8204	0.7495	0.8430	0.3333	0.3333
30	0.3382	0.8301	0.7278	0.8688	0.3333	0.5440	0.8301	0.7629	0.8286	0.3333	0.3333
35	0.3332	0.8391	0.7420	0.8630	0.3333	0.5608	0.8391	0.7743	0.8201	0.3333	0.3333
40	0.3332	0.8469	0.7549	0.8415	0.3333	0.5750	0.8469	0.7846	0.7857	0.3333	0.3333
45	0.3333	0.8537	0.7657	0.8119	0.3333	0.5874	0.8537	0.7934	0.7467	0.3333	0.3333
48	0.3333	0.8577	0.7720	0.7937	0.3333	0.5944	0.8577	0.7984	0.7287	0.3333	0.3333
50	0.3333	0.8602	0.7758	0.7733	0.3333	0.5989	0.8602	0.8016	0.7067	0.3333	0.3333

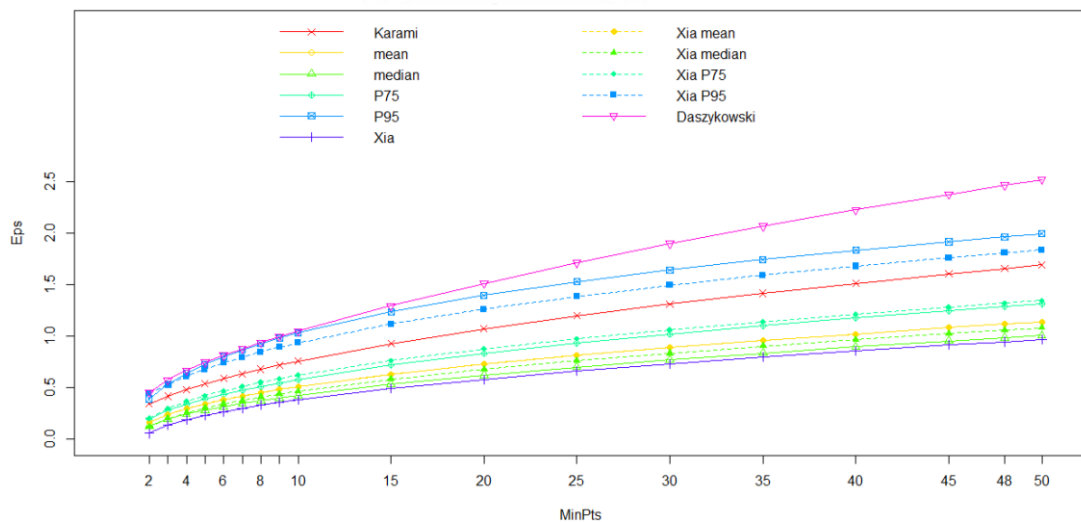
4.5.1.10 ขนาดตัวอย่าง $N=1,200$ กลุ่มละ 400, 400, 400 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

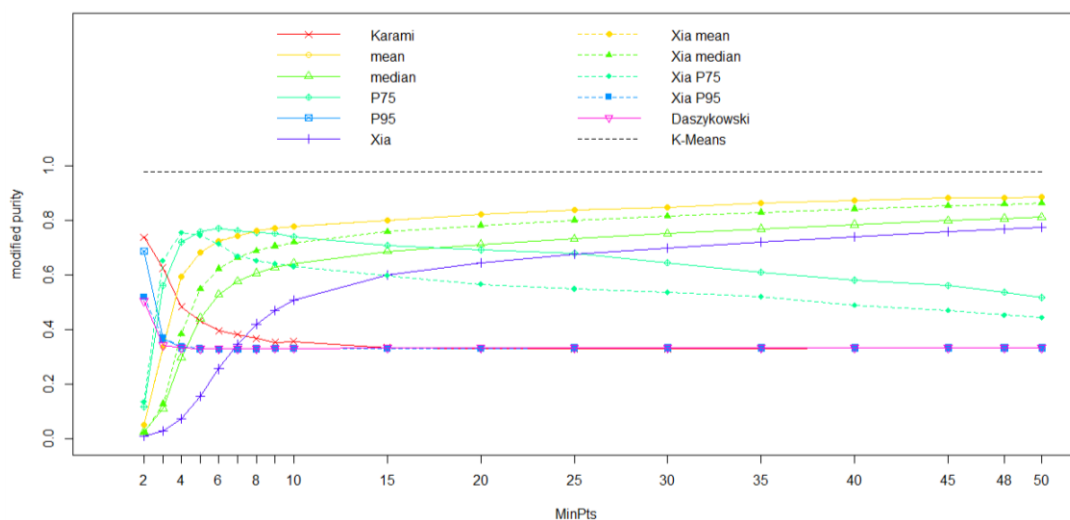
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



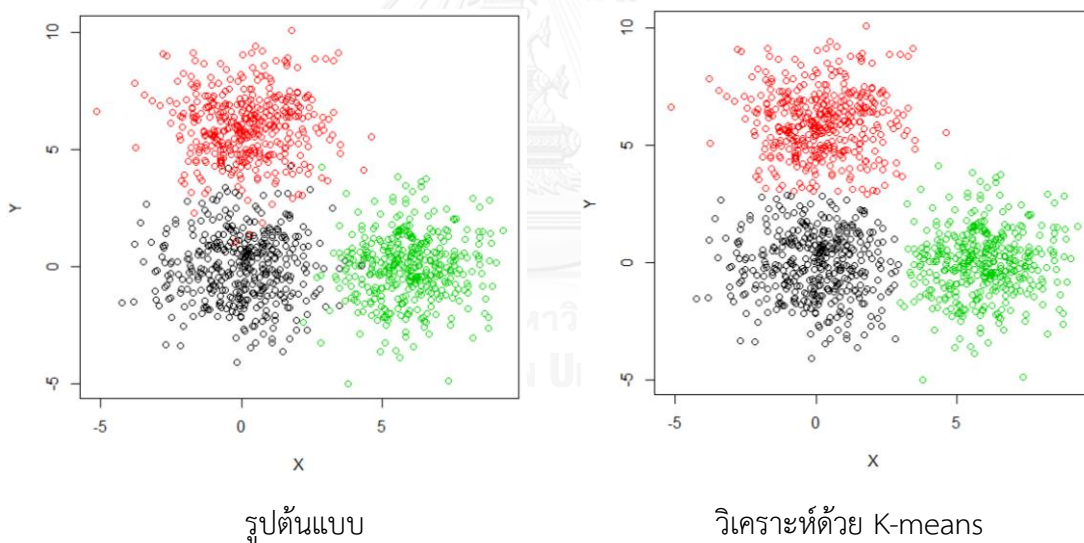
รูปที่ 4.5.1.10.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.1.10



รูปที่ 4.5.1.10.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.10



รูปที่ 4.5.1.10.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.10



รูปที่ 4.5.1.10.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.10

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.10.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี

Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.1.10.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-3 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.333 (400/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-3 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 4 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (400/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-5 เป็นต้นไป ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.333 (400/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 6 สำหรับวิธี P75 และสูงสุดที่ MinPts 4 สำหรับวิธี Xia P75 จากนั้น modified purity มีค่าลดลง

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-6 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 10-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.10.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.10.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9771767

ตารางที่ 4.5.1.10.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.10

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.3380	0.1581	0.1243	0.1920	0.3878	0.0531	0.1581	0.1108	0.1978	0.4460	0.4502
3	0.4139	0.2349	0.1909	0.2735	0.5302	0.1313	0.2349	0.1957	0.2955	0.5272	0.5715
4	0.4779	0.2916	0.2394	0.3348	0.6376	0.1854	0.2916	0.2529	0.3638	0.6068	0.6637
5	0.5344	0.3381	0.2789	0.3861	0.7245	0.2281	0.3381	0.2986	0.4191	0.6760	0.7417
6	0.5854	0.3785	0.3132	0.4311	0.7990	0.2645	0.3785	0.3379	0.4671	0.7375	0.8115
7	0.6323	0.4146	0.3441	0.4715	0.8656	0.2969	0.4146	0.3730	0.5098	0.7928	0.8754
8	0.6759	0.4475	0.3725	0.5086	0.9262	0.3265	0.4475	0.4049	0.5487	0.8430	0.9351
9	0.7169	0.4778	0.3986	0.5432	0.9801	0.3536	0.4778	0.4342	0.5843	0.8893	0.9916
10	0.7557	0.5061	0.4234	0.5756	1.0296	0.3789	0.5061	0.4615	0.6176	0.9326	1.0454
15	0.9255	0.6266	0.5295	0.7158	1.2349	0.4876	0.6266	0.5783	0.7588	1.1148	1.2912
20	1.0687	0.7249	0.6178	0.8309	1.3929	0.5772	0.7249	0.6738	0.8730	1.2603	1.5110
25	1.1948	0.8099	0.6957	0.9305	1.5264	0.6558	0.8099	0.7568	0.9712	1.3837	1.7115
30	1.3089	0.8858	0.7664	1.0192	1.6400	0.7268	0.8858	0.8312	1.0584	1.4918	1.8950
35	1.4138	0.9551	0.8318	1.1000	1.7406	0.7920	0.9551	0.8992	1.1376	1.5889	2.0654
40	1.5114	1.0192	0.8929	1.1744	1.8318	0.8529	1.0192	0.9623	1.2105	1.6775	2.2251
45	1.6031	1.0792	0.9505	1.2438	1.9140	0.9102	1.0792	1.0214	1.2784	1.7591	2.3765
48	1.6556	1.1135	0.9833	1.2829	1.9618	0.9432	1.1135	1.0553	1.3172	1.8055	2.4636
50	1.6898	1.1358	1.0048	1.3088	1.9923	0.9647	1.1358	1.0774	1.3424	1.8354	2.5206

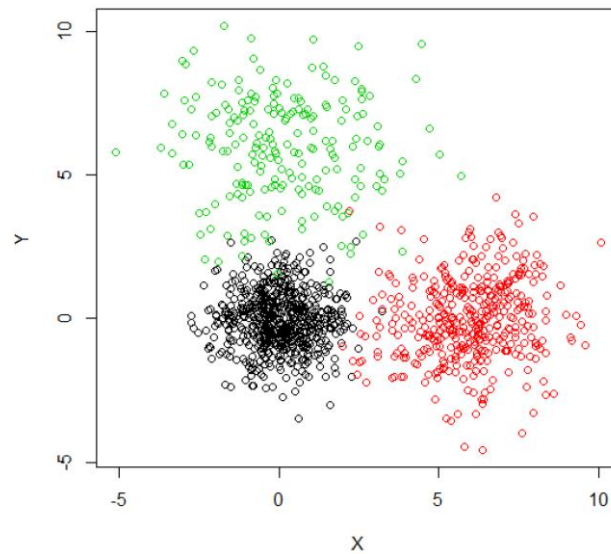
ตารางที่ 4.5.1.10.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี
ที่ 4.5.1.10

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7388	0.0486	0.0243	0.1162	0.6875	0.0082	0.0486	0.0192	0.1352	0.5196	0.5041
3	0.6255	0.3364	0.1109	0.5614	0.3681	0.0276	0.3364	0.1269	0.6514	0.3668	0.3425
4	0.4840	0.5926	0.2961	0.7208	0.3331	0.0729	0.5926	0.3827	0.7551	0.3383	0.3326
5	0.4328	0.6819	0.4414	0.7587	0.3295	0.1554	0.6819	0.5483	0.7465	0.3309	0.3296
6	0.3963	0.7227	0.5283	0.7710	0.3291	0.2553	0.7227	0.6226	0.7150	0.3288	0.3292
7	0.3816	0.7448	0.5775	0.7638	0.3295	0.3464	0.7448	0.6629	0.6694	0.3283	0.3295
8	0.3670	0.7613	0.6067	0.7548	0.3299	0.4204	0.7613	0.6887	0.6534	0.3287	0.3299
9	0.3525	0.7712	0.6277	0.7541	0.3303	0.4701	0.7712	0.7052	0.6417	0.3284	0.3304
10	0.3564	0.7774	0.6431	0.7416	0.3307	0.5070	0.7774	0.7193	0.6318	0.3288	0.3308
15	0.3340	0.8015	0.6872	0.7097	0.3319	0.6000	0.8015	0.7589	0.5983	0.3304	0.3322
20	0.3302	0.8218	0.7131	0.6924	0.3325	0.6448	0.8218	0.7819	0.5656	0.3313	0.3329
25	0.3294	0.8379	0.7334	0.6807	0.3328	0.6754	0.8379	0.8005	0.5491	0.3319	0.3332
30	0.3300	0.8495	0.7522	0.6455	0.3330	0.6999	0.8495	0.8159	0.5365	0.3323	0.3333
35	0.3309	0.8648	0.7693	0.6109	0.3331	0.7217	0.8648	0.8299	0.5204	0.3326	0.3333
40	0.3316	0.8727	0.7850	0.5811	0.3332	0.7412	0.8727	0.8417	0.4897	0.3328	0.3333
45	0.3320	0.8819	0.7992	0.5607	0.3332	0.7585	0.8819	0.8531	0.4704	0.3329	0.3333
48	0.3322	0.8837	0.8068	0.5379	0.3332	0.7684	0.8837	0.8592	0.4529	0.3330	0.3333
50	0.3324	0.8877	0.8121	0.5173	0.3333	0.7745	0.8877	0.8631	0.4428	0.3330	0.3333

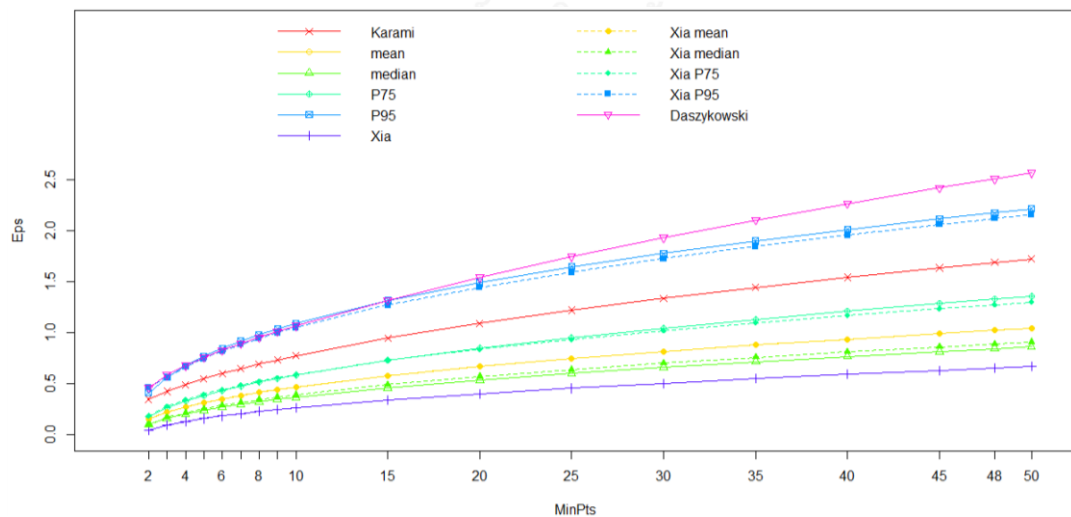
4.5.1.11 ขนาดตัวอย่าง $N=1,200$ กลุ่มละ 600, 400, 200 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

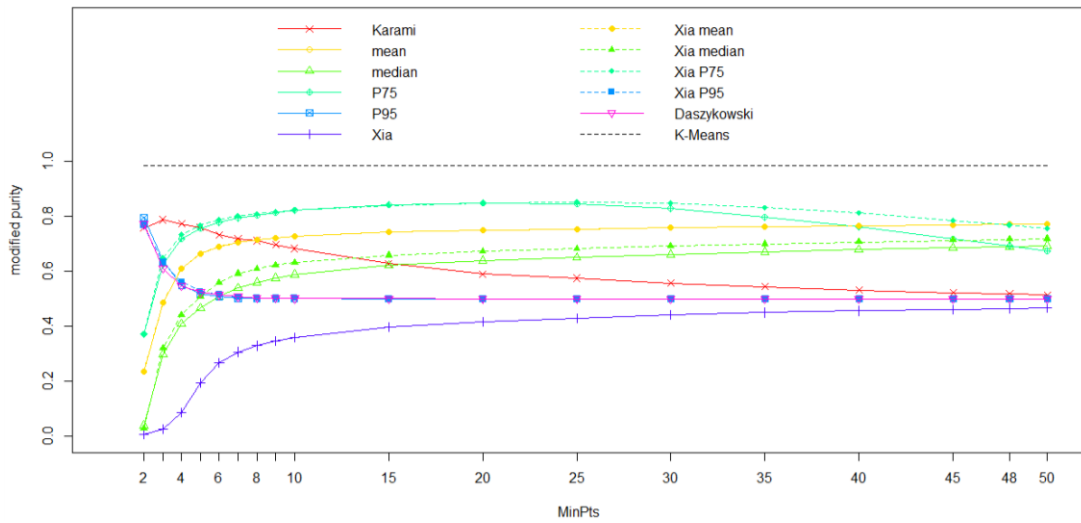
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



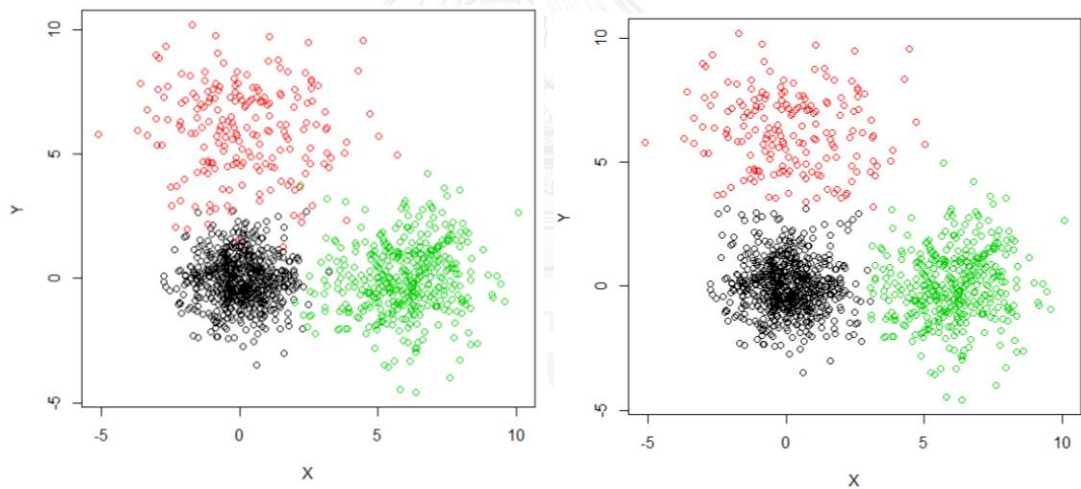
รูปที่ 4.5.1.11.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.1.11



รูปที่ 4.5.1.11.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.11



รูปที่ 4.5.1.11.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.11



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.11.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.5.1.11

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.11.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี

ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี P75, วิธี Xia P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.1.11.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่อู่เข้าสู่ 0.5 (600/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 6 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.5 (600/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 3 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 50 เป็นต้นไป ค่า modified purity มีแนวโน้มที่จะลู่อู่เข้าสู่ 0.5 (600/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 25 จากนั้น modified purity มีค่าลดลง

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 6-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง และค่า modified purity มีแนวโน้มที่จะลู่อู่เข้าสู่ 0.5 (600/1200)

จากรูปที่ 4.5.1.11.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.11.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.984385

ตารางที่ 4.5.1.11.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.11

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.3447	0.1466	0.1016	0.1790	0.4087	0.0370	0.1466	0.0919	0.1790	0.4571	0.4582
3	0.4221	0.2172	0.1586	0.2641	0.5593	0.0895	0.2172	0.1639	0.2743	0.5640	0.5819
4	0.4874	0.2692	0.2008	0.3288	0.6735	0.1259	0.2692	0.2119	0.3404	0.6608	0.6757
5	0.5450	0.3121	0.2359	0.3823	0.7654	0.1552	0.3121	0.2505	0.3943	0.7441	0.7551
6	0.5970	0.3492	0.2664	0.4299	0.8444	0.1804	0.3492	0.2838	0.4408	0.8170	0.8261
7	0.6448	0.3823	0.2938	0.4726	0.9148	0.2028	0.3823	0.3133	0.4822	0.8827	0.8912
8	0.6893	0.4124	0.3186	0.5115	0.9779	0.2231	0.4124	0.3401	0.5198	0.9424	0.9520
9	0.7311	0.4402	0.3417	0.5478	1.0358	0.2418	0.4402	0.3647	0.5545	0.9978	1.0094
10	0.7707	0.4661	0.3633	0.5820	1.0889	0.2593	0.4661	0.3877	0.5867	1.0490	1.0646
15	0.9439	0.5761	0.4558	0.7279	1.3121	0.3345	0.5761	0.4857	0.7235	1.2661	1.3147
20	1.0899	0.6658	0.5320	0.8474	1.4904	0.3967	0.6658	0.5659	0.8348	1.4415	1.5388
25	1.2186	0.7434	0.5986	0.9510	1.6442	0.4514	0.7434	0.6355	0.9308	1.5916	1.7425
30	1.3349	0.8127	0.6588	1.0441	1.7772	0.5007	0.8127	0.6979	1.0165	1.7248	1.9294
35	1.4418	0.8761	0.7138	1.1297	1.8981	0.5461	0.8761	0.7550	1.0947	1.8457	2.1028
40	1.5414	0.9347	0.7651	1.2090	2.0104	0.5883	0.9347	0.8080	1.1670	1.9573	2.2655
45	1.6349	0.9897	0.8132	1.2839	2.1148	0.6282	0.9897	0.8577	1.2347	2.0614	2.4197
48	1.6885	1.0212	0.8411	1.3267	2.1741	0.6510	1.0212	0.8863	1.2735	2.1211	2.5084
50	1.7233	1.0417	0.8589	1.3546	2.2129	0.6659	1.0417	0.9048	1.2988	2.1599	2.5664



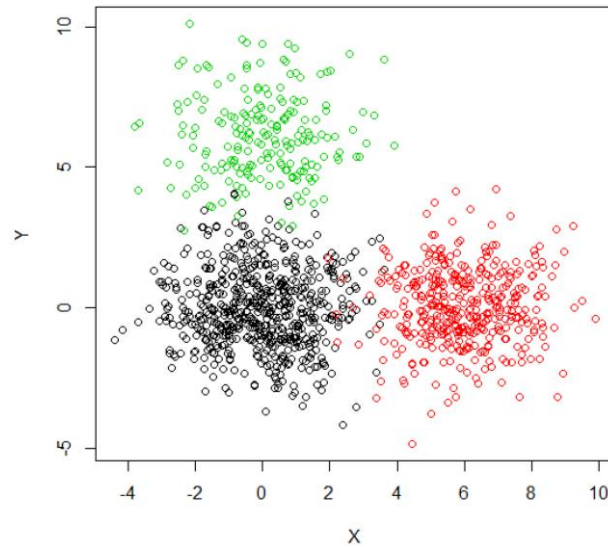
ตารางที่ 4.5.1.11.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี
ที่ 4.5.1.11

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7564	0.2356	0.0376	0.3706	0.7967	0.0070	0.2356	0.0262	0.3715	0.7666	0.7720
3	0.7881	0.4868	0.2971	0.6252	0.6377	0.0235	0.4868	0.3187	0.6490	0.6317	0.6102
4	0.7732	0.6095	0.4081	0.7183	0.5478	0.0855	0.6095	0.4399	0.7344	0.5609	0.5469
5	0.7597	0.6639	0.4656	0.7567	0.5165	0.1943	0.6639	0.5088	0.7680	0.5261	0.5240
6	0.7323	0.6890	0.5064	0.7782	0.5074	0.2664	0.6890	0.5575	0.7888	0.5157	0.5143
7	0.7183	0.7043	0.5389	0.7939	0.5019	0.3058	0.7043	0.5907	0.8011	0.5060	0.5081
8	0.7117	0.7145	0.5595	0.8045	0.5004	0.3284	0.7145	0.6085	0.8089	0.5028	0.5029
9	0.6965	0.7216	0.5755	0.8142	0.5007	0.3459	0.7216	0.6216	0.8170	0.5032	0.5029
10	0.6828	0.7277	0.5875	0.8215	0.5006	0.3585	0.7277	0.6311	0.8239	0.5024	0.5016
15	0.6273	0.7431	0.6209	0.8423	0.4998	0.3958	0.7431	0.6576	0.8391	0.5004	0.5001
20	0.5891	0.7497	0.6384	0.8489	0.4999	0.4158	0.7497	0.6719	0.8482	0.4999	0.4999
25	0.5749	0.7537	0.6507	0.8463	0.5000	0.4293	0.7537	0.6820	0.8514	0.5000	0.5000
30	0.5562	0.7577	0.6607	0.8290	0.5000	0.4409	0.7577	0.6910	0.8481	0.5000	0.5000
35	0.5438	0.7611	0.6702	0.7958	0.5000	0.4492	0.7611	0.6993	0.8315	0.5000	0.5000
40	0.5313	0.7653	0.6782	0.7623	0.5000	0.4557	0.7653	0.7064	0.8127	0.5000	0.5000
45	0.5215	0.7692	0.6856	0.7173	0.5000	0.4614	0.7692	0.7128	0.7854	0.5000	0.5000
48	0.5187	0.7715	0.6897	0.6939	0.5000	0.4635	0.7715	0.7163	0.7687	0.5000	0.5000
50	0.5136	0.7730	0.6922	0.6749	0.5000	0.4658	0.7730	0.7184	0.7560	0.5000	0.5000

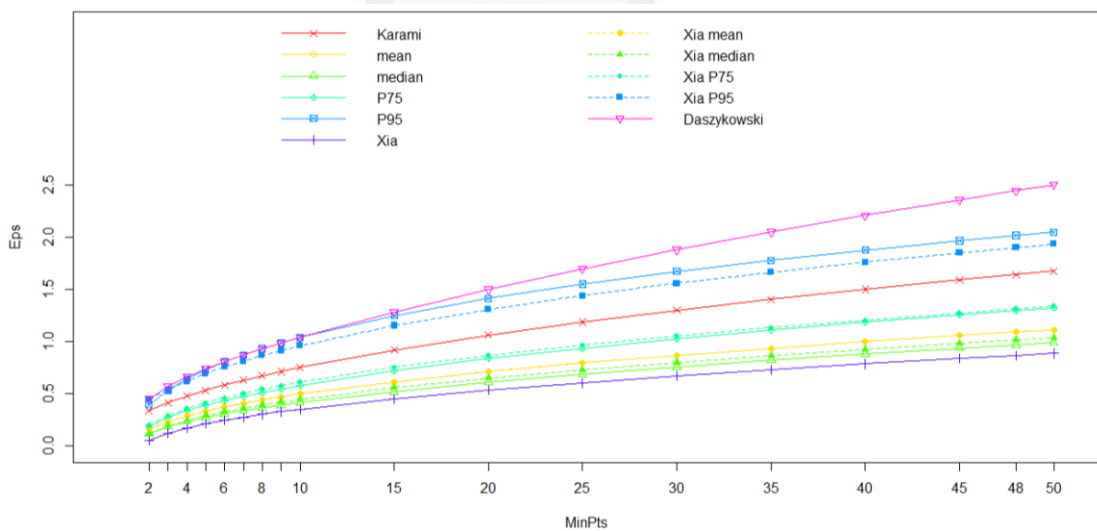
4.5.1.12 ขนาดตัวอย่าง $N=1,200$ กลุ่มละ 600, 400, 200 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$

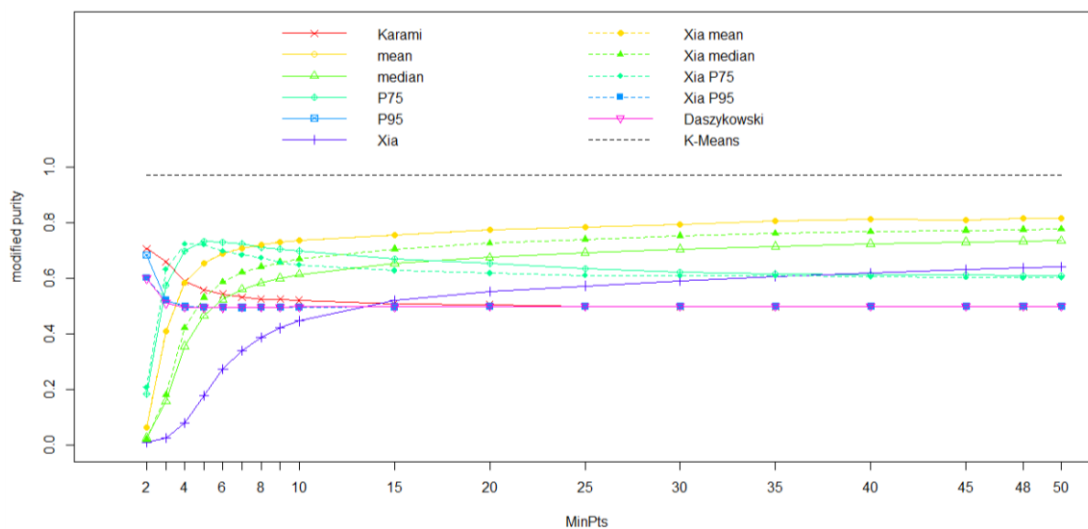
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



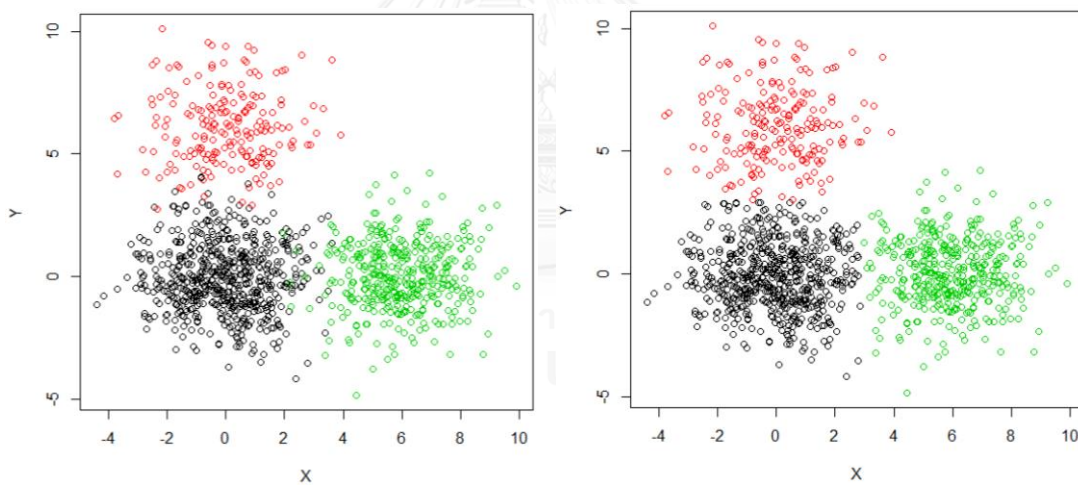
รูปที่ 4.5.1.12.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.1.12



รูปที่ 4.5.1.12.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.1.12



รูปที่ 4.5.1.12.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.1.12



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.1.12.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.1.12

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.1.12.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี

ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.1.12.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-3 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.5 (600/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-3 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นจนกระทั่งที่ MinPts 4 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (600/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-5 เป็นต้นไป ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.5 (600/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 5 สำหรับวิธี P75 และที่ MinPts 4 สำหรับวิธี Xia P75 จากนั้น modified purity มีค่าลดลง

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 15-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.1.12.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.1.12.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9735383

ตารางที่ 4.5.1.12.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.1.12

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.3351	0.1545	0.1186	0.1879	0.3887	0.0489	0.1545	0.1059	0.1925	0.4456	0.4467
3	0.4104	0.2294	0.1826	0.2702	0.5319	0.1204	0.2294	0.1873	0.2894	0.5323	0.5673
4	0.4739	0.2847	0.2296	0.3327	0.6404	0.1699	0.2847	0.2420	0.3569	0.6159	0.6586
5	0.5298	0.3303	0.2688	0.3852	0.7269	0.2090	0.3303	0.2860	0.4121	0.6891	0.7360
6	0.5804	0.3698	0.3026	0.4307	0.8027	0.2427	0.3698	0.3238	0.4598	0.7533	0.8051
7	0.6269	0.4050	0.3332	0.4718	0.8702	0.2724	0.4050	0.3573	0.5022	0.8114	0.8685
8	0.6701	0.4370	0.3611	0.5091	0.9306	0.2993	0.4370	0.3877	0.5408	0.8644	0.9278
9	0.7108	0.4666	0.3875	0.5440	0.9855	0.3242	0.4666	0.4158	0.5762	0.9130	0.9839
10	0.7492	0.4942	0.4118	0.5769	1.0368	0.3475	0.4942	0.4421	0.6094	0.9586	1.0375
15	0.9176	0.6118	0.5174	0.7172	1.2474	0.4471	0.6118	0.5538	0.7498	1.1507	1.2813
20	1.0596	0.7079	0.6058	0.8337	1.4111	0.5295	0.7079	0.6453	0.8640	1.3054	1.4996
25	1.1847	0.7911	0.6835	0.9343	1.5493	0.6016	0.7911	0.7250	0.9626	1.4379	1.6978
30	1.2977	0.8656	0.7534	1.0243	1.6693	0.6667	0.8656	0.7963	1.0505	1.5550	1.8801
35	1.4017	0.9336	0.8178	1.1069	1.7763	0.7264	0.9336	0.8616	1.1304	1.6608	2.0492
40	1.4985	0.9966	0.8782	1.1831	1.8755	0.7822	0.9966	0.9222	1.2043	1.7579	2.2078
45	1.5894	1.0556	0.9342	1.2544	1.9647	0.8347	1.0556	0.9791	1.2734	1.8483	2.3579
48	1.6415	1.0894	0.9668	1.2950	2.0161	0.8650	1.0894	1.0117	1.3129	1.8998	2.4446
50	1.6754	1.1113	0.9874	1.3219	2.0490	0.8846	1.1113	1.0329	1.3386	1.9332	2.5011



ตารางที่ 4.5.1.12. 2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี
ที่ 4.5.1.12

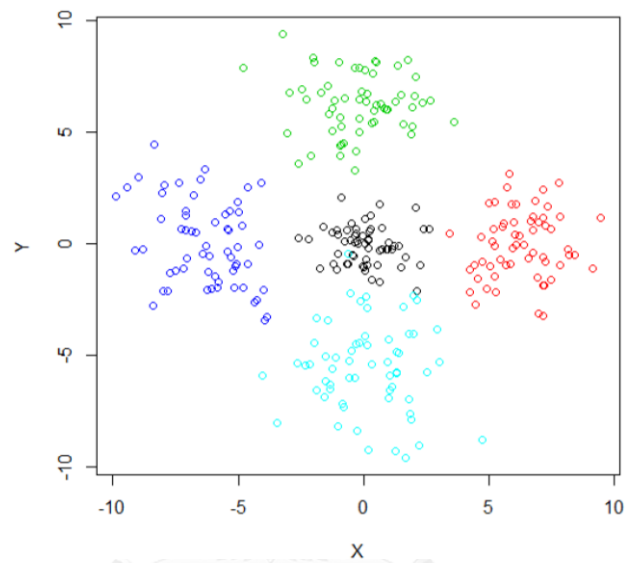
MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7068	0.0636	0.0245	0.1829	0.6833	0.0076	0.0636	0.0190	0.2076	0.6006	0.5994
3	0.6593	0.4094	0.1567	0.5752	0.5220	0.0254	0.4094	0.1799	0.6330	0.5203	0.5106
4	0.5902	0.5836	0.3555	0.6975	0.4987	0.0783	0.5836	0.4211	0.7246	0.5012	0.4968
5	0.5574	0.6531	0.4653	0.7329	0.4958	0.1778	0.6531	0.5309	0.7219	0.4968	0.4955
6	0.5429	0.6889	0.5246	0.7293	0.4956	0.2729	0.6889	0.5867	0.6992	0.4952	0.4955
7	0.5329	0.7077	0.5598	0.7253	0.4955	0.3401	0.7077	0.6208	0.6857	0.4949	0.4959
8	0.5254	0.7208	0.5834	0.7113	0.4960	0.3868	0.7208	0.6414	0.6744	0.4954	0.4959
9	0.5245	0.7312	0.6001	0.7064	0.4965	0.4207	0.7312	0.6580	0.6584	0.4954	0.4964
10	0.5209	0.7384	0.6140	0.6990	0.4969	0.4462	0.7384	0.6697	0.6492	0.4955	0.4969
15	0.5090	0.7575	0.6541	0.6686	0.4982	0.5197	0.7575	0.7056	0.6289	0.4972	0.4985
20	0.5048	0.7743	0.6759	0.6532	0.4989	0.5515	0.7743	0.7265	0.6193	0.4981	0.4993
25	0.4997	0.7841	0.6931	0.6342	0.4992	0.5731	0.7841	0.7411	0.6112	0.4987	0.4996
30	0.4984	0.7956	0.7059	0.6215	0.4994	0.5900	0.7956	0.7526	0.6101	0.4990	0.4998
35	0.4982	0.8059	0.7156	0.6162	0.4996	0.6053	0.8059	0.7615	0.6096	0.4993	0.4999
40	0.4988	0.8119	0.7246	0.6123	0.4997	0.6189	0.8119	0.7679	0.6063	0.4995	0.5000
45	0.4991	0.8091	0.7314	0.6126	0.4997	0.6309	0.8092	0.7726	0.6046	0.4996	0.5000
48	0.4986	0.8161	0.7341	0.6099	0.4998	0.6379	0.8161	0.7753	0.6027	0.4996	0.5000
50	0.4991	0.8164	0.7359	0.6089	0.4998	0.6423	0.8164	0.7769	0.6028	0.4997	0.5000

4.5.2 กรณีจำนวนกลุ่มเท่ากับ 5

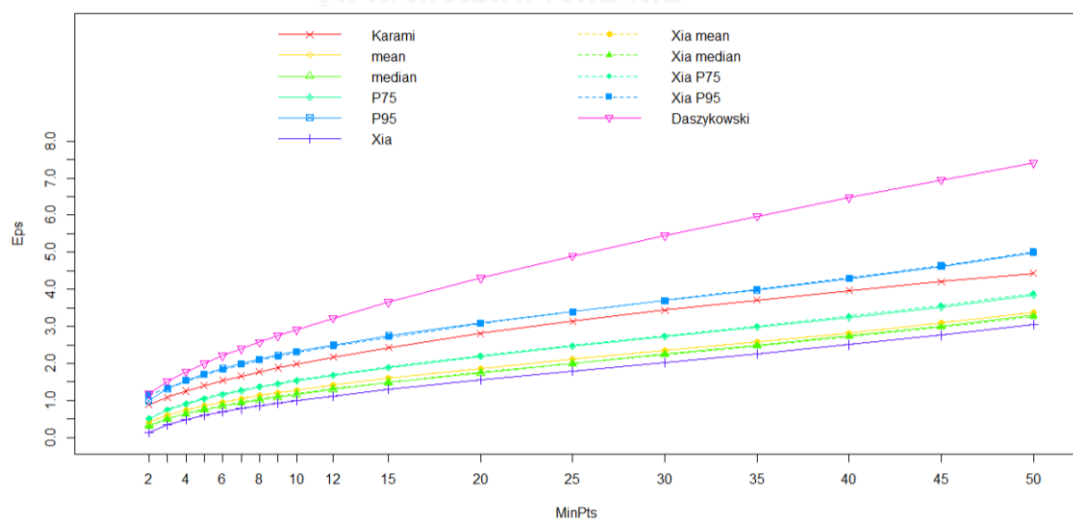
4.5.2.1 ขนาดตัวอย่าง $N=300$ กลุ่มละ 60, 60, 60, 60, 60 ตามลำดับ

$$\text{โดยที่ } \mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$$

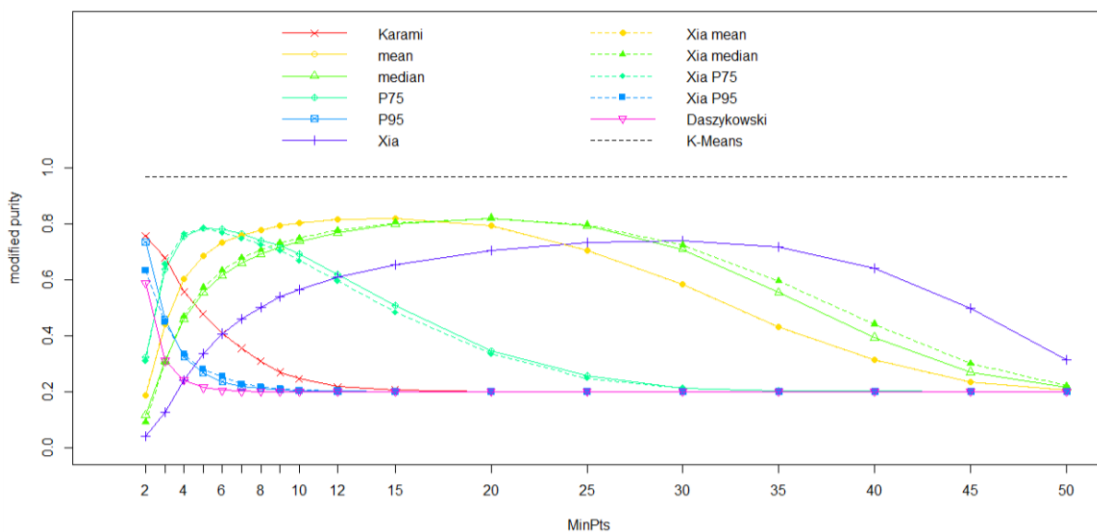
$$\text{และมีความแปรปรวนไม่เท่ากัน } \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_4 = \Sigma_5 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$



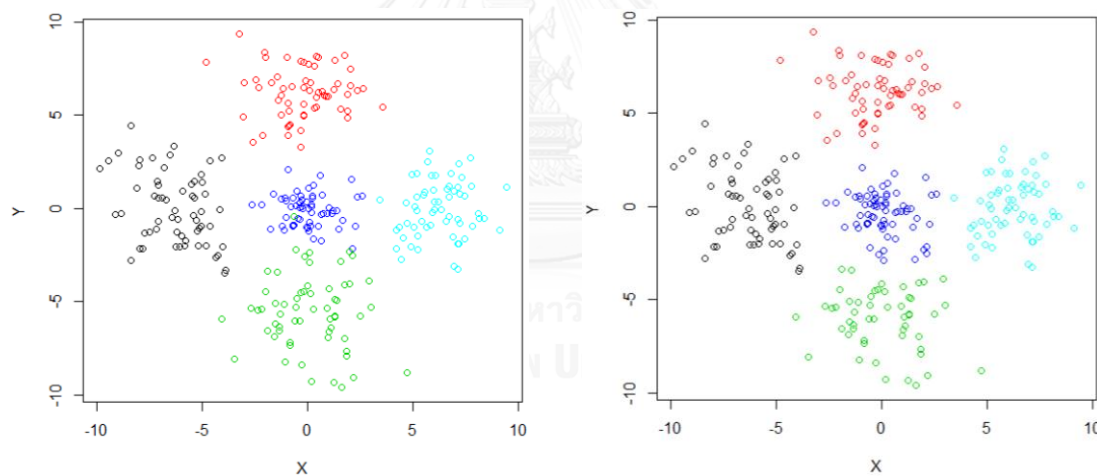
รูปที่ 4.5.2.1.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.1



รูปที่ 4.5.2.1.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.1



รูปที่ 4.5.2.1.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.2.1



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.1.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.1

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.1.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ ซึ่งวิธี Daszykowski ตั้งแต่ MinPts 20 เป็นต้นไปให้ค่า Eps สูงกว่าวิธีอื่น ๆ เป็นอย่างมาก

จากรูปที่ 4.5.2.1.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.2 (60/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-5 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 9 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.2 (60/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-6 เป็นต้นไป ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.2 (60/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 5 จากนั้น modified purity มีค่าลดลง จนกระทั่งที่ MinPts 35 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.2 (60/300)

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-6 แล้วค่อย ๆ เพิ่มสูงขึ้น จากนั้นลดลงอย่างรวดเร็วในช่วง MinPts 20 เป็นต้นมา และที่ค่า MinPts ตั้งแต่ 50 ค่า modified purity มีแนวโน้มลู่เข้าสู่ 0.2

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง จนกระทั่งสูงสุดที่ MinPts 30 จากนั้น modified purity มีค่าลดลง

จากรูปที่ 4.5.2.1.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.1.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9699933

ตารางที่ 4.5.2.1.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.1

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.8869	0.4067	0.3280	0.5174	0.9923	0.1412	0.4067	0.2870	0.5080	1.1375	1.1820
3	1.0862	0.5997	0.5083	0.7369	1.3123	0.3423	0.5997	0.5025	0.7535	1.3319	1.5169
4	1.2542	0.7414	0.6410	0.9007	1.5355	0.4819	0.7414	0.6471	0.9224	1.5206	1.7754
5	1.4022	0.8581	0.7509	1.0391	1.7140	0.5940	0.8581	0.7637	1.0599	1.6838	1.9990
6	1.5361	0.9590	0.8474	1.1553	1.8606	0.6895	0.9590	0.8636	1.1781	1.8274	2.2029
7	1.6592	1.0494	0.9352	1.2618	1.9961	0.7749	1.0494	0.9528	1.2835	1.9563	2.3933
8	1.7737	1.1321	1.0168	1.3579	2.1118	0.8535	1.1321	1.0346	1.3794	2.0731	2.5726
9	1.8813	1.2086	1.0914	1.4475	2.2208	0.9260	1.2086	1.1100	1.4679	2.1811	2.7445
10	1.9831	1.2805	1.1623	1.5309	2.3202	0.9943	1.2805	1.1809	1.5507	2.2818	2.9092
12	2.1723	1.4130	1.2944	1.6842	2.4997	1.1211	1.4130	1.3119	1.7027	2.4652	3.2221
15	2.4288	1.5926	1.4753	1.8886	2.7439	1.2940	1.5926	1.4897	1.9074	2.7099	3.6544
20	2.8045	1.8603	1.7492	2.1903	3.0853	1.5532	1.8603	1.7550	2.2103	3.0688	4.3064
25	3.1355	2.1067	2.0046	2.4579	3.3935	1.7941	2.1067	2.0000	2.4867	3.3910	4.8990
30	3.4348	2.3441	2.2535	2.7138	3.6907	2.0275	2.3441	2.2363	2.7511	3.6960	5.4505
35	3.7100	2.5810	2.5002	2.9698	3.9815	2.2609	2.5810	2.4722	3.0136	3.9971	5.9725
40	3.9661	2.8252	2.7507	3.2324	4.2872	2.5021	2.8252	2.7156	3.2830	4.3042	6.4722
45	4.2067	3.0860	3.0104	3.5154	4.6159	2.7585	3.0860	2.9751	3.5705	4.6332	6.9504
50	4.4343	3.3789	3.2961	3.8419	4.9912	3.0433	3.3788	3.2653	3.8946	5.0098	7.4097

ตารางที่ 4.5.2.1.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี

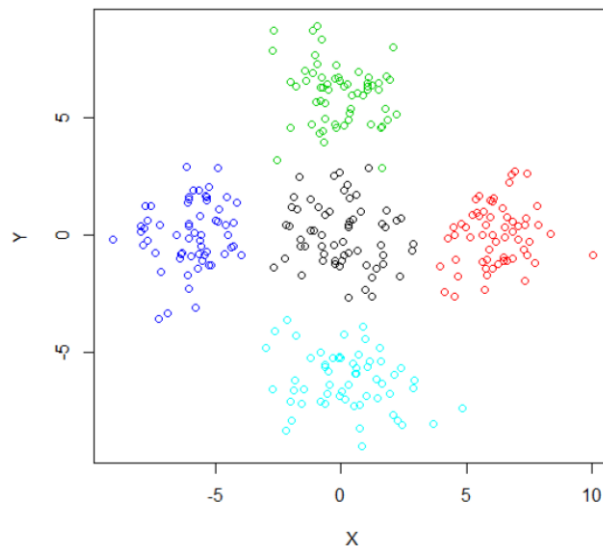
4.5.2.1

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7567	0.1857	0.1155	0.3223	0.7365	0.0403	0.1857	0.0905	0.3107	0.6327	0.5895
3	0.6791	0.4422	0.3094	0.6377	0.4592	0.1258	0.4422	0.3034	0.6577	0.4481	0.3110
4	0.5581	0.6034	0.4590	0.7561	0.3247	0.2406	0.6034	0.4683	0.7663	0.3337	0.2415
5	0.4767	0.6869	0.5547	0.7861	0.2673	0.3364	0.6870	0.5717	0.7853	0.2804	0.2145
6	0.4083	0.7326	0.6169	0.7829	0.2359	0.4077	0.7326	0.6345	0.7675	0.2548	0.2044
7	0.3555	0.7600	0.6606	0.7648	0.2180	0.4607	0.7600	0.6780	0.7490	0.2276	0.2016
8	0.3081	0.7795	0.6926	0.7411	0.2126	0.5013	0.7795	0.7071	0.7265	0.2176	0.2004
9	0.2694	0.7943	0.7176	0.7230	0.2071	0.5387	0.7943	0.7316	0.7039	0.2090	0.2004
10	0.2456	0.8031	0.7382	0.6926	0.2036	0.5661	0.8032	0.7485	0.6696	0.2051	0.2000
12	0.2171	0.8174	0.7678	0.6200	0.2016	0.6093	0.8174	0.7774	0.5972	0.2024	0.2000
15	0.2055	0.8206	0.8005	0.5087	0.2004	0.6549	0.8206	0.8053	0.4851	0.2004	0.2000
20	0.2008	0.7931	0.8207	0.3452	0.2000	0.7046	0.7931	0.8190	0.3346	0.2000	0.2000
25	0.2000	0.7065	0.7949	0.2574	0.2000	0.7349	0.7065	0.7982	0.2479	0.2000	0.2000
30	0.2000	0.5842	0.7094	0.2135	0.2000	0.7413	0.5840	0.7240	0.2117	0.2000	0.2000
35	0.2000	0.4328	0.5544	0.2032	0.2000	0.7188	0.4328	0.5955	0.2017	0.2000	0.2000
40	0.2000	0.3130	0.3944	0.2011	0.2000	0.6404	0.3130	0.4412	0.2011	0.2000	0.2000
45	0.2000	0.2356	0.2699	0.2000	0.2000	0.4995	0.2356	0.3002	0.2000	0.2000	0.2000
50	0.2000	0.2065	0.2142	0.2000	0.2000	0.3144	0.2065	0.2213	0.2000	0.2000	0.2000

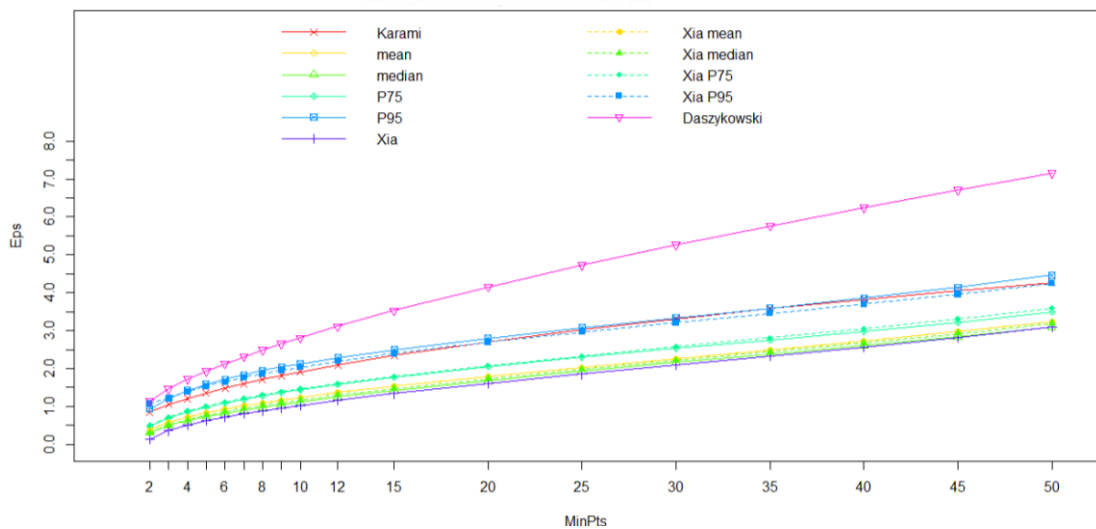
4.5.2.2 ขนาดตัวอย่าง N=300 กลุ่มละ 60, 60, 60, 60, 60 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

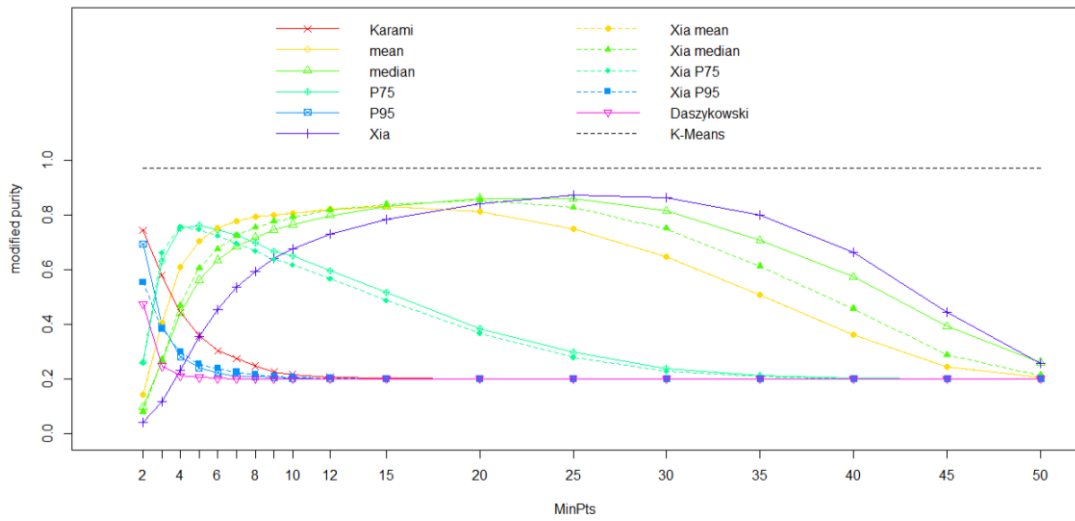
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



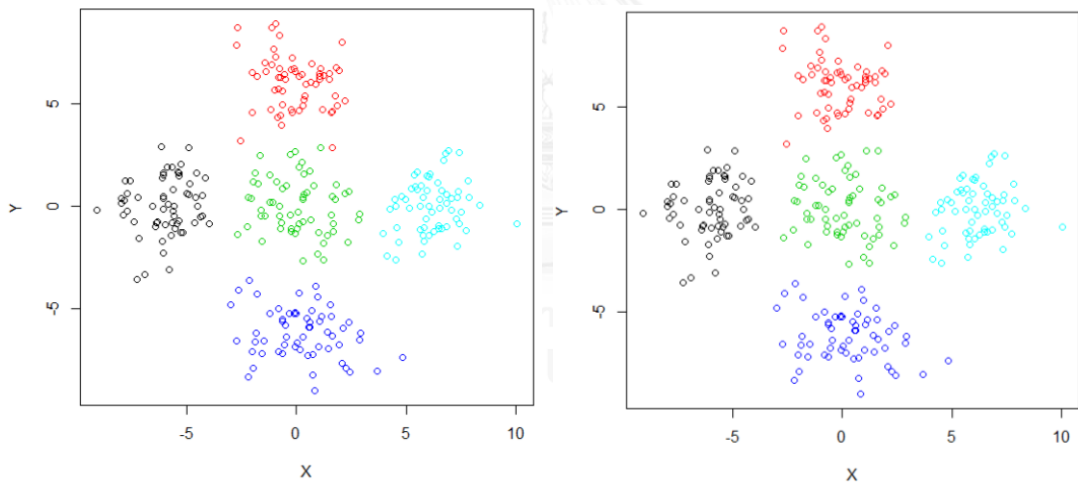
รูปที่ 4.5.2.2.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.2



รูปที่ 4.5.2.2.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.2



รูปที่ 4.5.2.2.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.2.2



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.2.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.2

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.2.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski กลุ่มที่ประกอบด้วยวิธี P95, วิธี Xia P95 และวิธี Karami กลุ่มที่ประกอบด้วยวิธี P75 และวิธี Xia P75 กลุ่มที่ประกอบด้วยวิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ ซึ่งวิธี Daszykowski ตั้งแต่ MinPts 20 เป็นต้นไปให้ค่า Eps สูงกว่าวิธีอื่น ๆ เป็นอย่างมาก

จากรูปที่ 4.5.2.2.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-3 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.2 (60/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 8 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.2 (60/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-6 เป็นต้นไป ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 12 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.2 (60/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 5 สำหรับวิธี P75 และที่ MinPts 4 สำหรับวิธี Xia P75 จากนั้น modified purity มีค่าลดลง จนกระทั่งที่ MinPts 35 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.2 (60/300)

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-6 แล้วค่อย ๆ เพิ่มสูงขึ้น จากนั้นลดลงอย่างรวดเร็วในช่วง MinPts 20 เป็นต้นมา และที่ค่า MinPts ตั้งแต่ 50 ค่า modified purity มีแนวโน้มลู่เข้าสู่ 0.2

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง จนกระทั่งสูงสุดที่ MinPts 25 จากนั้น modified purity มีค่าลดลง

จากรูปที่ 4.5.2.2.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.2.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9714667

ตารางที่ 4.5.2.2.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.2

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.8544	0.3909	0.3235	0.4927	0.9330	0.1449	0.3909	0.2826	0.4899	1.0649	1.1409
3	1.0465	0.5772	0.4973	0.7000	1.2127	0.3564	0.5772	0.4957	0.7213	1.2236	1.4640
4	1.2084	0.7141	0.6248	0.8541	1.4190	0.5023	0.7141	0.6387	0.8800	1.3842	1.7137
5	1.3510	0.8260	0.7287	0.9813	1.5777	0.6178	0.8260	0.7530	1.0079	1.5232	1.9303
6	1.4799	0.9230	0.8207	1.0917	1.7135	0.7164	0.9230	0.8511	1.1179	1.6454	2.1266
7	1.5985	1.0098	0.9035	1.1901	1.8329	0.8044	1.0098	0.9387	1.2156	1.7545	2.3103
8	1.7089	1.0895	0.9790	1.2792	1.9414	0.8851	1.0895	1.0190	1.3049	1.8542	2.4845
9	1.8125	1.1636	1.0501	1.3643	2.0359	0.9600	1.1636	1.0936	1.3875	1.9464	2.6493
10	1.9106	1.2331	1.1184	1.4415	2.1255	1.0304	1.2331	1.1636	1.4647	2.0321	2.8085
12	2.0929	1.3608	1.2441	1.5847	2.2849	1.1596	1.3608	1.2921	1.6058	2.1885	3.1095
15	2.3400	1.5340	1.4159	1.7745	2.4978	1.3362	1.5339	1.4667	1.7953	2.3950	3.5278
20	2.7020	1.7930	1.6760	2.0523	2.8004	1.6019	1.7930	1.7283	2.0752	2.6946	4.1563
25	3.0209	2.0317	1.9194	2.2992	3.0727	1.8485	2.0318	1.9699	2.3297	2.9610	4.7270
30	3.3092	2.2608	2.1525	2.5276	3.3348	2.0858	2.2608	2.2019	2.5711	3.2094	5.2593
35	3.5744	2.4884	2.3807	2.7513	3.5898	2.3223	2.4884	2.4325	2.8085	3.4498	5.7622
40	3.8211	2.7214	2.6091	2.9768	3.8538	2.5642	2.7214	2.6686	3.0498	3.6923	6.2439
45	4.0529	2.9688	2.8419	3.2196	4.1385	2.8196	2.9688	2.9187	3.3053	3.9502	6.7065
50	4.2722	3.2437	3.0894	3.4959	4.4659	3.0991	3.2437	3.1952	3.5918	4.2475	7.1489

ตารางที่ 4.5.2.2.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

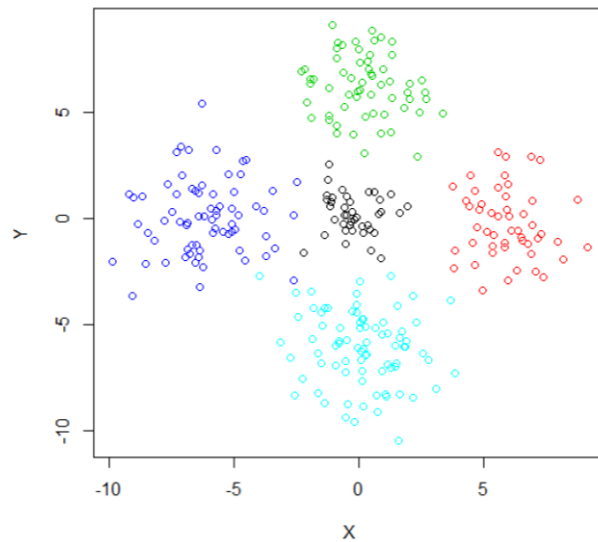
4.5.2.2

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7446	0.1435	0.0969	0.2612	0.6944	0.0404	0.1435	0.0784	0.2576	0.5549	0.4750
3	0.5791	0.4051	0.2671	0.6306	0.3951	0.1177	0.4051	0.2660	0.6631	0.3839	0.2470
4	0.4439	0.6083	0.4412	0.7525	0.2811	0.2327	0.6083	0.4690	0.7582	0.2988	0.2109
5	0.3583	0.7048	0.5628	0.7640	0.2406	0.3561	0.7048	0.6057	0.7474	0.2541	0.2051
6	0.3030	0.7519	0.6355	0.7461	0.2221	0.4539	0.7519	0.6761	0.7254	0.2382	0.2012
7	0.2743	0.7789	0.6855	0.7284	0.2098	0.5351	0.7790	0.7240	0.6965	0.2249	0.2004
8	0.2471	0.7928	0.7194	0.6976	0.2082	0.5949	0.7928	0.7553	0.6687	0.2152	0.2000
9	0.2262	0.7996	0.7473	0.6675	0.2043	0.6417	0.7996	0.7766	0.6373	0.2098	0.2000
10	0.2159	0.8079	0.7661	0.6514	0.2028	0.6770	0.8079	0.7921	0.6188	0.2070	0.2000
12	0.2059	0.8210	0.7984	0.5957	0.2008	0.7315	0.8213	0.8183	0.5672	0.2032	0.2000
15	0.2011	0.8312	0.8314	0.5171	0.2000	0.7843	0.8312	0.8387	0.4881	0.2000	0.2000
20	0.2000	0.8129	0.8625	0.3840	0.2000	0.8431	0.8133	0.8544	0.3681	0.2000	0.2000
25	0.2000	0.7492	0.8602	0.2987	0.2000	0.8723	0.7492	0.8277	0.2792	0.2000	0.2000
30	0.2000	0.6481	0.8154	0.2386	0.2000	0.8634	0.6481	0.7507	0.2268	0.2000	0.2000
35	0.2000	0.5077	0.7074	0.2120	0.2000	0.8005	0.5070	0.6137	0.2075	0.2000	0.2000
40	0.2000	0.3604	0.5744	0.2022	0.2000	0.6629	0.3604	0.4575	0.2004	0.2000	0.2000
45	0.2000	0.2431	0.3924	0.2000	0.2000	0.4434	0.2431	0.2877	0.2000	0.2000	0.2000
50	0.2000	0.2063	0.2597	0.2000	0.2000	0.2579	0.2063	0.2132	0.2000	0.2000	0.2000

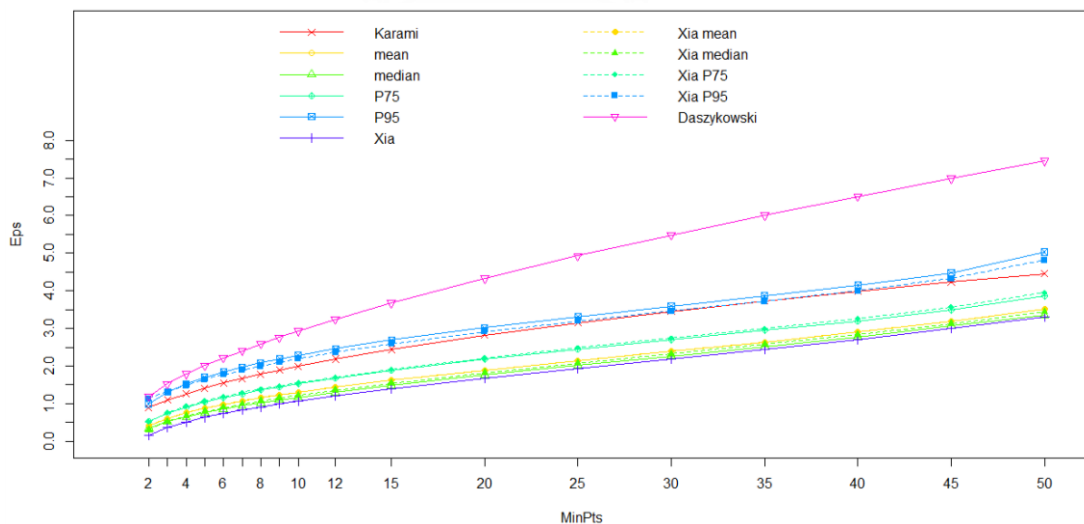
4.5.2.3 ขนาดตัวอย่าง N=300 กลุ่มละ 40, 50, 60, 70, 80 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

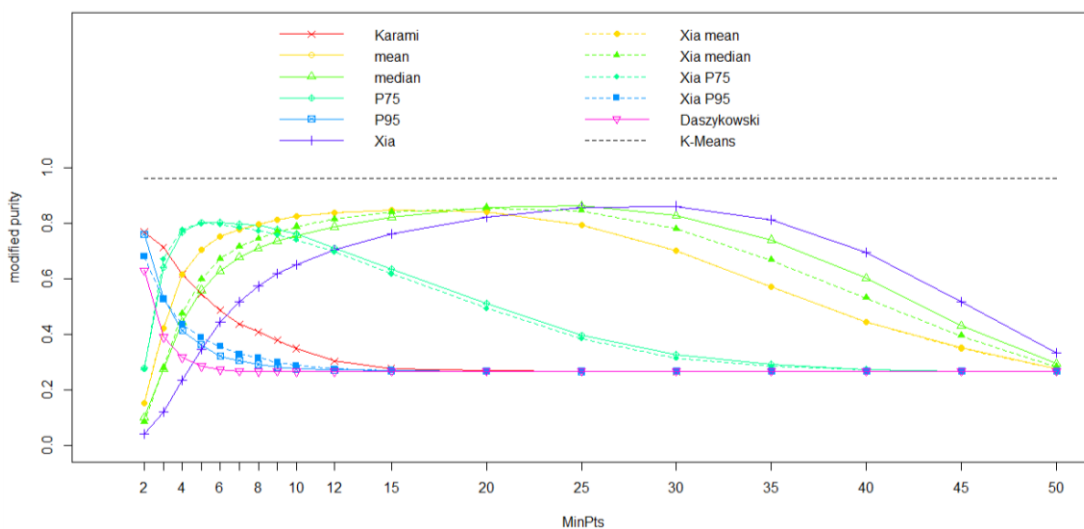
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_4 = \Sigma_5 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



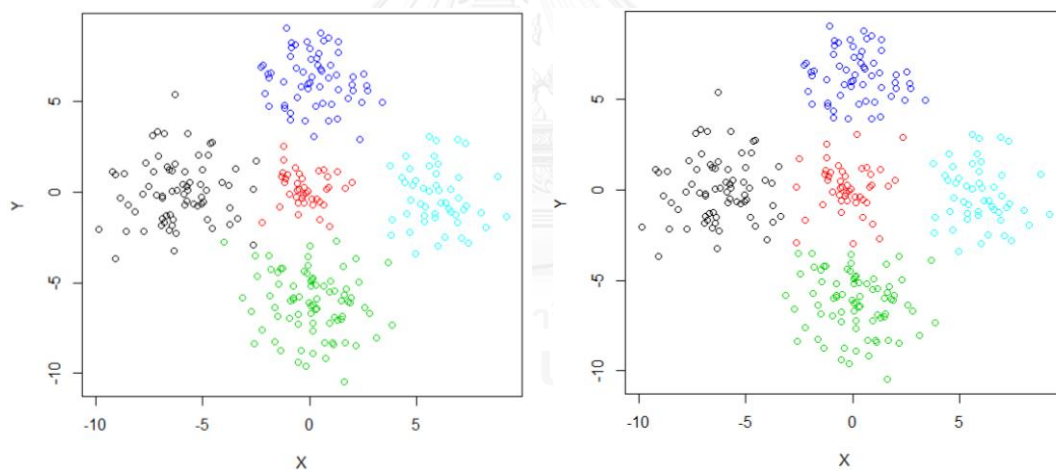
รูปที่ 4.5.2.3.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.3



รูปที่ 4.5.2.3.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.3



รูปที่ 4.5.2.3.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.2.3



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.3.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.3

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.3.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski กลุ่มที่ประกอบด้วยวิธี P95, วิธี Xia P95 และวิธี Karami กลุ่มที่ประกอบด้วยวิธี P75 และวิธี Xia P75 กลุ่มที่ประกอบด้วยวิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ ซึ่งวิธี Daszykowski ตั้งแต่ MinPts 20 เป็นต้นไปให้ค่า Eps สูงกว่าวิธีอื่น ๆ เป็นอย่างมาก

จากรูปที่ 4.5.2.3.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่อู่เข้าสู่ 0.267 (80/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-5 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.267 (80/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-6 เป็นต้นไป ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.267 (80/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 5 จากนั้น modified purity มีค่าลดลง จนกระทั่งที่ MinPts 45 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.267 (80/300)

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-6 แล้วค่อย ๆ เพิ่มสูงขึ้น จากนั้นลดลงอย่างรวดเร็วในช่วง MinPts 25 เป็นต้นมา และที่ค่า MinPts ตั้งแต่ 50 ค่า modified purity มีแนวโน้มลู่อู่เข้าสู่ 0.267

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง จนกระทั่งสูงสุดที่ MinPts 30 จากนั้น modified purity มีค่าลดลง

จากรูปที่ 4.5.2.3.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.3.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.96404

ตารางที่ 4.5.2.3.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.3

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.8917	0.4133	0.3373	0.5211	0.9947	0.1520	0.4133	0.2980	0.5179	1.1292	1.1894
3	1.0921	0.6094	0.5202	0.7421	1.3061	0.3673	0.6094	0.5195	0.7631	1.3112	1.5250
4	1.2611	0.7528	0.6521	0.9035	1.5229	0.5165	0.7528	0.6681	0.9308	1.4867	1.7860
5	1.4099	0.8708	0.7617	1.0392	1.6967	0.6359	0.8708	0.7879	1.0670	1.6387	2.0114
6	1.5445	0.9734	0.8580	1.1580	1.8425	0.7380	0.9734	0.8911	1.1846	1.7733	2.2163
7	1.6683	1.0651	0.9448	1.2638	1.9700	0.8300	1.0651	0.9833	1.2888	1.8915	2.4068
8	1.7835	1.1491	1.0254	1.3603	2.0822	0.9142	1.1491	1.0678	1.3838	1.9993	2.5883
9	1.8916	1.2270	1.1012	1.4483	2.1863	0.9921	1.2270	1.1460	1.4715	2.0989	2.7615
10	1.9940	1.3000	1.1709	1.5314	2.2825	1.0651	1.3000	1.2192	1.5534	2.1920	2.9269
12	2.1843	1.4355	1.3039	1.6840	2.4589	1.2012	1.4355	1.3552	1.7044	2.3616	3.2410
15	2.4421	1.6197	1.4856	1.8891	2.6876	1.3873	1.6197	1.5404	1.9080	2.5878	3.6762
20	2.8199	1.8940	1.7600	2.1866	3.0176	1.6674	1.8940	1.8171	2.2070	2.9123	4.3316
25	3.1527	2.1466	2.0157	2.4524	3.3143	1.9278	2.1466	2.0726	2.4783	3.1989	4.9279
30	3.4537	2.3908	2.2654	2.7011	3.5881	2.1818	2.3908	2.3203	2.7367	3.4650	5.4836
35	3.7304	2.6379	2.5137	2.9438	3.8641	2.4401	2.6379	2.5713	2.9949	3.7251	6.0077
40	3.9879	2.8972	2.7761	3.1965	4.1451	2.7095	2.8972	2.8341	3.2649	3.9981	6.5097
45	4.2298	3.1806	3.0614	3.4848	4.4783	2.9941	3.1806	3.1180	3.5666	4.3239	6.9912
50	4.4586	3.5131	3.3498	3.8562	5.0312	3.2967	3.5131	3.4403	3.9485	4.8098	7.4545

ตารางที่ 4.5.2.3.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

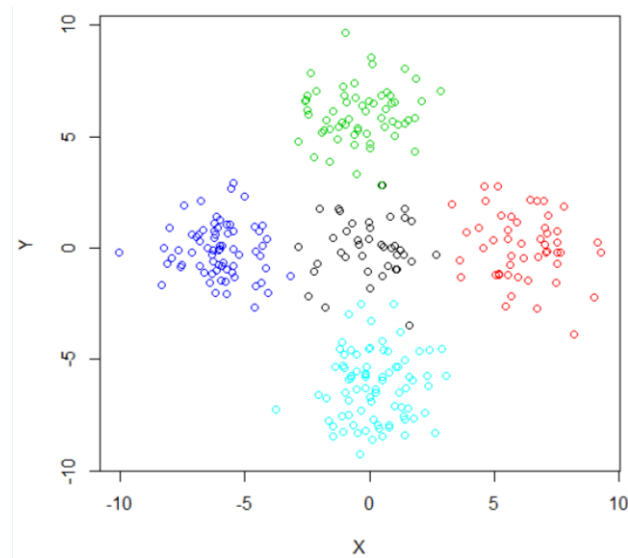
4.5.2.3

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7701	0.1529	0.0997	0.2785	0.7617	0.0410	0.1529	0.0828	0.2739	0.6804	0.6310
3	0.7139	0.4228	0.2759	0.6425	0.5297	0.1196	0.4228	0.2781	0.6720	0.5276	0.3899
4	0.6173	0.6151	0.4436	0.7683	0.4138	0.2333	0.6151	0.4748	0.7782	0.4348	0.3163
5	0.5455	0.7048	0.5593	0.8026	0.3621	0.3463	0.7048	0.5986	0.7998	0.3898	0.2848
6	0.4863	0.7530	0.6290	0.8046	0.3216	0.4438	0.7530	0.6730	0.7971	0.3560	0.2717
7	0.4375	0.7798	0.6781	0.7981	0.3056	0.5179	0.7798	0.7158	0.7858	0.3285	0.2673
8	0.4088	0.7977	0.7110	0.7941	0.2915	0.5732	0.7977	0.7451	0.7742	0.3157	0.2662
9	0.3779	0.8125	0.7367	0.7790	0.2821	0.6195	0.8125	0.7692	0.7584	0.2983	0.2667
10	0.3487	0.8269	0.7571	0.7615	0.2752	0.6507	0.8269	0.7876	0.7411	0.2873	0.2665
12	0.3030	0.8382	0.7881	0.7108	0.2719	0.7045	0.8382	0.8150	0.6978	0.2749	0.2665
15	0.2756	0.8476	0.8227	0.6349	0.2681	0.7614	0.8476	0.8405	0.6193	0.2703	0.2666
20	0.2678	0.8411	0.8568	0.5109	0.2666	0.8243	0.8410	0.8559	0.4944	0.2672	0.2667
25	0.2666	0.7934	0.8628	0.3956	0.2664	0.8572	0.7934	0.8466	0.3851	0.2671	0.2667
30	0.2665	0.7010	0.8278	0.3266	0.2666	0.8625	0.7010	0.7799	0.3151	0.2665	0.2667
35	0.2666	0.5719	0.7411	0.2932	0.2666	0.8132	0.5719	0.6680	0.2840	0.2666	0.2667
40	0.2666	0.4440	0.6020	0.2738	0.2666	0.6954	0.4440	0.5318	0.2712	0.2666	0.2667
45	0.2667	0.3507	0.4311	0.2673	0.2667	0.5174	0.3504	0.3948	0.2672	0.2666	0.2667
50	0.2667	0.2755	0.2933	0.2664	0.2667	0.3328	0.2755	0.2823	0.2665	0.2667	0.2667

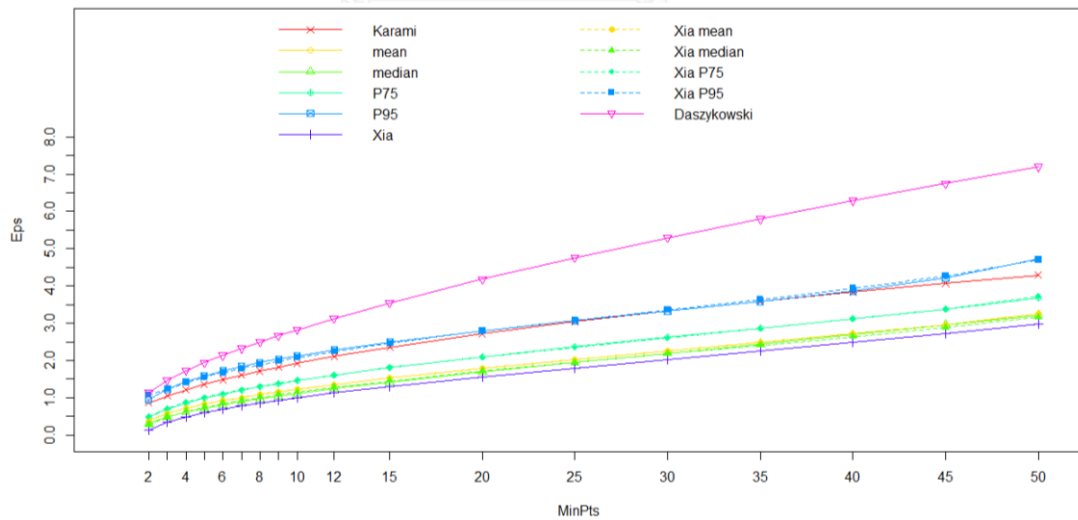
4.5.2.4 ขนาดตัวอย่าง N=300 กลุ่มละ 40, 50, 60, 70, 80 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

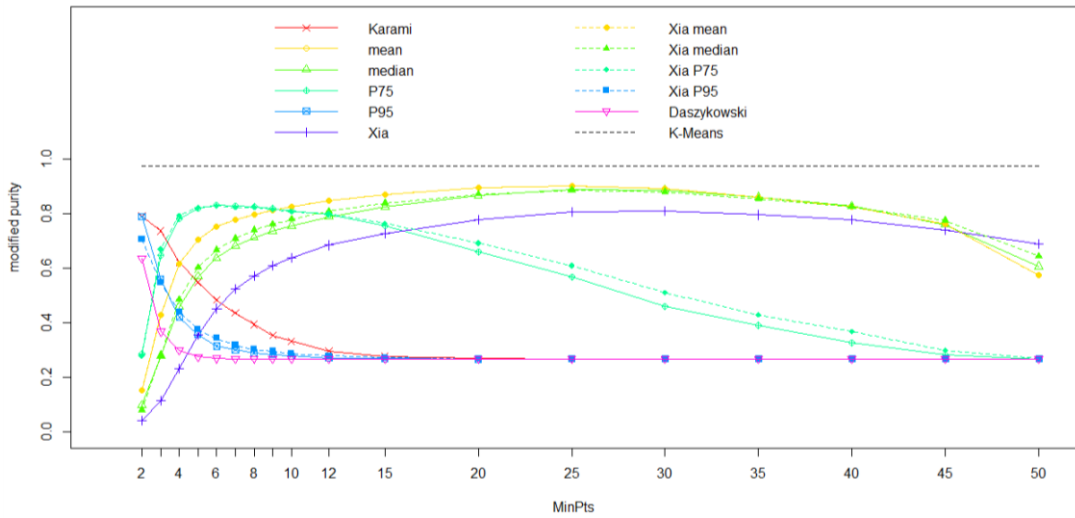
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



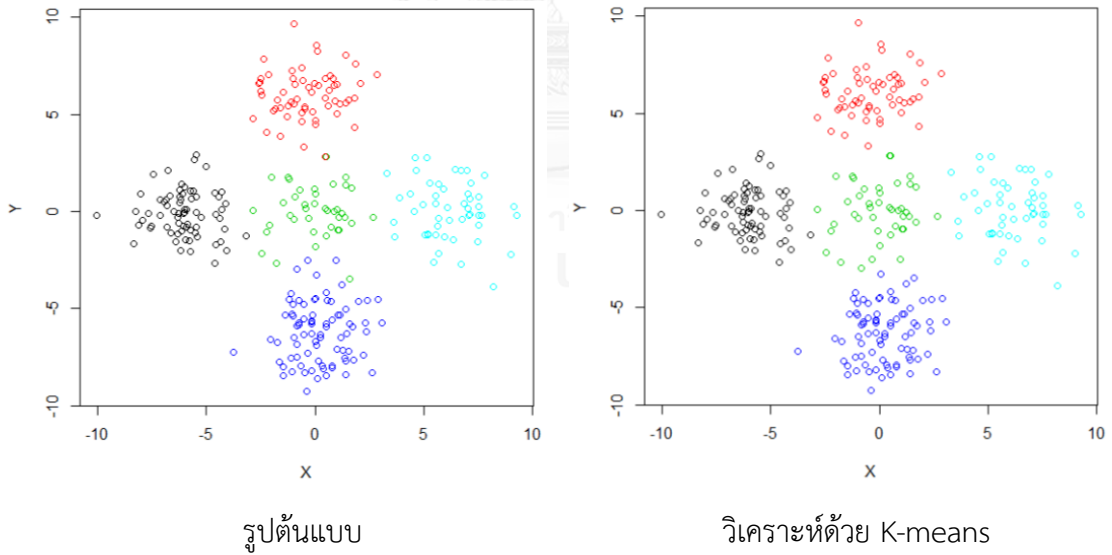
รูปที่ 4.5.2.4.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.4



รูปที่ 4.5.2.4.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.4



รูปที่ 4.5.2.4.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.2.4



รูปที่ 4.5.2.4.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.4

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.4.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski กลุ่มที่ประกอบด้วยวิธี P95, วิธี Xia P95 และวิธี Karami กลุ่มที่ประกอบด้วยวิธี P75 และวิธี Xia P75 กลุ่มที่ประกอบด้วยวิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ ซึ่งวิธี Daszykowski ตั้งแต่ MinPts 20 เป็นต้นไปให้ค่า Eps สูงกว่าวิธีอื่น ๆ เป็นอย่างมาก

จากรูปที่ 4.5.2.4.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.267 (80/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-5 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (80/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-6 เป็นต้นไป ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (80/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 6 จากนั้น modified purity มีค่าลดลง จนกระทั่งที่ MinPts 45 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (80/300)

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้น จากนั้นที่ค่า MinPts 25 เป็นต้นมา modified purity มีค่าลดลง

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง จนกระทั่งสูงสุดที่ MinPts 30 จากนั้น modified purity มีค่าลดลง

จากรูปที่ 4.5.2.4.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.4.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.97426

ตารางที่ 4.5.2.4.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.4

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.8591	0.3898	0.3186	0.4932	0.9420	0.1409	0.3898	0.2793	0.4880	1.0723	1.1477
3	1.0521	0.5740	0.4896	0.7020	1.2281	0.3450	0.5740	0.4889	0.7189	1.2371	1.4724
4	1.2149	0.7102	0.6167	0.8586	1.4337	0.4869	0.7102	0.6302	0.8783	1.4035	1.7242
5	1.3583	0.8216	0.7221	0.9898	1.5889	0.5999	0.8216	0.7434	1.0067	1.5462	1.9422
6	1.4879	0.9185	0.8155	1.1010	1.7224	0.6962	0.9185	0.8407	1.1178	1.6735	2.1398
7	1.6071	1.0050	0.8992	1.2029	1.8377	0.7824	1.0050	0.9276	1.2164	1.7866	2.3250
8	1.7181	1.0839	0.9760	1.2949	1.9425	0.8601	1.0839	1.0064	1.3061	1.8909	2.4994
9	1.8223	1.1575	1.0481	1.3804	2.0397	0.9328	1.1575	1.0800	1.3894	1.9870	2.6663
10	1.9209	1.2265	1.1157	1.4610	2.1257	1.0010	1.2265	1.1489	1.4672	2.0768	2.8265
12	2.1042	1.3541	1.2439	1.6106	2.2861	1.1275	1.3541	1.2764	1.6103	2.2411	3.1294
15	2.3526	1.5272	1.4212	1.8095	2.4958	1.2996	1.5272	1.4495	1.8033	2.4610	3.5501
20	2.7166	1.7860	1.6900	2.1039	2.7966	1.5577	1.7860	1.7084	2.0892	2.7835	4.1818
25	3.0372	2.0243	1.9444	2.3689	3.0640	1.7959	2.0243	1.9469	2.3507	3.0762	4.7557
30	3.3271	2.2538	2.1937	2.6190	3.3254	2.0245	2.2538	2.1762	2.6016	3.3572	5.2925
35	3.5937	2.4818	2.4412	2.8678	3.5822	2.2505	2.4818	2.4036	2.8507	3.6376	5.8001
40	3.8418	2.7160	2.6917	3.1220	3.8635	2.4794	2.7160	2.6361	3.1080	3.9330	6.2840
45	4.0748	2.9658	2.9550	3.3853	4.2044	2.7171	2.9658	2.8819	3.3869	4.2666	6.7492
50	4.2953	3.2538	3.2162	3.6812	4.7242	2.9704	3.2538	3.1581	3.7232	4.7114	7.1957

ตารางที่ 4.5.2.4.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

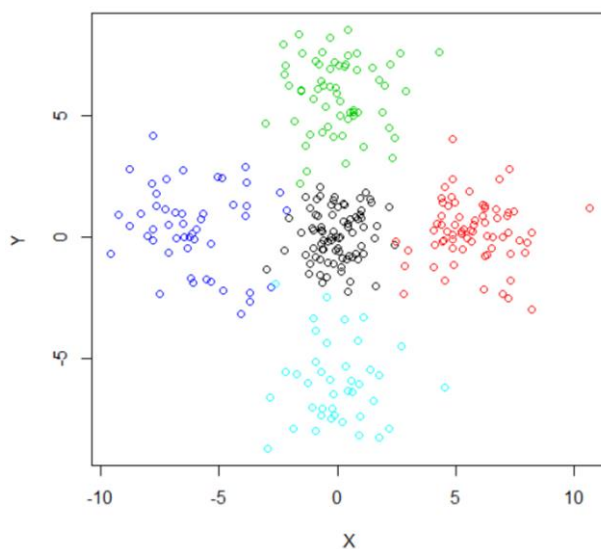
4.5.2.4

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7911	0.1506	0.0968	0.2855	0.7896	0.0394	0.1506	0.0791	0.2783	0.7059	0.6361
3	0.7368	0.4298	0.2791	0.6477	0.5591	0.1140	0.4298	0.2787	0.6699	0.5468	0.3678
4	0.6217	0.6161	0.4581	0.7819	0.4196	0.2314	0.6161	0.4844	0.7938	0.4405	0.2987
5	0.5488	0.7042	0.5698	0.8209	0.3560	0.3555	0.7042	0.6027	0.8224	0.3753	0.2748
6	0.4838	0.7514	0.6373	0.8310	0.3140	0.4521	0.7514	0.6669	0.8283	0.3438	0.2694
7	0.4358	0.7769	0.6815	0.8288	0.3005	0.5230	0.7769	0.7090	0.8243	0.3173	0.2672
8	0.3934	0.7978	0.7125	0.8263	0.2885	0.5717	0.7977	0.7400	0.8232	0.3022	0.2669
9	0.3525	0.8129	0.7374	0.8210	0.2825	0.6096	0.8129	0.7618	0.8158	0.2947	0.2666
10	0.3315	0.8273	0.7570	0.8114	0.2777	0.6392	0.8273	0.7794	0.8085	0.2851	0.2666
12	0.2964	0.8485	0.7893	0.7988	0.2698	0.6847	0.8485	0.8096	0.8006	0.2775	0.2667
15	0.2763	0.8717	0.8250	0.7562	0.2679	0.7280	0.8717	0.8392	0.7648	0.2714	0.2667
20	0.2683	0.8960	0.8670	0.6596	0.2666	0.7794	0.8960	0.8719	0.6926	0.2676	0.2667
25	0.2670	0.9033	0.8882	0.5677	0.2666	0.8065	0.9033	0.8868	0.6099	0.2666	0.2667
30	0.2667	0.8939	0.8876	0.4597	0.2667	0.8093	0.8939	0.8800	0.5106	0.2667	0.2667
35	0.2667	0.8624	0.8617	0.3900	0.2667	0.7961	0.8624	0.8544	0.4289	0.2667	0.2667
40	0.2667	0.8263	0.8285	0.3254	0.2667	0.7792	0.8259	0.8252	0.3673	0.2667	0.2667
45	0.2667	0.7627	0.7606	0.2833	0.2667	0.7398	0.7627	0.7750	0.2988	0.2667	0.2667
50	0.2667	0.5750	0.6073	0.2676	0.2667	0.6900	0.5750	0.6444	0.2690	0.2667	0.2667

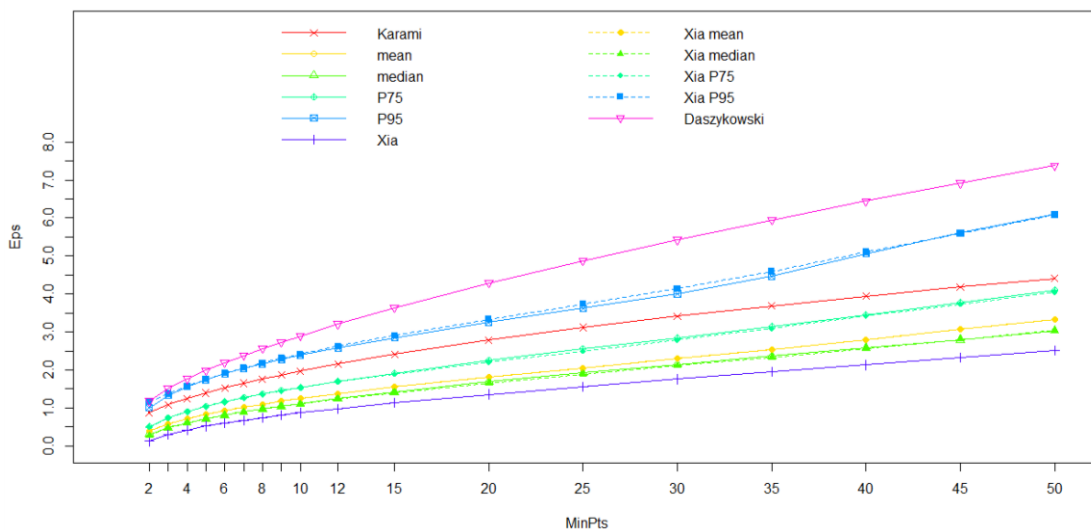
4.5.2.5 ขนาดตัวอย่าง N=300 กลุ่มละ 80, 70, 60, 50, 40 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

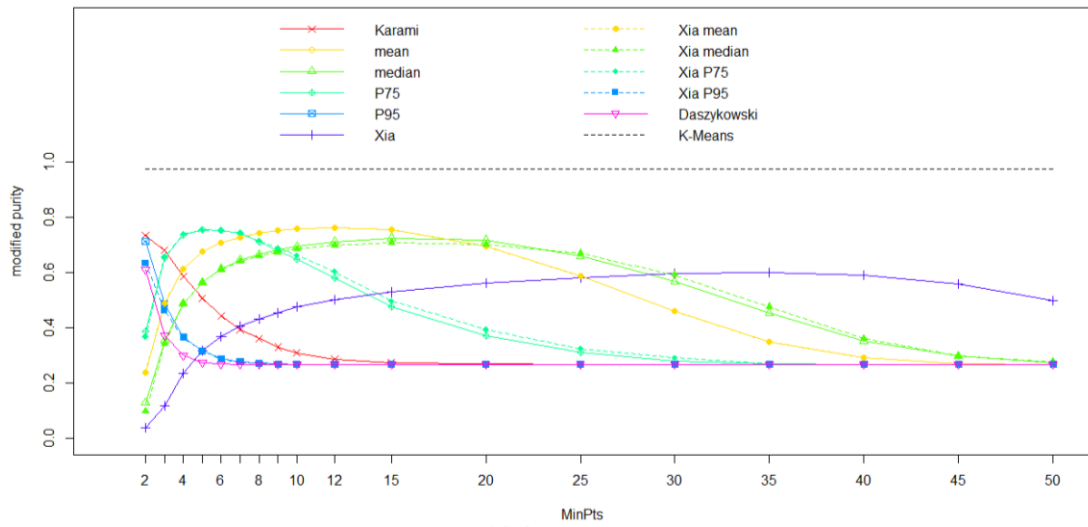
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_4 = \Sigma_5 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



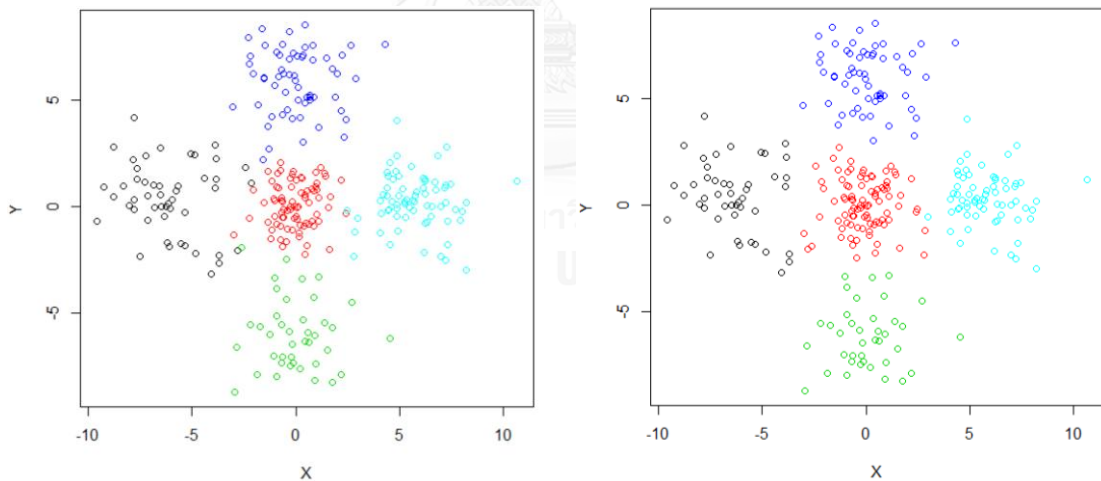
รูปที่ 4.5.2.5.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.5



รูปที่ 4.5.2.5.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.5



รูปที่ 4.5.2.5.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.2.5



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.5.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.5

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.5.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี Xia P95, วิธี P95, วิธี Karami, วิธี P75, วิธี Xia P75, วิธี mean, วิธี Xia mean, วิธี median, วิธี Xia median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.2.5.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.267 (80/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-5 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 9 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (80/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-6 เป็นต้นไป ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (80/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 5 จากนั้น modified purity มีค่าลดลง จนกระทั่งที่ MinPts 35 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (80/300)

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้น จากนั้นลดลงอย่างรวดเร็วในช่วง MinPts 20 เป็นต้นมา และที่ค่า MinPts ตั้งแต่ 50 ค่า modified purity มีแนวโน้มลู่เข้าสู่ 0.267

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง จนกระทั่งสูงสุดที่ MinPts 35 จากนั้น modified purity มีค่าลดลง

จากรูปที่ 4.5.2.5.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.5.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.97488

ตารางที่ 4.5.2.5.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.5

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.8817	0.3949	0.3075	0.5052	1.0061	0.1235	0.3949	0.2680	0.4898	1.1480	1.1771
3	1.0798	0.5824	0.4799	0.7306	1.3288	0.2989	0.5824	0.4719	0.7348	1.3681	1.5097
4	1.2469	0.7201	0.6106	0.8975	1.5621	0.4212	0.7201	0.6083	0.9037	1.5765	1.7667
5	1.3940	0.8328	0.7187	1.0371	1.7443	0.5196	0.8328	0.7177	1.0401	1.7537	1.9898
6	1.5271	0.9308	0.8135	1.1587	1.8991	0.6031	0.9308	0.8116	1.1585	1.9130	2.1926
7	1.6494	1.0184	0.8980	1.2651	2.0397	0.6783	1.0184	0.8956	1.2638	2.0541	2.3821
8	1.7633	1.0977	0.9761	1.3636	2.1621	0.7466	1.0977	0.9717	1.3590	2.1817	2.5618
9	1.8703	1.1718	1.0494	1.4552	2.2787	0.8099	1.1718	1.0424	1.4477	2.3017	2.7324
10	1.9715	1.2413	1.1179	1.5405	2.3842	0.8699	1.2413	1.1090	1.5307	2.4131	2.8962
12	2.1596	1.3701	1.2462	1.6995	2.5809	0.9808	1.3701	1.2321	1.6843	2.6200	3.2068
15	2.4145	1.5456	1.4227	1.9156	2.8462	1.1312	1.5456	1.3995	1.8932	2.9030	3.6379
20	2.7881	1.8096	1.6857	2.2444	3.2476	1.3558	1.8096	1.6504	2.2074	3.3317	4.2880
25	3.1171	2.0544	1.9239	2.5506	3.6204	1.5610	2.0544	1.8819	2.4992	3.7354	4.8772
30	3.4147	2.2934	2.1489	2.8502	4.0055	1.7556	2.2933	2.1057	2.7857	4.1418	5.4260
35	3.6883	2.5393	2.3666	3.1520	4.4633	1.9436	2.5393	2.3314	3.0846	4.5866	5.9458
40	3.9429	2.8029	2.5810	3.4502	5.0603	2.1275	2.8029	2.5667	3.4106	5.1015	6.4425
45	4.1821	3.0642	2.7990	3.7695	5.6066	2.3168	3.0642	2.8027	3.7315	5.5967	6.9177
50	4.4083	3.3284	3.0352	4.0983	6.0987	2.5167	3.3285	3.0444	4.0534	6.0792	7.3759

ตารางที่ 4.5.2.5.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

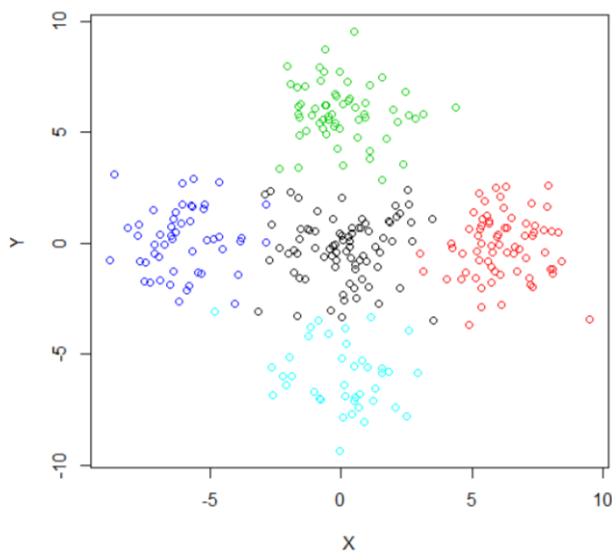
4.5.2.5

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7347	0.2370	0.1277	0.3869	0.7137	0.0367	0.2370	0.0959	0.3679	0.6338	0.6111
3	0.6802	0.4889	0.3534	0.6538	0.4869	0.1168	0.4889	0.3435	0.6575	0.4624	0.3727
4	0.5871	0.6144	0.4868	0.7367	0.3660	0.2352	0.6144	0.4855	0.7400	0.3631	0.3001
5	0.5071	0.6763	0.5643	0.7556	0.3162	0.3161	0.6762	0.5639	0.7539	0.3164	0.2733
6	0.4431	0.7076	0.6136	0.7544	0.2873	0.3670	0.7076	0.6095	0.7541	0.2864	0.2679
7	0.3937	0.7279	0.6460	0.7432	0.2771	0.4061	0.7279	0.6399	0.7442	0.2757	0.2667
8	0.3603	0.7429	0.6662	0.7101	0.2717	0.4321	0.7429	0.6598	0.7156	0.2700	0.2667
9	0.3302	0.7533	0.6824	0.6789	0.2684	0.4535	0.7534	0.6742	0.6883	0.2682	0.2667
10	0.3086	0.7584	0.6944	0.6493	0.2672	0.4752	0.7584	0.6849	0.6633	0.2672	0.2667
12	0.2857	0.7636	0.7123	0.5809	0.2667	0.5009	0.7636	0.6995	0.6045	0.2667	0.2667
15	0.2735	0.7555	0.7251	0.4752	0.2667	0.5289	0.7555	0.7092	0.4972	0.2667	0.2667
20	0.2679	0.6958	0.7180	0.3697	0.2667	0.5624	0.6958	0.7027	0.3934	0.2667	0.2667
25	0.2667	0.5882	0.6594	0.3094	0.2667	0.5812	0.5882	0.6704	0.3247	0.2667	0.2667
30	0.2667	0.4609	0.5669	0.2806	0.2667	0.5962	0.4609	0.5918	0.2911	0.2667	0.2667
35	0.2667	0.3479	0.4529	0.2685	0.2667	0.5993	0.3479	0.4767	0.2695	0.2667	0.2667
40	0.2667	0.2906	0.3508	0.2667	0.2667	0.5901	0.2906	0.3606	0.2667	0.2667	0.2667
45	0.2667	0.2691	0.2974	0.2667	0.2667	0.5580	0.2691	0.2988	0.2667	0.2667	0.2667
50	0.2667	0.2667	0.2738	0.2667	0.2667	0.4992	0.2667	0.2754	0.2667	0.2667	0.2667

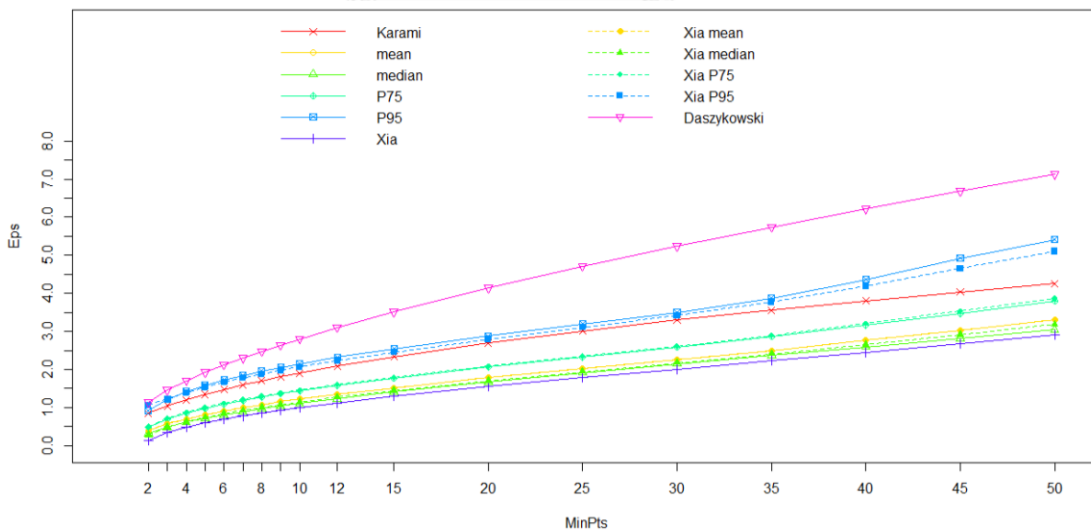
4.5.2.6 ขนาดตัวอย่าง N=300 กลุ่มละ 80, 70, 60, 50, 40 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

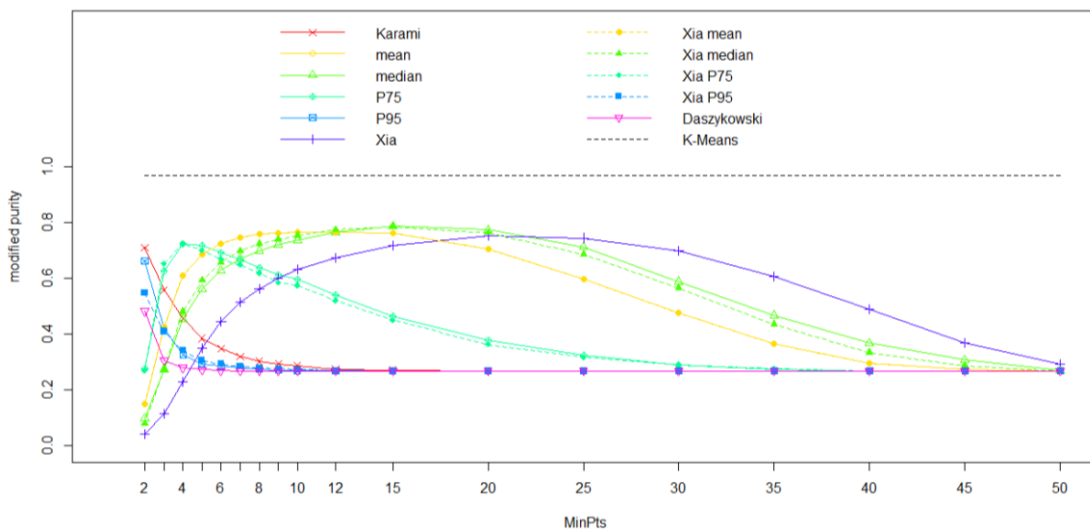
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



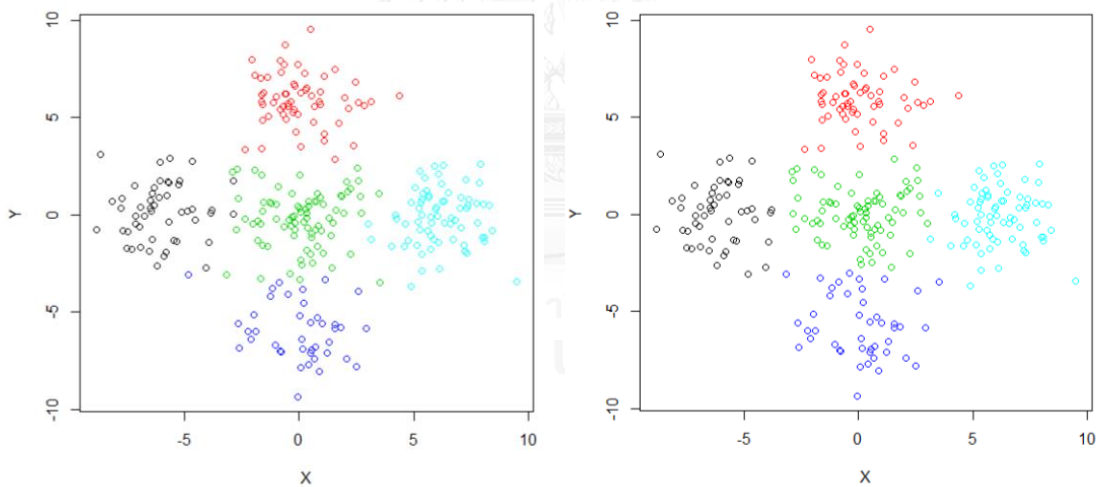
รูปที่ 4.5.2.6.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.6



รูปที่ 4.5.2.6.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.6



รูปที่ 4.5.2.6.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.2.6



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.6.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.6

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.6.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.2.6.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-3 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.267 (80/300) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (80/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-5 เป็นต้นไป ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (80/300) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 4 จากนั้น modified purity มีค่าลดลง จนกระทั่งที่ MinPts 35 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (80/300)

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้น จากนั้นลดลงอย่างรวดเร็วในช่วง MinPts 15 เป็นต้นมา และที่ค่า MinPts ตั้งแต่ 50 ค่า modified purity มีแนวโน้มลู่เข้าสู่ 0.267

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มขึ้นอย่างต่อเนื่อง จนกระทั่งสูงสุดที่ MinPts 20 จากนั้น modified purity มีค่าลดลง

จากรูปที่ 4.5.2.6.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.6.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9677467

ตารางที่ 4.5.2.6.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.6

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.8506	0.3853	0.3162	0.4862	0.9196	0.1400	0.3853	0.2767	0.4826	1.0582	1.1371
3	1.0418	0.5690	0.4864	0.6909	1.2128	0.3428	0.5690	0.4850	0.7124	1.2245	1.4588
4	1.2030	0.7038	0.6120	0.8457	1.4197	0.4837	0.7038	0.6250	0.8700	1.3881	1.7071
5	1.3450	0.8147	0.7157	0.9722	1.5815	0.5946	0.8147	0.7370	0.9983	1.5336	1.9229
6	1.4733	0.9113	0.8082	1.0833	1.7193	0.6905	0.9113	0.8341	1.1093	1.6613	2.1185
7	1.5914	0.9977	0.8900	1.1820	1.8430	0.7759	0.9977	0.9206	1.2078	1.7753	2.3020
8	1.7013	1.0766	0.9670	1.2730	1.9511	0.8540	1.0766	0.9995	1.2975	1.8788	2.4751
9	1.8045	1.1499	1.0391	1.3580	2.0515	0.9261	1.1499	1.0726	1.3805	1.9753	2.6399
10	1.9021	1.2187	1.1065	1.4360	2.1488	0.9941	1.2187	1.1413	1.4580	2.0648	2.7976
12	2.0836	1.3462	1.2341	1.5794	2.3170	1.1198	1.3462	1.2686	1.6014	2.2306	3.0985
15	2.3296	1.5193	1.4068	1.7714	2.5416	1.2907	1.5193	1.4412	1.7949	2.4533	3.5146
20	2.6899	1.7795	1.6714	2.0595	2.8762	1.5470	1.7795	1.7004	2.0843	2.7862	4.1410
25	3.0074	2.0207	1.9145	2.3271	3.1910	1.7816	2.0207	1.9396	2.3529	3.0996	4.7091
30	3.2945	2.2554	2.1466	2.5909	3.5018	2.0064	2.2554	2.1710	2.6156	3.4125	5.2399
35	3.5584	2.4977	2.3710	2.8711	3.8743	2.2283	2.4977	2.4065	2.8927	3.7620	5.7418
40	3.8041	2.7605	2.5902	3.1596	4.3627	2.4510	2.7605	2.6555	3.2042	4.1892	6.2217
45	4.0349	3.0350	2.8139	3.4612	4.9208	2.6771	3.0350	2.9135	3.5332	4.6522	6.6824
50	4.2532	3.3135	3.0406	3.8066	5.4069	2.9138	3.3135	3.1778	3.8628	5.1031	7.1249

ตารางที่ 4.5.2.6.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีที่

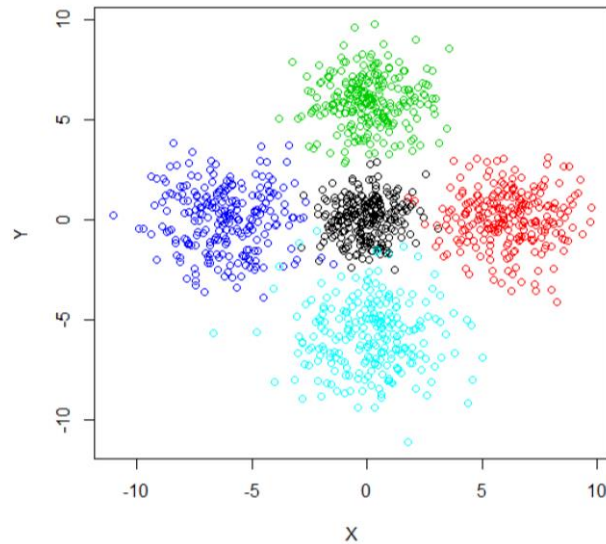
4.5.2.6

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7091	0.1474	0.0967	0.2744	0.6621	0.0393	0.1473	0.0776	0.2682	0.5469	0.4828
3	0.5583	0.4252	0.2725	0.6271	0.4177	0.1141	0.4252	0.2733	0.6538	0.4096	0.3053
4	0.4564	0.6082	0.4536	0.7235	0.3257	0.2282	0.6082	0.4793	0.7228	0.3398	0.2782
5	0.3840	0.6862	0.5608	0.7183	0.2923	0.3492	0.6862	0.5921	0.7009	0.3048	0.2712
6	0.3479	0.7244	0.6280	0.6920	0.2849	0.4453	0.7244	0.6580	0.6692	0.2924	0.2680
7	0.3193	0.7464	0.6700	0.6669	0.2786	0.5134	0.7464	0.6977	0.6486	0.2834	0.2675
8	0.3011	0.7579	0.6991	0.6379	0.2733	0.5627	0.7579	0.7241	0.6190	0.2785	0.2669
9	0.2928	0.7629	0.7209	0.6118	0.2703	0.6014	0.7629	0.7394	0.5853	0.2742	0.2667
10	0.2855	0.7651	0.7359	0.5964	0.2694	0.6313	0.7651	0.7530	0.5746	0.2717	0.2667
12	0.2734	0.7652	0.7643	0.5414	0.2682	0.6745	0.7652	0.7738	0.5193	0.2685	0.2667
15	0.2695	0.7617	0.7869	0.4641	0.2675	0.7175	0.7617	0.7851	0.4499	0.2674	0.2667
20	0.2669	0.7052	0.7748	0.3763	0.2667	0.7530	0.7052	0.7637	0.3630	0.2669	0.2667
25	0.2667	0.5963	0.7126	0.3232	0.2667	0.7448	0.5963	0.6865	0.3167	0.2667	0.2667
30	0.2667	0.4772	0.5884	0.2900	0.2667	0.6978	0.4772	0.5652	0.2875	0.2667	0.2667
35	0.2667	0.3660	0.4671	0.2729	0.2667	0.6060	0.3660	0.4349	0.2744	0.2667	0.2667
40	0.2667	0.2952	0.3676	0.2669	0.2667	0.4887	0.2952	0.3335	0.2669	0.2667	0.2667
45	0.2667	0.2712	0.3064	0.2667	0.2667	0.3672	0.2712	0.2866	0.2667	0.2667	0.2667
50	0.2667	0.2669	0.2679	0.2667	0.2667	0.2907	0.2669	0.2674	0.2667	0.2667	0.2667

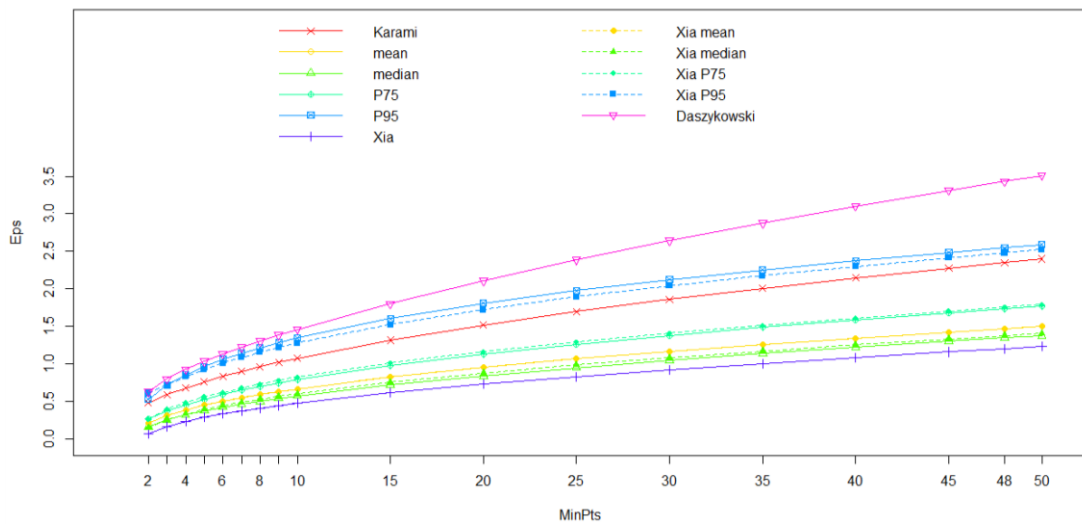
4.5.2.7 ขนาดตัวอย่าง N=1,200 กลุ่มละ 240, 240, 240, 240, 240 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

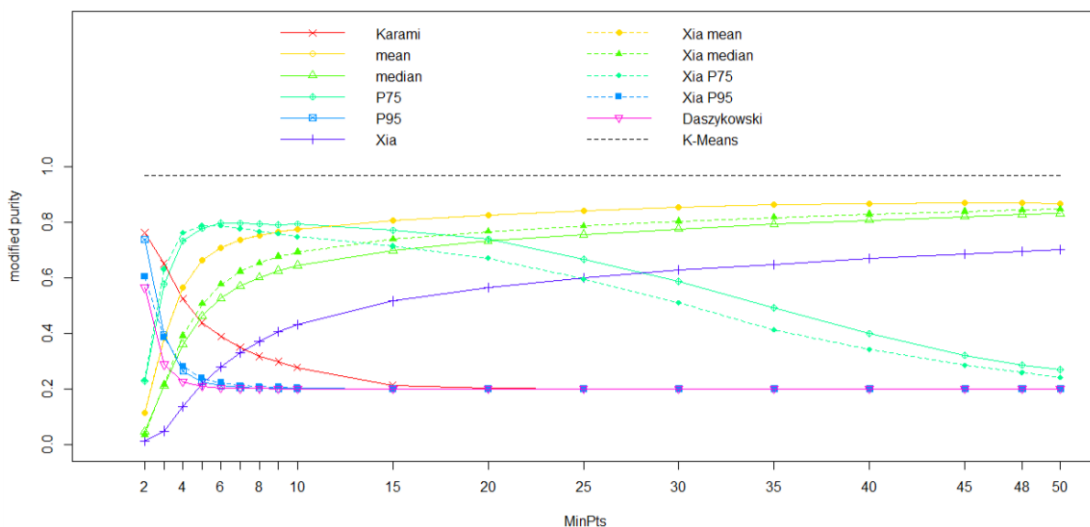
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_4 = \Sigma_5 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



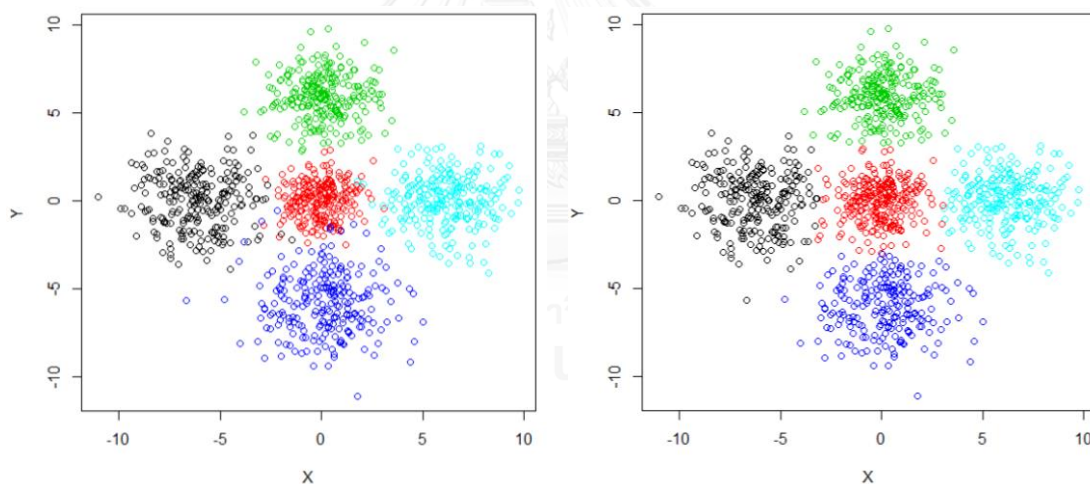
รูปที่ 4.5.2.7.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.7



รูปที่ 4.5.2.7.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.7



รูปที่ 4.5.2.7.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.2.7



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.7.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.7

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.7.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี

ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.2.7.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่อู่เข้าสู่ 0.2 (240/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 8 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.2 (240/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 2 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.2 (240/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 6 จากนั้น modified purity มีค่าลดลงและมีแนวโน้มที่จะลู่อู่เข้าสู่ 0.2

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 15-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.2.7.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.7.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.97052

ตารางที่ 4.5.2.7.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.7

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.4797	0.2096	0.1633	0.2593	0.5272	0.0663	0.2096	0.1436	0.2611	0.6046	0.6273
3	0.5875	0.3109	0.2534	0.3728	0.7147	0.1649	0.3109	0.2547	0.3921	0.7179	0.7967
4	0.6784	0.3854	0.3190	0.4571	0.8510	0.2323	0.3854	0.3287	0.4828	0.8288	0.9249
5	0.7585	0.4465	0.3734	0.5272	0.9609	0.2864	0.4465	0.3882	0.5565	0.9237	1.0336
6	0.8309	0.4995	0.4205	0.5887	1.0571	0.3324	0.4995	0.4393	0.6201	1.0080	1.1309
7	0.8974	0.5468	0.4629	0.6439	1.1404	0.3733	0.5468	0.4846	0.6768	1.0837	1.2201
8	0.9594	0.5898	0.5015	0.6947	1.2151	0.4105	0.5898	0.5259	0.7282	1.1526	1.3032
9	1.0176	0.6295	0.5377	0.7419	1.2829	0.4449	0.6295	0.5639	0.7756	1.2161	1.3818
10	1.0726	0.6664	0.5716	0.7855	1.3461	0.4769	0.6664	0.5993	0.8195	1.2748	1.4571
15	1.3137	0.8236	0.7174	0.9729	1.6034	0.6140	0.8236	0.7501	1.0059	1.5231	1.7998
20	1.5169	0.9525	0.8393	1.1268	1.8068	0.7281	0.9525	0.8742	1.1577	1.7229	2.1059
25	1.6960	1.0641	0.9467	1.2600	1.9743	0.8284	1.0641	0.9822	1.2884	1.8923	2.3848
30	1.8578	1.1642	1.0443	1.3780	2.1208	0.9193	1.1642	1.0794	1.4049	2.0416	2.6408
35	2.0067	1.2559	1.1345	1.4853	2.2512	1.0034	1.2559	1.1686	1.5112	2.1760	2.8781
40	2.1453	1.3411	1.2189	1.5848	2.3723	1.0822	1.3411	1.2518	1.6097	2.2997	3.1008
45	2.2754	1.4214	1.2993	1.6778	2.4842	1.1571	1.4214	1.3304	1.7021	2.4146	3.3116
48	2.3500	1.4676	1.3457	1.7308	2.5469	1.2003	1.4676	1.3757	1.7550	2.4800	3.4332
50	2.3985	1.4977	1.3760	1.7651	2.5880	1.2286	1.4977	1.4052	1.7895	2.5225	3.5125



ตารางที่ 4.5.2.7.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีที่

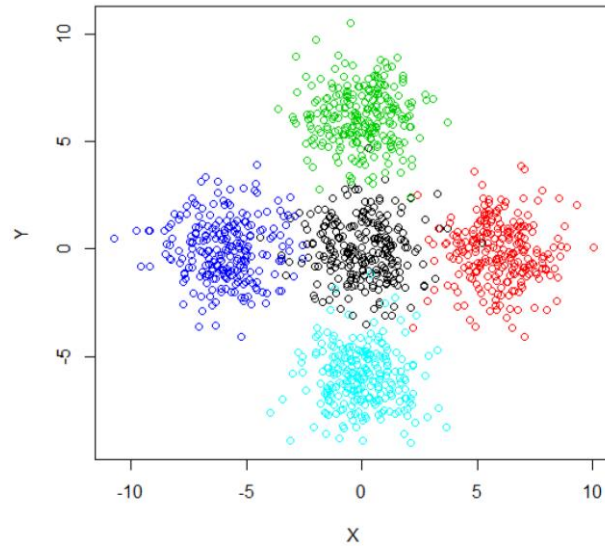
4.5.2.7

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7624	0.1139	0.0465	0.2276	0.7391	0.0124	0.1139	0.0339	0.2322	0.6044	0.5643
3	0.6509	0.3827	0.2121	0.5781	0.3942	0.0481	0.3827	0.2154	0.6320	0.3853	0.2882
4	0.5244	0.5667	0.3598	0.7322	0.2641	0.1363	0.5667	0.3904	0.7635	0.2817	0.2260
5	0.4380	0.6625	0.4622	0.7798	0.2259	0.2170	0.6625	0.5063	0.7890	0.2386	0.2098
6	0.3895	0.7079	0.5267	0.7988	0.2110	0.2795	0.7079	0.5763	0.7882	0.2215	0.2032
7	0.3505	0.7361	0.5707	0.7967	0.2059	0.3301	0.7361	0.6234	0.7774	0.2109	0.2024
8	0.3172	0.7536	0.6016	0.7948	0.2035	0.3717	0.7536	0.6536	0.7672	0.2082	0.2012
9	0.2978	0.7660	0.6261	0.7925	0.2016	0.4051	0.7660	0.6765	0.7603	0.2063	0.2004
10	0.2770	0.7757	0.6449	0.7929	0.2012	0.4323	0.7757	0.6922	0.7489	0.2024	0.2000
15	0.2123	0.8064	0.6997	0.7713	0.2000	0.5162	0.8064	0.7396	0.7146	0.2004	0.2000
20	0.2011	0.8252	0.7325	0.7395	0.2000	0.5650	0.8252	0.7667	0.6709	0.2000	0.2000
25	0.2000	0.8414	0.7563	0.6672	0.2000	0.5991	0.8414	0.7871	0.5965	0.2000	0.2000
30	0.2000	0.8533	0.7762	0.5872	0.2000	0.6277	0.8533	0.8040	0.5104	0.2000	0.2000
35	0.2000	0.8629	0.7933	0.4920	0.2000	0.6488	0.8629	0.8176	0.4141	0.2000	0.2000
40	0.2000	0.8687	0.8081	0.3989	0.2000	0.6694	0.8687	0.8294	0.3420	0.2000	0.2000
45	0.2000	0.8694	0.8208	0.3213	0.2000	0.6870	0.8694	0.8390	0.2845	0.2000	0.2000
48	0.2000	0.8713	0.8278	0.2867	0.2000	0.6969	0.8713	0.8441	0.2593	0.2000	0.2000
50	0.2000	0.8666	0.8320	0.2694	0.2000	0.7027	0.8666	0.8470	0.2420	0.2000	0.2000

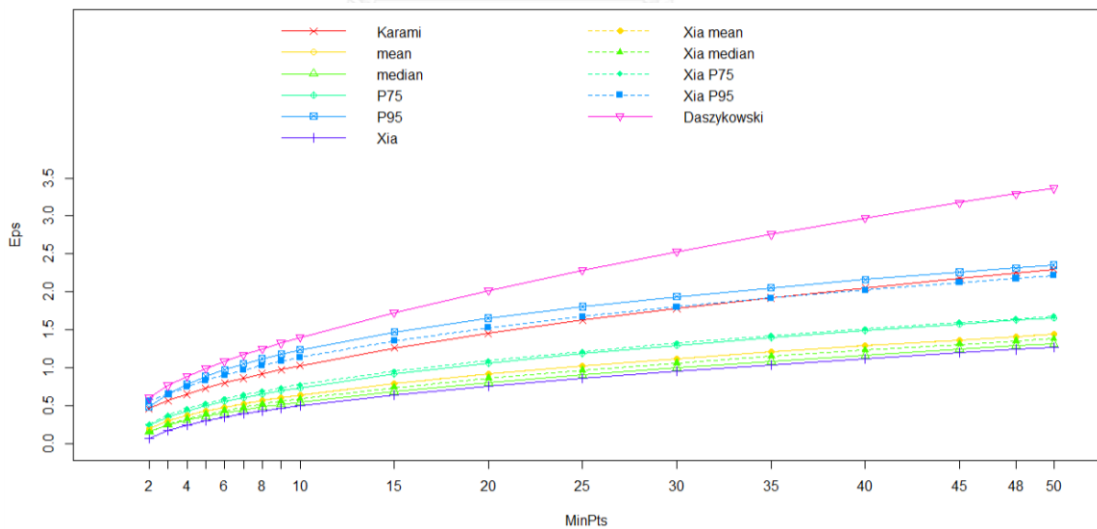
4.5.2.8 ขนาดตัวอย่าง N=1,200 กลุ่มละ 240, 240, 240, 240, 240 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

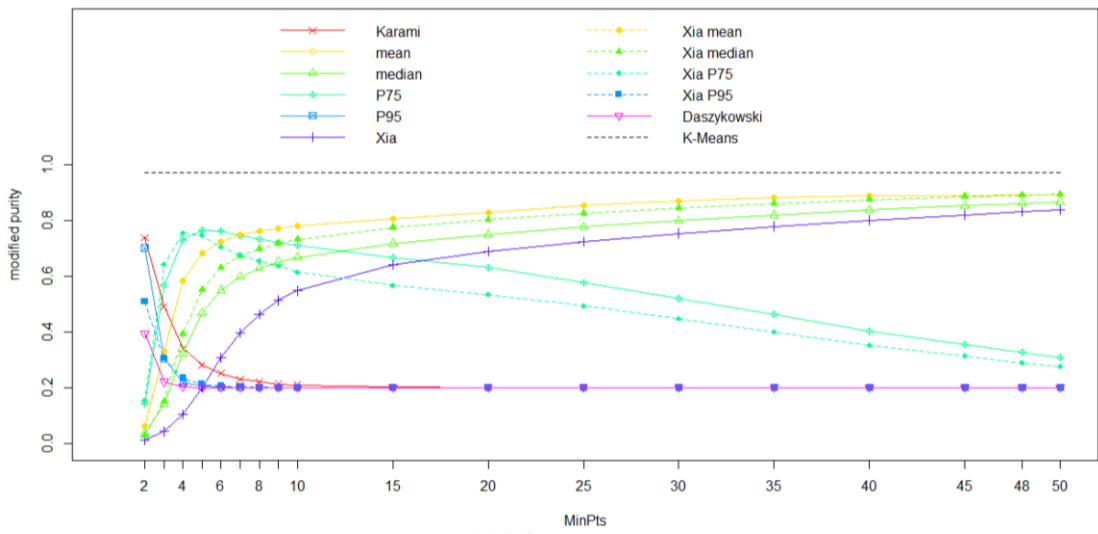
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



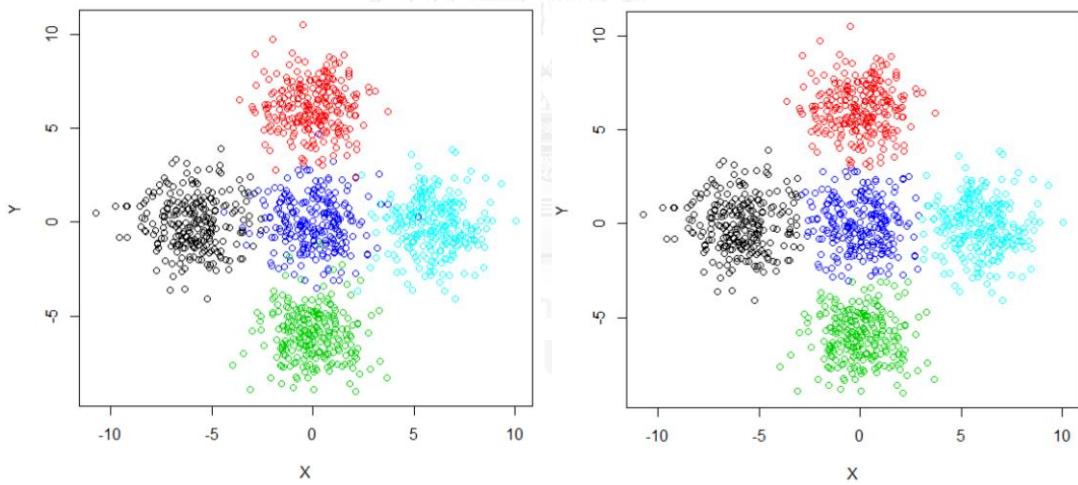
รูปที่ 4.5.2.8.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.8



รูปที่ 4.5.2.8.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.8



รูปที่ 4.5.2.8.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.2.8



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.8.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.8

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.8.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski กลุ่มที่ประกอบด้วยวิธี P95, วิธี Xia P95 และวิธี Karami กลุ่มที่ประกอบด้วยวิธี P75 และวิธี Xia P75 กลุ่มที่ประกอบด้วยวิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.2.8.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-3 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่อู่เข้าสู่ 0.2 (240/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 6 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.2 (240/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่อู่เข้าสู่ 0.2 (240/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 5 สำหรับวิธี P75 และที่ MinPts 4 สำหรับวิธี Xia P75 จากนั้น modified purity มีค่าลดลงและมีแนวโน้มที่จะลู่อู่เข้าสู่ 0.2

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 15-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.2.8.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.8.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.971875

ตารางที่ 4.5.2.8.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.8

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.4594	0.1997	0.1599	0.2462	0.4835	0.0693	0.1997	0.1416	0.2503	0.5560	0.6015
3	0.5626	0.2964	0.2462	0.3500	0.6552	0.1720	0.2964	0.2498	0.3722	0.6516	0.7640
4	0.6497	0.3680	0.3088	0.4288	0.7827	0.2427	0.3680	0.3227	0.4573	0.7474	0.8868
5	0.7263	0.4267	0.3602	0.4943	0.8840	0.2984	0.4267	0.3810	0.5265	0.8308	0.9911
6	0.7957	0.4775	0.4048	0.5517	0.9708	0.3461	0.4775	0.4311	0.5860	0.9042	1.0844
7	0.8594	0.5229	0.4446	0.6039	1.0475	0.3884	0.5229	0.4757	0.6391	0.9702	1.1697
8	0.9188	0.5642	0.4815	0.6514	1.1160	0.4268	0.5642	0.5162	0.6872	1.0302	1.2495
9	0.9745	0.6023	0.5154	0.6953	1.1792	0.4622	0.6023	0.5535	0.7315	1.0856	1.3250
10	1.0272	0.6378	0.5477	0.7365	1.2362	0.4953	0.6378	0.5882	0.7726	1.1369	1.3973
15	1.2581	0.7894	0.6859	0.9138	1.4718	0.6369	0.7894	0.7368	0.9475	1.3538	1.7256
20	1.4527	0.9135	0.8022	1.0588	1.6533	0.7544	0.9135	0.8589	1.0894	1.5271	2.0191
25	1.6241	1.0213	0.9043	1.1838	1.8050	0.8575	1.0213	0.9652	1.2116	1.6739	2.2865
30	1.7792	1.1180	0.9971	1.2954	1.9367	0.9510	1.1180	1.0610	1.3205	1.8028	2.5321
35	1.9217	1.2065	1.0839	1.3961	2.0535	1.0374	1.2065	1.1489	1.4196	1.9187	2.7599
40	2.0544	1.2889	1.1648	1.4887	2.1611	1.1183	1.2889	1.2308	1.5113	2.0244	2.9731
45	2.1790	1.3665	1.2417	1.5750	2.2601	1.1950	1.3665	1.3082	1.5972	2.1225	3.1751
48	2.2505	1.4111	1.2864	1.6244	2.3150	1.2393	1.4111	1.3528	1.6464	2.1783	3.2918
50	2.2969	1.4402	1.3157	1.6567	2.3522	1.2683	1.4402	1.3819	1.6784	2.2143	3.3678

ตารางที่ 4.5.2.8.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีนี้

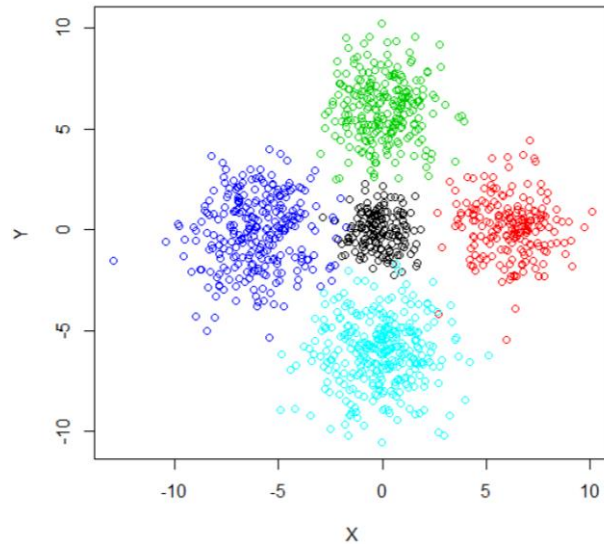
4.5.2.8

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7375	0.0634	0.0354	0.1429	0.7018	0.0126	0.0634	0.0279	0.1544	0.5087	0.3951
3	0.4912	0.3308	0.1419	0.5697	0.3032	0.0421	0.3308	0.1512	0.6428	0.3052	0.2219
4	0.3410	0.5852	0.3233	0.7314	0.2177	0.1042	0.5852	0.3914	0.7555	0.2344	0.2037
5	0.2808	0.6817	0.4689	0.7670	0.2046	0.1988	0.6817	0.5512	0.7475	0.2133	0.2004
6	0.2512	0.7248	0.5500	0.7629	0.2013	0.3082	0.7248	0.6311	0.7064	0.2057	0.2003
7	0.2306	0.7492	0.5984	0.7459	0.2009	0.3962	0.7492	0.6722	0.6780	0.2033	0.1999
8	0.2207	0.7623	0.6301	0.7328	0.1998	0.4635	0.7623	0.6988	0.6551	0.2018	0.2000
9	0.2127	0.7722	0.6515	0.7205	0.1999	0.5129	0.7722	0.7178	0.6380	0.2000	0.2000
10	0.2096	0.7809	0.6682	0.7105	0.1999	0.5478	0.7809	0.7321	0.6157	0.1997	0.2000
15	0.2012	0.8084	0.7184	0.6677	0.2000	0.6413	0.8084	0.7758	0.5670	0.1999	0.2000
20	0.1999	0.8304	0.7499	0.6311	0.2000	0.6893	0.8304	0.8028	0.5343	0.2000	0.2000
25	0.2000	0.8537	0.7768	0.5766	0.2000	0.7251	0.8537	0.8262	0.4942	0.2000	0.2000
30	0.2000	0.8706	0.7998	0.5208	0.2000	0.7544	0.8706	0.8455	0.4485	0.2000	0.2000
35	0.2000	0.8824	0.8205	0.4646	0.2000	0.7796	0.8824	0.8611	0.4012	0.2000	0.2000
40	0.2000	0.8889	0.8382	0.4029	0.2000	0.8020	0.8889	0.8751	0.3525	0.2000	0.2000
45	0.2000	0.8907	0.8537	0.3556	0.2000	0.8212	0.8907	0.8863	0.3153	0.2000	0.2000
48	0.2000	0.8935	0.8621	0.3275	0.2000	0.8317	0.8935	0.8912	0.2900	0.2000	0.2000
50	0.2000	0.8918	0.8674	0.3082	0.2000	0.8380	0.8918	0.8951	0.2769	0.2000	0.2000

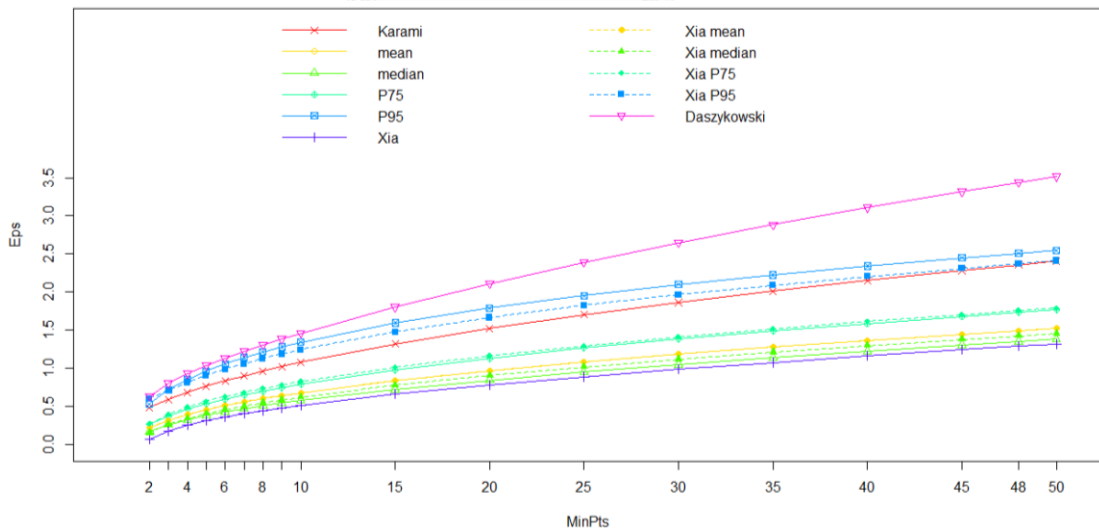
4.5.2.9 ขนาดตัวอย่าง N=1,200 กลุ่มละ 160, 200, 240, 280, 320 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

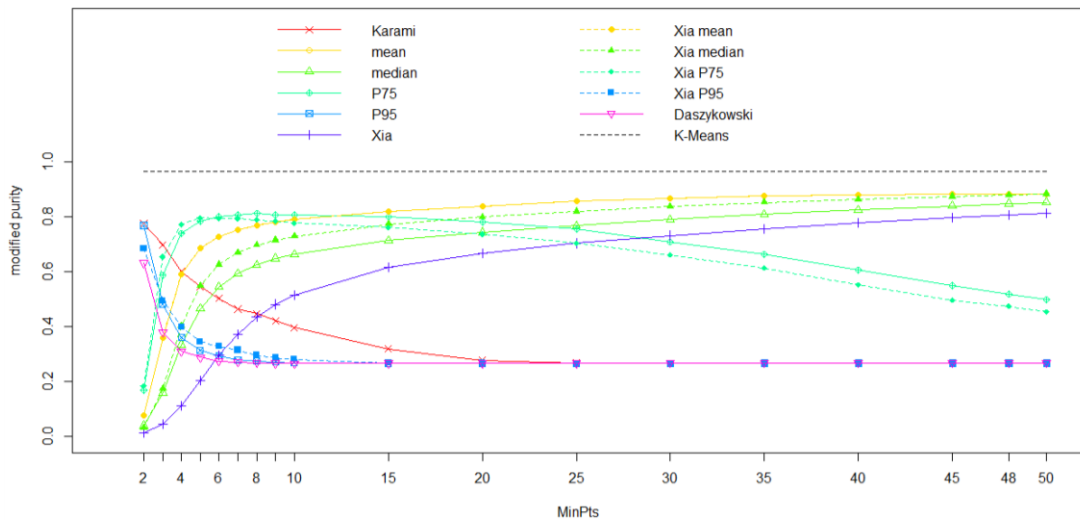
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_4 = \Sigma_5 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



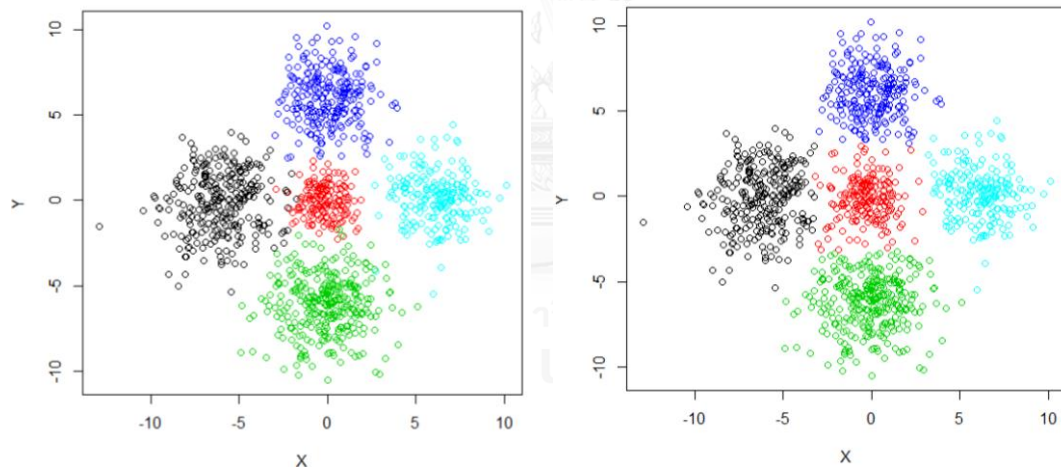
รูปที่ 4.5.2.9.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.9



รูปที่ 4.5.2.9.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.9



รูปที่ 4.5.2.9.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกัน ของข้อมูลกรณีที่ 4.5.2.9



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.9.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.9

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.9.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski กลุ่มที่ประกอบด้วยวิธี P95, วิธี Xia P95 และวิธี Karami กลุ่มที่ประกอบด้วยวิธี P75 และวิธี Xia P75 กลุ่มที่ประกอบด้วยวิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.2.9.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.267 (320/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (320/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 2 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 25 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (320/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 8 สำหรับวิธี P75 และที่ MinPts 5 สำหรับวิธี Xia P75 จากนั้น modified purity มีค่าลดลง

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 15-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.2.9.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.9.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9648117

ตารางที่ 4.5.2.9.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.9

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.4811	0.2130	0.1681	0.2616	0.5283	0.0722	0.2130	0.1498	0.2666	0.5982	0.6287
3	0.5893	0.3158	0.2586	0.3741	0.7145	0.1771	0.3158	0.2633	0.3974	0.7080	0.7983
4	0.6804	0.3917	0.3244	0.4581	0.8512	0.2500	0.3917	0.3400	0.4883	0.8129	0.9269
5	0.7607	0.4540	0.3786	0.5281	0.9622	0.3077	0.4540	0.4015	0.5624	0.9045	1.0358
6	0.8334	0.5078	0.4252	0.5897	1.0567	0.3570	0.5078	0.4542	0.6261	0.9848	1.1330
7	0.9001	0.5560	0.4671	0.6452	1.1394	0.4007	0.5560	0.5011	0.6829	1.0575	1.2223
8	0.9623	0.5998	0.5057	0.6959	1.2136	0.4404	0.5998	0.5437	0.7345	1.1236	1.3056
9	1.0206	0.6400	0.5413	0.7430	1.2793	0.4769	0.6400	0.5828	0.7817	1.1842	1.3845
10	1.0759	0.6776	0.5749	0.7871	1.3399	0.5111	0.6776	0.6194	0.8258	1.2404	1.4602
15	1.3177	0.8377	0.7202	0.9748	1.5934	0.6582	0.8377	0.7754	1.0121	1.4762	1.8035
20	1.5215	0.9687	0.8416	1.1285	1.7903	0.7804	0.9687	0.9037	1.1632	1.6641	2.1103
25	1.7011	1.0824	0.9487	1.2620	1.9541	0.8878	1.0824	1.0155	1.2934	1.8236	2.3897
30	1.8634	1.1844	1.0465	1.3796	2.0978	0.9851	1.1844	1.1161	1.4094	1.9635	2.6460
35	2.0127	1.2779	1.1370	1.4876	2.2249	1.0752	1.2779	1.2085	1.5150	2.0892	2.8841
40	2.1517	1.3649	1.2224	1.5867	2.3400	1.1598	1.3649	1.2949	1.6128	2.2042	3.1074
45	2.2822	1.4468	1.3031	1.6798	2.4478	1.2400	1.4468	1.3763	1.7044	2.3105	3.3185
48	2.3571	1.4940	1.3498	1.7325	2.5092	1.2864	1.4940	1.4232	1.7569	2.3711	3.4403
50	2.4057	1.5247	1.3805	1.7669	2.5491	1.3167	1.5247	1.4539	1.7910	2.4103	3.5195

ตารางที่ 4.5.2.9.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีที่

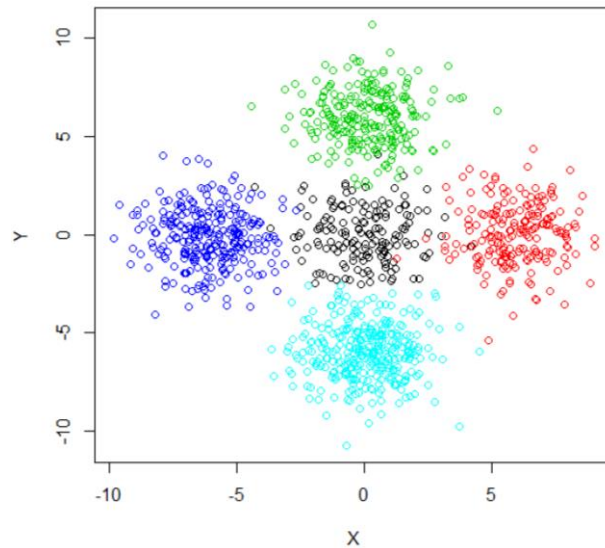
4.5.2.9

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7751	0.0755	0.0381	0.1678	0.7688	0.0127	0.0755	0.0298	0.1821	0.6847	0.6319
3	0.6980	0.3601	0.1582	0.5862	0.4799	0.0434	0.3601	0.1723	0.6544	0.4943	0.3789
4	0.6014	0.5913	0.3340	0.7392	0.3599	0.1108	0.5913	0.4009	0.7722	0.3984	0.3095
5	0.5448	0.6863	0.4645	0.7837	0.3131	0.2032	0.6863	0.5489	0.7966	0.3443	0.2874
6	0.5029	0.7286	0.5444	0.8020	0.2927	0.2950	0.7286	0.6264	0.7941	0.3271	0.2730
7	0.4635	0.7518	0.5931	0.8070	0.2774	0.3715	0.7518	0.6693	0.7941	0.3141	0.2699
8	0.4468	0.7678	0.6241	0.8122	0.2737	0.4353	0.7678	0.6972	0.7885	0.2950	0.2672
9	0.4211	0.7803	0.6479	0.8075	0.2700	0.4808	0.7803	0.7150	0.7846	0.2853	0.2655
10	0.3954	0.7897	0.6629	0.8067	0.2669	0.5142	0.7897	0.7290	0.7771	0.2795	0.2652
15	0.3183	0.8194	0.7133	0.8020	0.2661	0.6163	0.8194	0.7729	0.7613	0.2675	0.2657
20	0.2760	0.8395	0.7435	0.7802	0.2652	0.6670	0.8395	0.7994	0.7368	0.2660	0.2662
25	0.2667	0.8563	0.7682	0.7551	0.2656	0.7036	0.8563	0.8204	0.7044	0.2652	0.2664
30	0.2652	0.8675	0.7896	0.7088	0.2658	0.7322	0.8675	0.8375	0.6613	0.2656	0.2666
35	0.2651	0.8767	0.8087	0.6626	0.2660	0.7566	0.8767	0.8525	0.6134	0.2655	0.2666
40	0.2655	0.8813	0.8256	0.6053	0.2661	0.7781	0.8813	0.8637	0.5512	0.2657	0.2666
45	0.2658	0.8843	0.8401	0.5495	0.2663	0.7971	0.8843	0.8740	0.4968	0.2659	0.2666
48	0.2660	0.8846	0.8480	0.5188	0.2663	0.8076	0.8846	0.8786	0.4738	0.2660	0.2667
50	0.2661	0.8845	0.8529	0.4972	0.2663	0.8141	0.8845	0.8830	0.4535	0.2661	0.2667

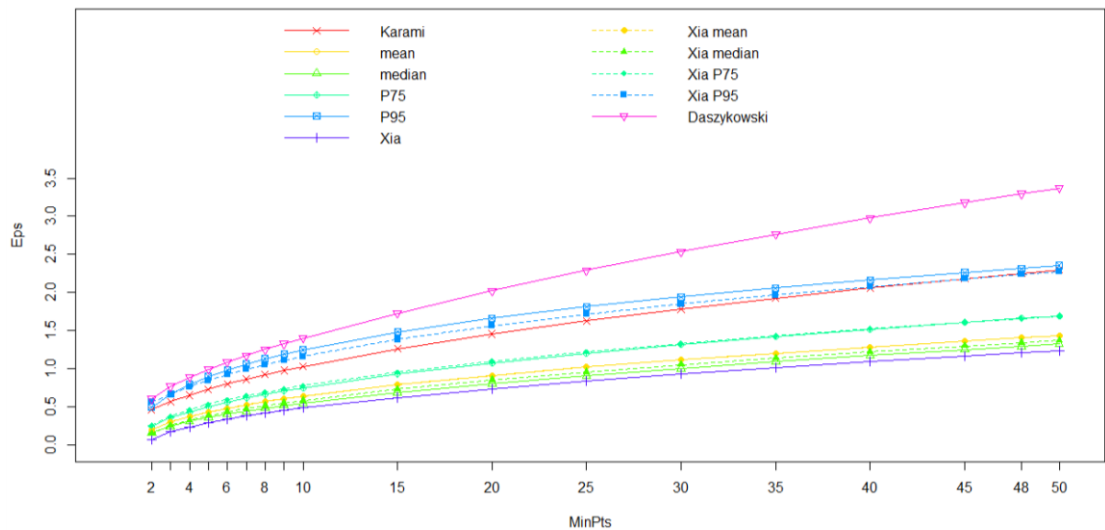
4.5.2.10 ขนาดตัวอย่าง $N=1,200$ กลุ่มละ 160, 200, 240, 280, 320 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

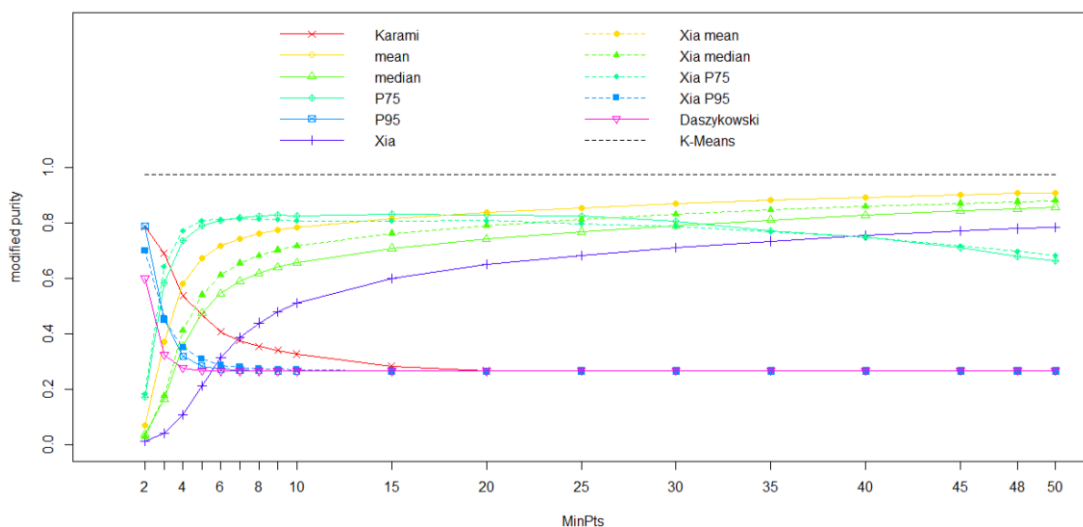
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



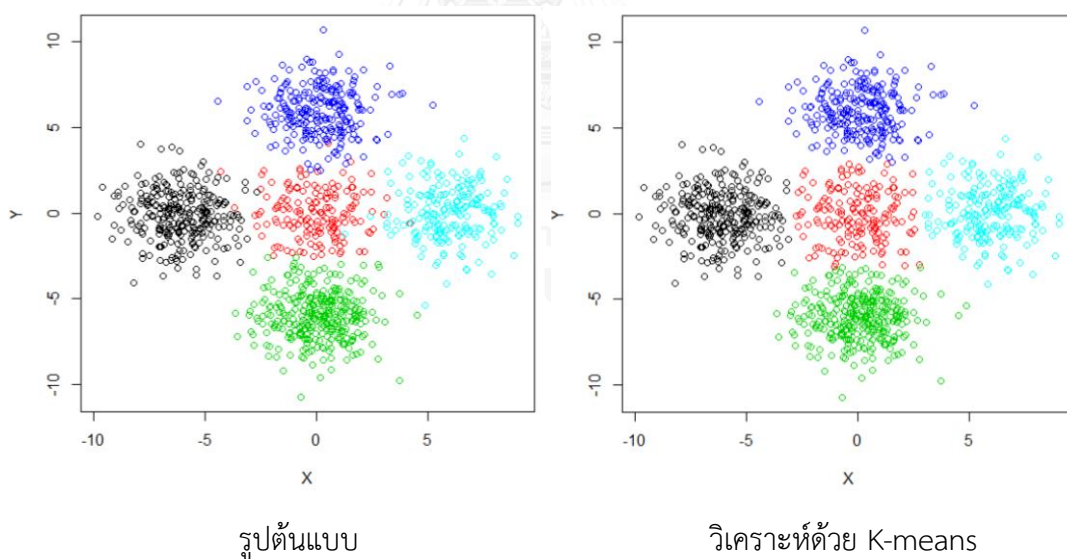
รูปที่ 4.5.2.10.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.10



รูปที่ 4.5.2.10.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.10



รูปที่ 4.5.2.10.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.10



รูปที่ 4.5.2.10.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณี 4.5.2.10

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.10.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี

ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski กลุ่มที่ประกอบด้วยวิธี P95, วิธี Xia P95 และวิธี Karami กลุ่มที่ประกอบด้วยวิธี P75 และวิธี Xia P75 กลุ่มที่ประกอบด้วยวิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.2.10.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.267 (320/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (320/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 2 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (320/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 จนกระทั่งสูงสุดที่ MinPts 15 สำหรับวิธี P75 และสูงสุดที่ MinPts 7 สำหรับวิธี Xia P75 จากนั้น modified purity มีค่าลดลง

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 15-50

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.2.10.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.10.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9747183

ตารางที่ 4.5.2.10.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี 4.5.2.10

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.4599	0.1996	0.1587	0.2468	0.4914	0.0680	0.1996	0.1406	0.2498	0.5600	0.6022
3	0.5633	0.2965	0.2444	0.3529	0.6662	0.1673	0.2965	0.2477	0.3729	0.6623	0.7647
4	0.6504	0.3677	0.3063	0.4332	0.7917	0.2356	0.3677	0.3196	0.4583	0.7612	0.8876
5	0.7272	0.4262	0.3578	0.4998	0.8942	0.2901	0.4262	0.3774	0.5277	0.8466	0.9920
6	0.7966	0.4769	0.4029	0.5587	0.9818	0.3367	0.4769	0.4271	0.5877	0.9221	1.0851
7	0.8604	0.5221	0.4433	0.6112	1.0579	0.3779	0.5221	0.4712	0.6409	0.9898	1.1707
8	0.9199	0.5632	0.4799	0.6594	1.1269	0.4153	0.5632	0.5112	0.6892	1.0514	1.2505
9	0.9757	0.6012	0.5139	0.7044	1.1888	0.4498	0.6012	0.5481	0.7338	1.1084	1.3261
10	1.0284	0.6365	0.5460	0.7465	1.2452	0.4821	0.6365	0.5826	0.7751	1.1609	1.3984
15	1.2596	0.7874	0.6854	0.9267	1.4800	0.6204	0.7874	0.7295	0.9507	1.3838	1.7272
20	1.4544	0.9111	0.8021	1.0746	1.6612	0.7353	0.9111	0.8505	1.0936	1.5622	2.0214
25	1.6261	1.0184	0.9056	1.2024	1.8110	0.8362	1.0184	0.9558	1.2165	1.7136	2.2893
30	1.7813	1.1146	0.9999	1.3168	1.9411	0.9274	1.1146	1.0504	1.3261	1.8469	2.5349
35	1.9240	1.2028	1.0875	1.4208	2.0574	1.0118	1.2028	1.1374	1.4261	1.9671	2.7628
40	2.0569	1.2849	1.1703	1.5172	2.1633	1.0909	1.2849	1.2186	1.5187	2.0773	2.9765
45	2.1816	1.3621	1.2489	1.6073	2.2600	1.1656	1.3621	1.2951	1.6056	2.1799	3.1787
48	2.2532	1.4066	1.2941	1.6591	2.3157	1.2088	1.4066	1.3392	1.6555	2.2386	3.2953
50	2.2996	1.4356	1.3238	1.6926	2.3515	1.2370	1.4356	1.3680	1.6879	2.2765	3.3713

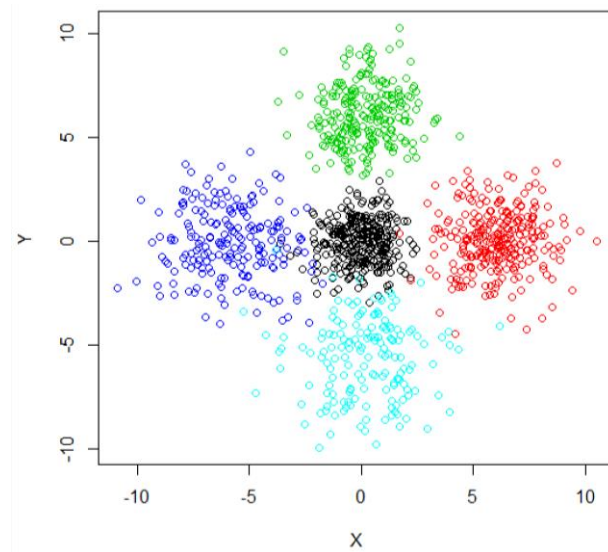
ตารางที่ 4.5.2.10.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี
ที่ 4.5.2.10

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7876	0.0691	0.0361	0.1723	0.7889	0.0124	0.0691	0.0284	0.1832	0.7005	0.6001
3	0.6905	0.3702	0.1645	0.5842	0.4522	0.0411	0.3702	0.1751	0.6433	0.4537	0.3239
4	0.5381	0.5810	0.3541	0.7361	0.3194	0.1076	0.5810	0.4107	0.7729	0.3493	0.2765
5	0.4686	0.6723	0.4757	0.7895	0.2829	0.2131	0.6723	0.5398	0.8087	0.3095	0.2673
6	0.4071	0.7171	0.5453	0.8100	0.2718	0.3139	0.7171	0.6114	0.8121	0.2857	0.2665
7	0.3749	0.7440	0.5892	0.8198	0.2692	0.3869	0.7440	0.6544	0.8166	0.2792	0.2658
8	0.3543	0.7614	0.6189	0.8250	0.2677	0.4393	0.7614	0.6817	0.8145	0.2741	0.2654
9	0.3399	0.7746	0.6401	0.8294	0.2665	0.4798	0.7746	0.7019	0.8127	0.2718	0.2657
10	0.3269	0.7840	0.6562	0.8264	0.2657	0.5122	0.7840	0.7168	0.8076	0.2705	0.2656
15	0.2837	0.8170	0.7090	0.8327	0.2655	0.6001	0.8170	0.7625	0.8080	0.2663	0.2663
20	0.2670	0.8381	0.7424	0.8305	0.2659	0.6500	0.8381	0.7920	0.8095	0.2655	0.2665
25	0.2665	0.8561	0.7690	0.8256	0.2662	0.6842	0.8561	0.8132	0.7989	0.2659	0.2666
30	0.2659	0.8713	0.7916	0.8067	0.2663	0.7119	0.8713	0.8317	0.7879	0.2661	0.2666
35	0.2662	0.8831	0.8114	0.7726	0.2664	0.7353	0.8831	0.8472	0.7690	0.2663	0.2667
40	0.2663	0.8942	0.8283	0.7482	0.2665	0.7549	0.8942	0.8601	0.7504	0.2663	0.2667
45	0.2664	0.9030	0.8439	0.7107	0.2665	0.7712	0.9030	0.8714	0.7175	0.2664	0.2667
48	0.2665	0.9075	0.8521	0.6799	0.2666	0.7801	0.9075	0.8778	0.6981	0.2665	0.2667
50	0.2665	0.9100	0.8576	0.6637	0.2666	0.7856	0.9100	0.8815	0.6826	0.2665	0.2667

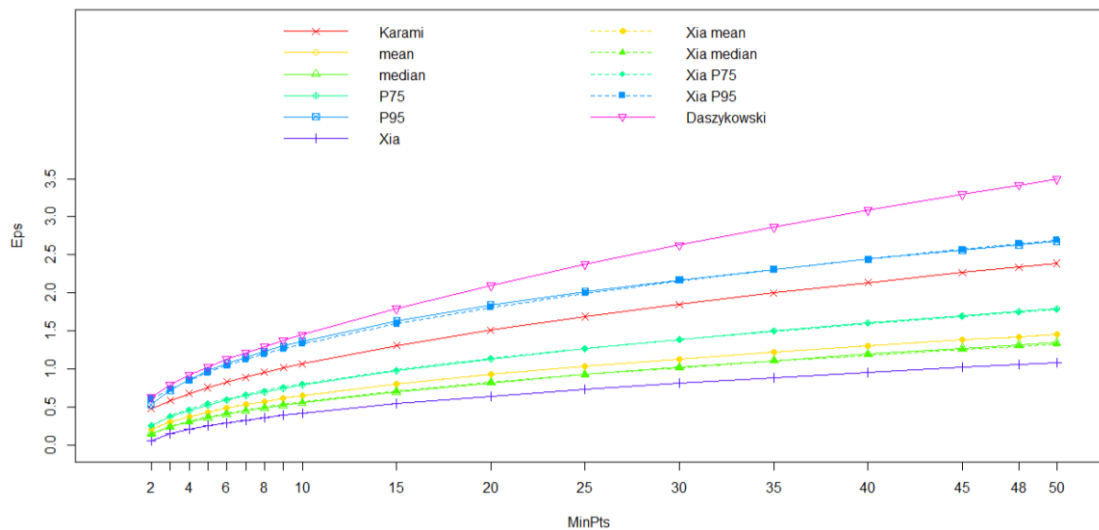
4.5.2.11 ขนาดตัวอย่าง $N=1,200$ กลุ่มละ 320, 280, 240, 200, 160 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

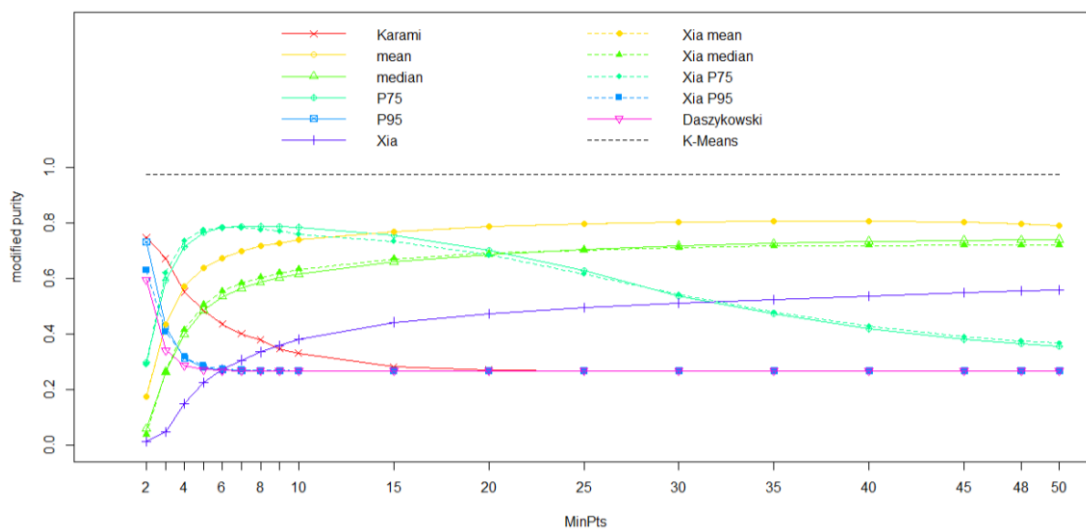
และมีความแปรปรวนไม่เท่ากัน $\Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \Sigma_3 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_4 = \Sigma_5 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$



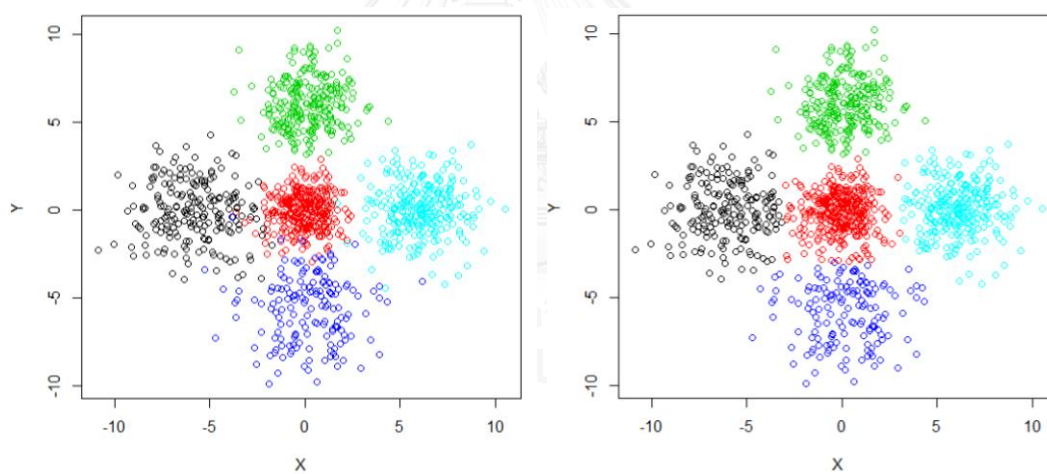
รูปที่ 4.5.2.11.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณีที่ 4.5.2.11



รูปที่ 4.5.2.11.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.11



รูปที่ 4.5.2.11.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.11



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.11.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.11

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.11.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี

ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski, วิธี P95, วิธี Xia P95, วิธี Karami, วิธี Xia P75, วิธี P75, วิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ โดยที่วิธี P95 ใกล้เคียงกับวิธี Xia P95 วิธี P75 ใกล้เคียงกับวิธี Xia P75 วิธี mean เหมือนกับวิธี Xia mean และวิธี median ใกล้เคียงกับวิธี Xia median

จากรูปที่ 4.5.2.11.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.267 (320/1200) ซึ่งคือกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 10 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (320/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าตั้งแต่ช่วง MinPts 2 เป็นต้นไป ค่า modified purity ลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 20 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (320/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 8 สำหรับวิธี P75 และที่ MinPts 7 สำหรับวิธี Xia P75 จากนั้น modified purity มีค่าลดลง

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 15-50 ยกเว้นวิธี mean และวิธี Xia mean ลดลงเล็กน้อยที่ช่วงปลายของ MinPts

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.2.11.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.11.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9749033

ตารางที่ 4.5.2.11.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีที่ 4.5.2.11

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.4775	0.2034	0.1535	0.2538	0.5308	0.0585	0.2034	0.1344	0.2516	0.6067	0.6241
3	0.5848	0.3021	0.2403	0.3691	0.7208	0.1450	0.3021	0.2397	0.3818	0.7327	0.7926
4	0.6753	0.3742	0.3039	0.4545	0.8579	0.2042	0.3742	0.3094	0.4714	0.8510	0.9202
5	0.7550	0.4335	0.3574	0.5259	0.9704	0.2516	0.4335	0.3654	0.5444	0.9530	1.0285
6	0.8270	0.4849	0.4040	0.5877	1.0690	0.2922	0.4849	0.4135	0.6075	1.0430	1.1248
7	0.8933	0.5307	0.4457	0.6439	1.1535	0.3282	0.5307	0.4563	0.6636	1.1236	1.2136
8	0.9550	0.5724	0.4842	0.6945	1.2307	0.3609	0.5724	0.4950	0.7145	1.1972	1.2962
9	1.0129	0.6108	0.5200	0.7422	1.2996	0.3912	0.6108	0.5309	0.7614	1.2649	1.3747
10	1.0677	0.6467	0.5535	0.7862	1.3643	0.4195	0.6467	0.5642	0.8050	1.3279	1.4495
15	1.3077	0.7992	0.6977	0.9752	1.6308	0.5410	0.7992	0.7065	0.9903	1.5939	1.7906
20	1.5099	0.9240	0.8182	1.1314	1.8411	0.6417	0.9240	0.8233	1.1412	1.8089	2.0951
25	1.6882	1.0324	0.9251	1.2661	2.0178	0.7301	1.0324	0.9251	1.2718	1.9932	2.3724
30	1.8493	1.1295	1.0211	1.3869	2.1716	0.8103	1.1295	1.0166	1.3885	2.1567	2.6271
35	1.9975	1.2185	1.1098	1.4972	2.3114	0.8842	1.2185	1.1005	1.4952	2.3052	2.8635
40	2.1354	1.3015	1.1926	1.6003	2.4407	0.9535	1.3015	1.1790	1.5944	2.4428	3.0851
45	2.2649	1.3798	1.2710	1.6969	2.5601	1.0191	1.3798	1.2530	1.6880	2.5723	3.2948
48	2.3392	1.4248	1.3161	1.7523	2.6291	1.0569	1.4248	1.2956	1.7417	2.6466	3.4159
50	2.3874	1.4541	1.3454	1.7882	2.6737	1.0816	1.4541	1.3234	1.7767	2.6947	3.4949

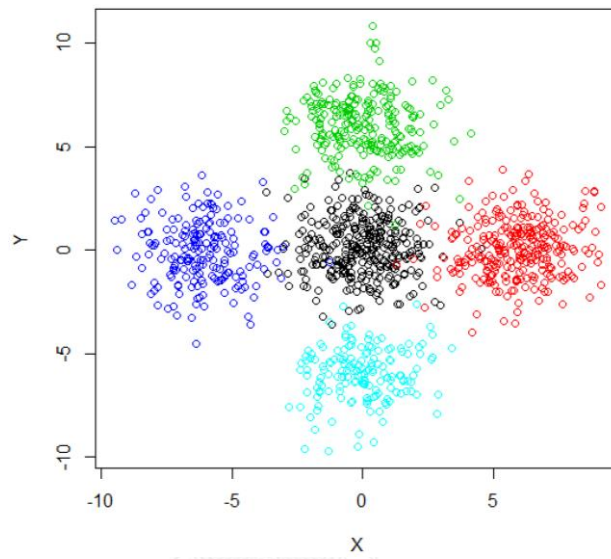
ตารางที่ 4.5.2.11.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี
ที่ 4.5.2.11

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.7493	0.1736	0.0575	0.2972	0.7323	0.0118	0.1736	0.0374	0.2919	0.6292	0.5960
3	0.6724	0.4333	0.2638	0.5951	0.4265	0.0473	0.4333	0.2626	0.6211	0.4090	0.3411
4	0.5522	0.5730	0.3995	0.7148	0.3129	0.1497	0.5730	0.4154	0.7373	0.3182	0.2862
5	0.4851	0.6371	0.4856	0.7661	0.2795	0.2247	0.6371	0.5046	0.7750	0.2852	0.2711
6	0.4362	0.6743	0.5350	0.7832	0.2707	0.2715	0.6743	0.5539	0.7831	0.2727	0.2682
7	0.4005	0.6986	0.5651	0.7864	0.2683	0.3060	0.6986	0.5832	0.7845	0.2700	0.2670
8	0.3786	0.7169	0.5865	0.7894	0.2678	0.3347	0.7169	0.6044	0.7779	0.2691	0.2669
9	0.3471	0.7289	0.6040	0.7872	0.2672	0.3595	0.7289	0.6203	0.7711	0.2682	0.2669
10	0.3313	0.7388	0.6169	0.7845	0.2675	0.3811	0.7388	0.6331	0.7598	0.2678	0.2669
15	0.2810	0.7699	0.6613	0.7572	0.2667	0.4426	0.7699	0.6711	0.7353	0.2667	0.2667
20	0.2701	0.7868	0.6875	0.7012	0.2667	0.4727	0.7868	0.6916	0.6848	0.2667	0.2667
25	0.2671	0.7984	0.7045	0.6277	0.2667	0.4943	0.7984	0.7027	0.6172	0.2667	0.2667
30	0.2667	0.8053	0.7178	0.5350	0.2667	0.5108	0.8053	0.7115	0.5433	0.2667	0.2667
35	0.2667	0.8058	0.7264	0.4742	0.2667	0.5246	0.8058	0.7168	0.4787	0.2667	0.2667
40	0.2667	0.8073	0.7338	0.4191	0.2667	0.5376	0.8073	0.7191	0.4300	0.2667	0.2667
45	0.2667	0.8027	0.7373	0.3811	0.2667	0.5484	0.8027	0.7200	0.3906	0.2667	0.2667
48	0.2667	0.7960	0.7390	0.3659	0.2667	0.5547	0.7960	0.7201	0.3749	0.2667	0.2667
50	0.2667	0.7903	0.7405	0.3568	0.2667	0.5581	0.7903	0.7208	0.3674	0.2667	0.2667

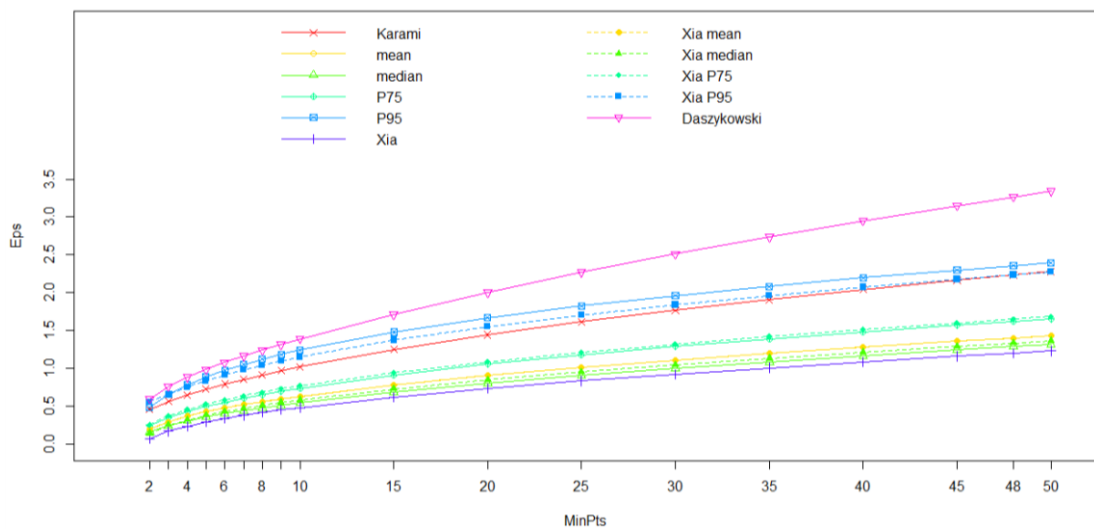
4.5.2.12 ขนาดตัวอย่าง N=1,200 กลุ่มละ 320, 280, 240, 200, 160 ตามลำดับ

โดยที่ $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}, \mu_4 = \begin{bmatrix} -6 \\ 0 \end{bmatrix}, \mu_5 = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$

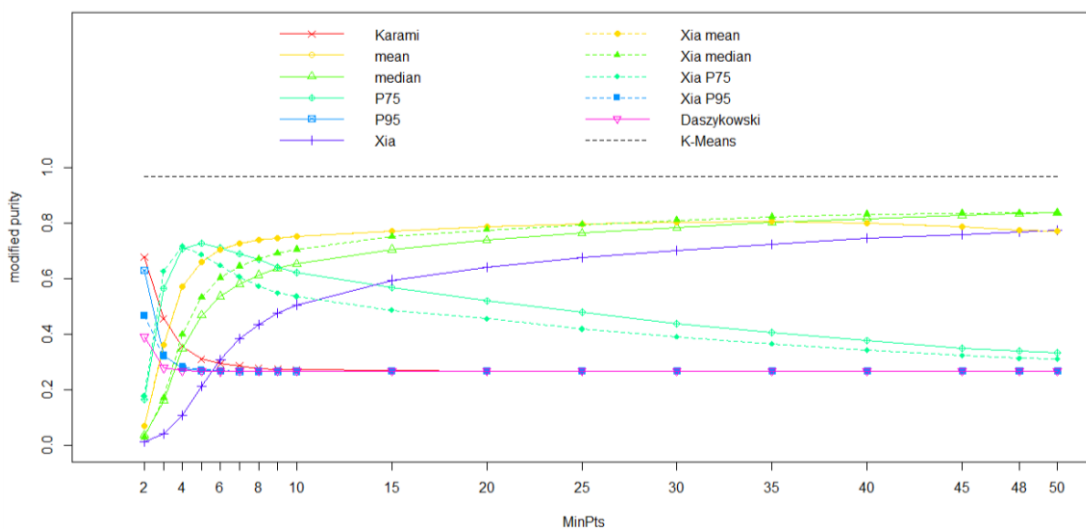
และมีความแปรปรวนเท่ากัน $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4 = \Sigma_5 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$



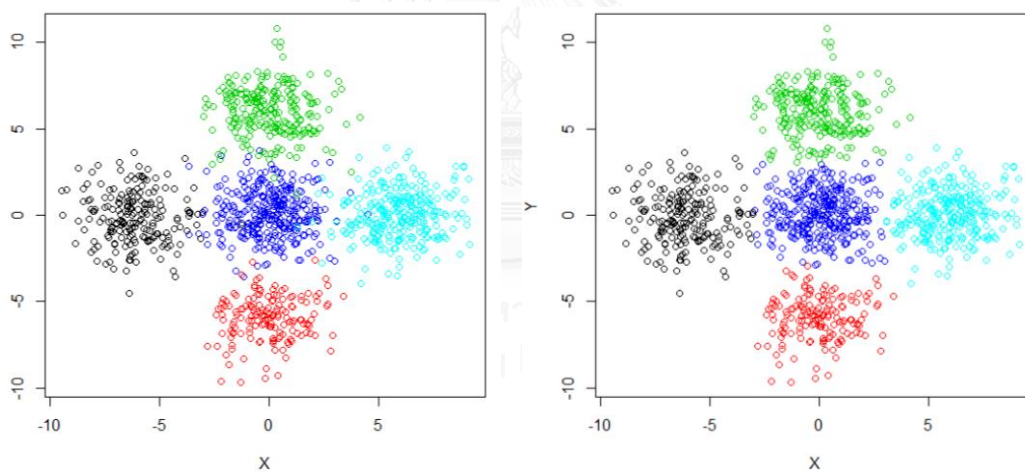
รูปที่ 4.5.2.12.1 ตัวอย่างแผนภาพการกระจายของข้อมูลกรณี 4.5.2.12



รูปที่ 4.5.2.12.2 กราฟแสดงการเปรียบเทียบค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณี 4.5.2.12



รูปที่ 4.5.2.12.3 กราฟแสดงการเปรียบเทียบค่า modified purity ของ 12 วิธี ที่ค่า MinPts ต่างกันของข้อมูลกรณีที่ 4.5.2.12



รูปต้นแบบ

วิเคราะห์ด้วย K-means

รูปที่ 4.5.2.12.4 แผนภาพกระจายของข้อมูล (ซ้าย) และแผนภาพผลการวิเคราะห์กลุ่มด้วย K-means (ขวา) ของข้อมูลกรณีที่ 4.5.2.12

สำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปรตามขนาดตัวอย่าง ค่าเฉลี่ย และความแปรปรวนข้างต้น พบว่าจากรูปที่ 4.5.2.12.2 เมื่อค่า MinPts สูงขึ้น จะทำให้ค่า Eps สูงขึ้นทุกกรณี ซึ่งมีระดับค่า Eps เรียงลำดับจากสูงไปต่ำได้ดังนี้ วิธี Daszykowski กลุ่มที่ประกอบด้วยวิธี P95, วิธี Xia P95 และวิธี Karami กลุ่มที่ประกอบด้วยวิธี P75 และวิธี Xia P75 กลุ่มที่ประกอบด้วยวิธี mean, วิธี Xia mean, วิธี Xia median, วิธี median, และวิธี Xia ตามลำดับ

จากรูปที่ 4.5.2.12.3 แสดงการเปรียบเทียบค่า modified purity ของวิธีต่าง ๆ ที่ค่า MinPts ต่างกัน พบว่า

สำหรับวิธี Daszykowski พบว่าเมื่อค่า MinPts สูงขึ้น ค่า modified purity จะลดลง ช่วง MinPts 2-3 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นค่า modified purity ค่อย ๆ ลดลง และลู่เข้าสู่ 0.267 (320/1200) ซึ่งเป็นกรณีที่ทุกจุดถูกรวมเข้าเป็นกลุ่มเดียวกัน เนื่องจาก Eps มีค่าสูงเกินไป

สำหรับวิธี P95 และ Xia P95 พบว่าช่วง MinPts 2-4 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 6 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (320/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี Karami พบว่าช่วง MinPts 2-5 ค่า modified purity ลดลงอย่างรวดเร็ว จากนั้นลดลงอย่างต่อเนื่อง จนกระทั่งที่ MinPts 15 เป็นต้นไป ค่า modified purity ลู่เข้าสู่ 0.267 (320/1200) ด้วยเหตุผลเดียวกับวิธี Daszykowski

สำหรับวิธี P75 และ Xia P75 พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 2 วิธีอยู่ในระดับที่ต่ำ จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-4 จนกระทั่งสูงสุดที่ MinPts 5 สำหรับวิธี P75 และที่ MinPts 4 สำหรับวิธี Xia P75 จากนั้น modified purity มีค่าลดลงและมีแนวโน้มที่จะลู่เข้าสู่ 0.267

สำหรับวิธี mean, median Xia mean และ Xia median พบว่าที่ค่า MinPts 2 ค่า modified purity ของทั้ง 4 วิธีอยู่ในระดับที่ต่ำมาก จากนั้นเพิ่มขึ้นอย่างรวดเร็วที่ MinPts 2-5 แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างช้า ๆ ตั้งแต่ MinPts 15-50 ยกเว้นวิธี mean และวิธี Xia mean ลดลงเล็กน้อยในช่วงปลายของ MinPts

สำหรับวิธี Xia พบว่าที่ค่า MinPts 2 ค่า modified purity อยู่ในระดับที่ต่ำมาก แล้วค่อย ๆ เพิ่มสูงขึ้นอย่างต่อเนื่อง

จากรูปที่ 4.5.2.12.4 พบว่าการวิเคราะห์ข้อมูลด้วย K-means ให้ผลการแบ่งกลุ่มใกล้เคียงกับแผนภาพกระจายของข้อมูลต้นแบบ ซึ่งสอดคล้องกับเมื่อพิจารณาจากรูปที่ 4.5.2.12.3 พบว่าค่า modified purity ของ K-means ทั้ง 500 รอบ มีค่าเฉลี่ยเป็น 0.9687867

ตารางที่ 4.5.2.12.1 ค่า Eps ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณีที่ 4.5.2.12

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.4560	0.1978	0.1574	0.2439	0.4813	0.0673	0.1978	0.1393	0.2475	0.5548	0.5967
3	0.5585	0.2936	0.2423	0.3483	0.6529	0.1662	0.2936	0.2455	0.3692	0.6545	0.7576
4	0.6449	0.3642	0.3046	0.4272	0.7795	0.2347	0.3642	0.3171	0.4537	0.7511	0.8796
5	0.7211	0.4224	0.3560	0.4925	0.8830	0.2890	0.4224	0.3746	0.5227	0.8360	0.9833
6	0.7899	0.4728	0.4007	0.5496	0.9718	0.3353	0.4728	0.4240	0.5821	0.9107	1.0756
7	0.8532	0.5177	0.4412	0.6008	1.0497	0.3763	0.5177	0.4678	0.6351	0.9781	1.1602
8	0.9121	0.5586	0.4779	0.6482	1.1179	0.4134	0.5586	0.5076	0.6831	1.0396	1.2394
9	0.9674	0.5964	0.5122	0.6917	1.1822	0.4478	0.5964	0.5444	0.7275	1.0964	1.3143
10	1.0197	0.6317	0.5441	0.7328	1.2408	0.4797	0.6317	0.5786	0.7688	1.1494	1.3858
15	1.2489	0.7819	0.6831	0.9086	1.4788	0.6171	0.7819	0.7248	0.9437	1.3717	1.7117
20	1.4421	0.9051	0.8001	1.0526	1.6655	0.7310	0.9051	0.8451	1.0861	1.5507	2.0031
25	1.6123	1.0121	0.9033	1.1768	1.8222	0.8310	1.0121	0.9499	1.2091	1.7031	2.2685
30	1.7662	1.1083	0.9970	1.2877	1.9614	0.9215	1.1083	1.0443	1.3190	1.8381	2.5120
35	1.9077	1.1964	1.0838	1.3890	2.0827	1.0050	1.1964	1.1309	1.4192	1.9601	2.7381
40	2.0395	1.2785	1.1660	1.4824	2.1950	1.0832	1.2785	1.2118	1.5122	2.0722	2.9499
45	2.1632	1.3556	1.2430	1.5692	2.2990	1.1572	1.3556	1.2879	1.5993	2.1764	3.1504
48	2.2341	1.4001	1.2880	1.6196	2.3581	1.1999	1.4001	1.3318	1.6493	2.2360	3.2658
50	2.2802	1.4290	1.3172	1.6523	2.3961	1.2278	1.4290	1.3604	1.6819	2.2747	3.3413

ตารางที่ 4.5.2.12.2 ค่า modified purity ที่ได้จาก 11 วิธี ที่ค่า MinPts ต่างกันสำหรับข้อมูลกรณี
ที่ 4.5.2.12

MinPts	วิธี										
	Karami	mean	median	P75	P95	Xia	Xia mean	Xia median	Xia P75	Xia P95	Daszykowski
2	0.6771	0.0681	0.0356	0.1643	0.6292	0.0123	0.0681	0.0281	0.1770	0.4668	0.3895
3	0.4556	0.3611	0.1595	0.5664	0.3235	0.0409	0.3611	0.1700	0.6277	0.3217	0.2791
4	0.3561	0.5704	0.3481	0.7077	0.2764	0.1073	0.5704	0.3995	0.7183	0.2820	0.2682
5	0.3099	0.6619	0.4677	0.7280	0.2689	0.2110	0.6619	0.5322	0.6868	0.2723	0.2667
6	0.2941	0.7049	0.5372	0.7115	0.2669	0.3086	0.7049	0.6037	0.6490	0.2690	0.2665
7	0.2865	0.7268	0.5815	0.6908	0.2664	0.3837	0.7268	0.6447	0.6070	0.2669	0.2666
8	0.2773	0.7413	0.6133	0.6688	0.2665	0.4359	0.7413	0.6729	0.5734	0.2666	0.2666
9	0.2738	0.7464	0.6372	0.6423	0.2665	0.4754	0.7464	0.6917	0.5481	0.2666	0.2666
10	0.2717	0.7541	0.6534	0.6227	0.2666	0.5063	0.7541	0.7059	0.5365	0.2667	0.2666
15	0.2679	0.7713	0.7058	0.5673	0.2666	0.5953	0.7713	0.7516	0.4865	0.2666	0.2667
20	0.2668	0.7892	0.7388	0.5195	0.2667	0.6425	0.7892	0.7766	0.4558	0.2666	0.2667
25	0.2666	0.7981	0.7643	0.4795	0.2667	0.6762	0.7981	0.7954	0.4192	0.2667	0.2667
30	0.2667	0.8026	0.7845	0.4379	0.2667	0.7026	0.8026	0.8116	0.3918	0.2667	0.2667
35	0.2667	0.8058	0.8029	0.4061	0.2667	0.7252	0.8058	0.8231	0.3654	0.2667	0.2667
40	0.2667	0.7993	0.8175	0.3766	0.2667	0.7454	0.7993	0.8321	0.3429	0.2667	0.2667
45	0.2667	0.7885	0.8291	0.3491	0.2667	0.7606	0.7885	0.8357	0.3226	0.2667	0.2667
48	0.2667	0.7753	0.8345	0.3399	0.2667	0.7695	0.7753	0.8382	0.3131	0.2667	0.2667
50	0.2667	0.7726	0.8379	0.3330	0.2667	0.7748	0.7726	0.8396	0.3104	0.2667	0.2667

4.6 การเปรียบเทียบภายใต้สถานการณ์จำนวนกลุ่มและขนาดตัวอย่างเท่ากันสำหรับข้อมูลที่มีการแจกแจงปกติ

4.6.1 จำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 300 ประกอบด้วย

4.6.1.1 ความแปรปรวนไม่เท่ากัน

4.6.1.1.1 กลุ่มย่อยละ 50, 100, 150 ตามลำดับ (กรณี 4.5.1.1) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 20, 25, 30, 35, 40, 45

วิธี median ที่ MinPts 30, 35, 40, 45

วิธี Xia median ที่ MinPts 25, 30, 35, 40, 45

วิธี Xia ที่ MinPts 40

4.6.1.1.2 กลุ่มย่อยละ 100, 100, 100 ตามลำดับ (กรณี 4.5.1.3) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 20, 25, 30, 35, 40

วิธี median และวิธี Xia median ที่ MinPts 30, 35, 40, 45

4.6.1.1.3 กลุ่มย่อยละ 150, 100, 50 ตามลำดับ (กรณี 4.5.1.5) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 25, 30, 35

วิธี P75, Xia P75 ที่ MinPts 5-10, 12

วิธี P95 ที่ MinPts 2

วิธี Karami ที่ MinPts 3-4

สรุปได้ว่า สำหรับจำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 300 โดยที่ความแปรปรวนไม่เท่ากัน วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ที่ MinPts 25, 30, 35

4.6.1.2 ความแปรปรวนเท่ากัน

4.6.1.2.1 กลุ่มย่อยละ 50, 100, 150 ตามลำดับ (กรณี 4.5.1.2) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 15, 20, 25, 30, 50

วิธี P75 ที่ MinPts 6-10, 12, 15, 20, 25, 30

วิธี Xia P75 ที่ MinPts 5-10, 12, 15, 20, 25, 30, 35, 40

วิธี P95 และวิธี Daszykowski ที่ MinPts 2

วิธี Karami ที่ MinPts 3-5

4.6.1.2.2 กลุ่มย่อยละ 100, 100, 100 ตามลำดับ (กรณี 4.5.1.4) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 15, 20, 25, 30

วิธี median ที่ MinPts 25, 30, 35, 40, 45

วิธี Xia median ที่ MinPts 20, 25, 30, 35, 40

วิธี Xia ที่ MinPts 30, 35, 40, 45, 50

4.6.1.2.3 กลุ่มย่อยละ 150, 100, 50 ตามลำดับ (กรณี 4.5.1.6) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 8-10, 12, 15, 20

วิธี median ที่ MinPts 20, 25, 30

วิธี Xia median ที่ MinPts 15, 20, 25

วิธี Xia ที่ MinPts 45, 50

สรุปได้ว่า สำหรับจำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 300 โดยที่
ความแปรปรวนเท่ากัน วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ที่ MinPts 15, 20

4.6.2 จำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 1200 ประกอบด้วย

4.6.2.1 ความแปรปรวนไม่เท่ากัน

4.6.2.1.1 กลุ่มย่อยละ 200, 400, 600 ตามลำดับ (กรณี 4.5.1.7) วิธีการ
ที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 30, 35, 40, 45, 48, 50

วิธี P75 ที่ MinPts 15, 20, 25, 30, 35, 40, 45, 48, 50

วิธี Xia P75 ที่ MinPts 15, 20, 25, 30, 35, 40

วิธี Xia median ที่ MinPts 45, 48, 50

4.6.2.1.2 กลุ่มย่อยละ 400, 400, 400 ตามลำดับ (กรณี 4.5.1.9) วิธีการ
ที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 25, 30, 35, 40, 45, 48,
50

วิธี P75 ที่ MinPts 8-10, 15, 20, 25, 30, 35, 40

วิธี Xia P75 ที่ MinPts 6-10, 15, 20, 25, 30, 35

4.6.2.1.3 กลุ่มย่อยละ 600, 400, 200 ตามลำดับ (กรณี 4.5.1.11)
วิธีการที่เหมาะสมคือ

วิธี P75 ที่ MinPts 8-10, 15, 20, 25, 30

วิธี Xia P75 ที่ MinPts 7-10, 15, 20, 25, 30, 35, 40

สรุปได้ว่า สำหรับจำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 1200 โดยที่
ความแปรปรวนไม่เท่ากัน วิธีการที่เหมาะสมคือวิธี P75 และวิธี Xia P75 ที่ MinPts 15, 20, 25, 30

4.6.2.2 ความแปรปรวนเท่ากัน

4.6.2.2.1 กลุ่มย่อยละ 200, 400, 600 ตามลำดับ (กรณี 4.5.1.8) วิธีการ
ที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 45, 48, 50

วิธี P75 ที่ MinPts 9-10, 15, 20, 25, 30, 35, 40, 45, 48, 50

วิธี Xia P75 ที่ MinPts 8-10, 15, 20, 25, 30, 35, 40, 45, 48,

50

4.6.2.2.2 กลุ่มย่อยละ 400, 400, 400 ตามลำดับ (กรณี 4.5.1.10)

วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 20, 25, 30, 35, 40, 45, 48, 50

วิธี Xia median ที่ MinPts 35, 40, 45, 48, 50

4.6.2.2.3 กลุ่มย่อยละ 600, 400, 200 ตามลำดับ (กรณี 4.5.1.12)

วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 15, 20, 25, 30, 35, 40, 45, 48, 50

วิธี Xia median ที่ MinPts 30, 35, 40, 45, 48, 50

สรุปได้ว่า สำหรับจำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 1200 โดยที่ความแปรปรวนเท่ากัน วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ที่ MinPts 45 เป็นต้นไป

4.6.3 จำนวนกลุ่มเท่ากับ 5 และมีขนาดตัวอย่าง 300 ประกอบด้วย

4.6.3.1 กลุ่มย่อยละ 60, 60, 60, 60, 60 ตามลำดับ โดยที่ความแปรปรวนไม่เท่ากัน (กรณี 4.5.2.1) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 9-10, 12, 15, 20

วิธี median และวิธี Xia median ที่ MinPts 15, 20, 25

4.6.3.2 กลุ่มย่อยละ 60, 60, 60, 60, 60 ตามลำดับ โดยที่ความแปรปรวนเท่ากัน (กรณี 4.5.2.2) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 8-10, 12, 15, 20

วิธี median ที่ MinPts 12, 15, 20, 25, 30

วิธี Xia median ที่ MinPts 10, 12, 15, 20, 25

4.6.3.3 กลุ่มย่อยละ 40, 50, 60, 70, 80 ตามลำดับ โดยที่ความแปรปรวนไม่เท่ากัน (กรณี 4.5.2.3) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 9-10, 12, 15, 20

วิธี median ที่ MinPts 15, 20, 25, 30

วิธี Xia median ที่ MinPts 12, 15, 20, 25

วิธี Xia ที่ MinPts 20, 25, 30, 35

4.6.3.4 กลุ่มย่อยละ 40, 50, 60, 70, 80 ตามลำดับ โดยที่ความแปรปรวนเท่ากัน (กรณี 4.5.2.4) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 12, 15, 20, 25, 30, 35

วิธี median และวิธี Xia median ที่ MinPts 20, 25, 30, 35

4.6.3.5 กลุ่มย่อยละ 80, 70, 60, 50, 40 ตามลำดับ โดยที่ความแปรปรวนไม่เท่ากัน (กรณี 4.5.2.5) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 8-10, 12, 15

วิธี P75 ที่ MinPts 5-7

วิธี Xia P75 ที่ MinPts 4-7

4.6.3.6 กลุ่มย่อยละ 80, 70, 60, 50, 40 ตามลำดับ โดยที่ความแปรปรวนเท่ากัน (กรณี 4.5.2.6) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 6-10, 12, 15

วิธี median ที่ MinPts 9-10, 12, 15, 20

วิธี Xia median ที่ MinPts 8-10, 12, 15, 20

วิธี Xia ที่ MinPts 20, 25

สรุปได้ว่า สำหรับจำนวนกลุ่มเท่ากับ 5 และมีขนาดตัวอย่าง 300 วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ที่ MinPts 12, 15

4.6.4 จำนวนกลุ่มเท่ากับ 5 และมีขนาดตัวอย่าง 1200 ประกอบด้วย

4.6.4.1 กลุ่มย่อยละ 240, 240, 240, 240, 240 ตามลำดับ โดยที่ความแปรปรวนไม่เท่ากัน (กรณี 4.5.2.7) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 20, 25, 30, 35, 40, 45, 48, 50

วิธี median ที่ MinPts 45, 48, 50

วิธี Xia median ที่ MinPts 40, 45, 48, 50

4.6.4.2 กลุ่มย่อยละ 240, 240, 240, 240, 240 ตามลำดับ โดยที่ความแปรปรวนเท่ากัน (กรณี 4.5.2.8) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 25, 30, 35, 40, 45, 48, 50

วิธี median ที่ MinPts 45, 48, 50

วิธี Xia median ที่ MinPts 35, 40, 45, 48, 50

4.6.4.3 กลุ่มย่อยละ 160, 200, 240, 280, 320 ตามลำดับ โดยที่ความแปรปรวนไม่เท่ากัน (กรณีนี้ที่ 4.5.2.9) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 25, 30, 35, 40, 45, 48, 50

วิธี median ที่ MinPts 45, 48, 50

วิธี Xia median ที่ MinPts 35, 40, 45, 48, 50

4.6.4.4 กลุ่มย่อยละ 160, 200, 240, 280, 320 ตามลำดับ โดยที่ความแปรปรวนเท่ากัน (กรณีนี้ที่ 4.5.2.10) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 25, 30, 35, 40, 45, 48, 50

วิธี median ที่ MinPts 45, 48, 50

วิธี Xia median ที่ MinPts 35, 40, 45, 48, 50

4.6.4.5 กลุ่มย่อยละ 320, 280, 240, 200, 160 ตามลำดับ โดยที่ความแปรปรวนไม่เท่ากัน (กรณีนี้ที่ 4.5.2.11) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 15, 20, 25, 30, 35, 40, 45, 48, 50

วิธี P75 ที่ MinPts 5-10

วิธี Xia P75 ที่ MinPts 5-9

4.6.4.6 กลุ่มย่อยละ 320, 280, 240, 200, 160 ตามลำดับ โดยที่ความแปรปรวนเท่ากัน (กรณีนี้ที่ 4.5.2.12) วิธีการที่เหมาะสมคือ

วิธี mean และวิธี Xia mean ที่ MinPts 15, 20, 25, 30, 35, 40, 45, 48, 50

วิธี median ที่ MinPts 30, 35, 40, 45, 48, 50

วิธี Xia median ที่ MinPts 20, 25, 30, 35, 40, 45, 48, 50

สรุปได้ว่า สำหรับจำนวนกลุ่มเท่ากับ 5 และมีขนาดตัวอย่าง 1200 วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ตั้งแต่ MinPts 25 ขึ้นไป

4.7 การตรวจสอบผลวิธีการคำนวณค่า MinPts ที่ได้จากวิธี Daszykowski

จากที่กล่าวในบทที่ 2 เกี่ยวกับวิธีของ Daszykowski นั้น วิธีที่ใช้หา MinPts โดยคำนวณจาก rule of thumb มีสูตรคือ $MinPts = integer(\frac{m}{25})$ เมื่อ m คือจำนวนจุดในชุดข้อมูล แต่ถ้าข้อมูลมีจำนวนข้อมูลเกิน 500 ควรใช้ MinPts เท่ากับ 20 ทั้งนี้ผู้วิจัยต้องการศึกษาผลการจัดกลุ่มด้วยวิธี DBSCAN เมื่อใช้ค่า MinPts จากวิธีการข้างต้น โดยใช้เกณฑ์การวัดประสิทธิภาพด้วย modified purity เพื่อตรวจสอบถึงวิธีที่เหมาะสมสำหรับชุดข้อมูลต่าง ๆ จะได้ผลดังแสดงในตารางที่ 4.7.1

ตารางที่ 4.7.1 วิธีการระบุค่าพารามิเตอร์ Eps ที่เหมาะสมและค่า modified purity เมื่อใช้ค่า MinPts ที่ได้จากวิธี Daszykowski

ลักษณะข้อมูล	MinPts	วิธีที่เหมาะสม	ค่า modified purity สูงสุด	
ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน	20	P75 Xia P75	0.9962 0.9962	
ข้อมูลวงแหวนรูปแบบที่ 1	20	ไม่มี	0.5128	
ข้อมูลวงแหวนรูปแบบที่ 2	20	ไม่มี	0.3333	
ข้อมูลมีลักษณะเป็นมุม	20	ไม่มี	0.4345	
ข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร				
K=3, n=50,100,150	ความแปรปรวนไม่เท่ากัน	12	P75	0.8779
	ความแปรปรวนเท่ากัน	12	P75	0.8804
K=3, n=100, 100, 100	ความแปรปรวนไม่เท่ากัน	12	P75	0.8223
	ความแปรปรวนเท่ากัน	12	mean Xia mean	0.8256 0.8256
K=3, n=150, 100, 50	ความแปรปรวนไม่เท่ากัน	12	Xia P75	0.8025
	ความแปรปรวนเท่ากัน	12	mean Xia mean	0.7710 0.7710
K=3, n=200, 400, 600	ความแปรปรวนไม่เท่ากัน	20	P75	0.8725
	ความแปรปรวนเท่ากัน	20	P75	0.8835
K=3, n=400, 400, 400	ความแปรปรวนไม่เท่ากัน	20	P75	0.8583
	ความแปรปรวนเท่ากัน	20	mean Xia mean	0.8218 0.8218

ตารางที่ 4.7.1 วิธีการระบุค่าพารามิเตอร์ Eps ที่เหมาะสมและค่า modified purity เมื่อใช้ค่า MinPts ที่ได้จากวิธี Daszykowski (ต่อ)

	ลักษณะข้อมูล	MinPts	วิธีที่เหมาะสม	ค่า modified purity สูงสุด
K=3, n=600, 400, 200	ความแปรปรวนไม่เท่ากัน	20	P75	0.8489
	ความแปรปรวนเท่ากัน	20	mean Xia mean	0.7743 0.7743
K=5, n=60, 60, 60, 60, 60	ความแปรปรวนไม่เท่ากัน	12	mean Xia mean	0.8174 0.8174
	ความแปรปรวนเท่ากัน	12	Xia mean	0.8213
K=5, n=40, 50, 60, 70, 80	ความแปรปรวนไม่เท่ากัน	12	mean Xia mean	0.8382 0.8382
	ความแปรปรวนเท่ากัน	12	mean Xia mean	0.8485 0.8485
K=5, n=80, 70, 60, 50, 40	ความแปรปรวนไม่เท่ากัน	12	mean Xia mean	0.7636 0.7636
	ความแปรปรวนเท่ากัน	12	Xia median	0.7738
K=5, n=240, 240, 240, 240, 240	ความแปรปรวนไม่เท่ากัน	20	mean Xia mean	0.8252 0.8252
	ความแปรปรวนเท่ากัน	20	mean Xia mean	0.8304 0.8304
K=5, n=160, 200, 240, 280, 320	ความแปรปรวนไม่เท่ากัน	20	mean Xia mean	0.8395 0.8395
	ความแปรปรวนเท่ากัน	20	mean Xia mean	0.8381 0.8381
K=5, n=320, 280, 240, 200, 160	ความแปรปรวนไม่เท่ากัน	20	mean Xia mean	0.7868 0.7868
	ความแปรปรวนเท่ากัน	20	mean Xia mean	0.7892 0.7892

จากตารางที่ 4.7.1 พบว่าสำหรับข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน เมื่อใช้ค่า MinPts ที่ได้จากวิธี Daszykowski มีวิธีที่เหมาะสมได้แก่ P75 และ Xia P75 เนื่องจากได้ค่า modified purity สูง แต่สำหรับข้อมูลวงแหวนรูปแบบที่ 1 ข้อมูลวงแหวนรูปแบบที่ 2 และข้อมูลมีลักษณะเป็นมุม เมื่อใช้ค่า MinPts ที่ได้จากวิธี Daszykowski พบว่าผลการจัดกลุ่ม DBSCAN

ไม่เหมาะสม เนื่องจากได้ค่า modified purity ต่ำ ส่วนข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร พบว่า โดยส่วนใหญ่วิธีระบุค่าพารามิเตอร์ Eps ที่เหมาะสมได้แก่วิธี mean และ Xia mean อย่างไรก็ตาม ผลการจัดกลุ่ม DBSCAN ก็ยังไม่เหมาะสมเท่ากับวิธี K-means



บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

จากการเปรียบเทียบประสิทธิภาพ 11 วิธี ได้แก่ วิธีของ Daszykowski วิธี mean วิธี median วิธี P75 วิธี P95 วิธีของ Xia วิธีที่ประยุกต์จากวิธีของ Xia (Xia mean Xia median Xia P75 และ Xia P95) และวิธีของ Karami ผลการวิเคราะห์สรุปได้ดังนี้

การดำเนินการของ DBSCAN ขึ้นอยู่กับพารามิเตอร์ 2 ตัว คือ Eps และ MinPts ถ้า Eps มีขนาดใหญ่มาก ชุดข้อมูลนั้น ๆ ก็อาจมีจำนวนกลุ่มน้อยกว่าที่ควรจะเป็น แต่ถ้าหาก Eps มีขนาดเล็กเกินไป ก็จะทำให้มีจำนวนกลุ่มมากเกินไปหรืออาจทำให้เกิดจุดรบกวนเกิดขึ้นมาก ในกรณีที่ Eps คงที่ ถ้า MinPts มีค่าสูง จะทำให้จุดข้อมูลยากต่อการรวมเป็นกลุ่มเดียวกัน อีกทั้งอาจมีจุดรบกวนเกิดขึ้นมาก แต่ถ้า MinPts มีค่าน้อย จะทำให้จุดข้อมูลสามารถรวมเป็นกลุ่มเดียวกันได้ง่าย ซึ่งจากการศึกษาพบว่าในชุดข้อมูลเดียวกันอาจมีค่าพารามิเตอร์ Eps และ MinPts ที่เหมาะสมได้หลายค่า

วิธี Xia จะเหมาะสมสำหรับข้อมูลที่มีการแจกแจงของระยะห่างแต่ละจุดไปยังจุดที่ใกล้กันเป็นลำดับที่ k ที่มีลักษณะเบ้ขวา แต่วิธี Xia จะไม่เหมาะสมกับข้อมูลที่มีการแจกแจงของระยะห่างแต่ละจุดไปยังจุดที่ใกล้กันเป็นลำดับที่ k ที่ไม่ใช่ลักษณะเบ้ขวา นอกจากนี้การเปรียบเทียบระหว่างวิธี mean และวิธี Xia mean หรือวิธี median และวิธี Xia median หรือวิธี P75 และวิธี Xia P75 หรือวิธี P95 และวิธี Xia P95 การเปรียบเทียบวิธีภายในแต่ละคู่จะให้ค่า modified purity ที่ใกล้เคียงกัน จากการศึกษาข้อมูลแต่ละรูปแบบพบว่าสำหรับชุดข้อมูลไม่ปกติวิธีการจัดกลุ่ม DBSCAN ดีกว่าวิธีจัดกลุ่ม K-means ซึ่งวิธีการระบุค่าพารามิเตอร์ที่เหมาะสมจะแตกต่างกันออกไปตามช่วงของ MinPts จะเห็นว่าวิธี Daszykowski เหมาะสมกับทุกชุดข้อมูลในช่วง MinPts ไม่สูงมาก ส่วนวิธี Karami เหมาะสมในช่วง MinPts สูงกว่าช่วง MinPts ของวิธี Daszykowski และทั้ง 2 วิธีต่างไม่เหมาะสมในช่วง MinPts สูง ๆ

สำหรับข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน วิธีการที่เหมาะสมคือ

วิธี Daszykowski ที่ค่า MinPts ต่ำ (MinPts 2-5)

วิธี Karami ที่ค่า MinPts ต่ำแต่ช่วงกว้าง (MinPts 3-10)

วิธี P95 ที่ค่า MinPts ใกล้เคียงกับวิธี Karami (MinPts 4-10)

สำหรับข้อมูลวงแหวนรูปแบบที่ 1 วิธีการที่เหมาะสมคือ

วิธี Daszykowski ที่ค่า MinPts ต่ำมาก (MinPts 3-4)

วิธี Karami ที่ค่า MinPts ต่ำแต่ช่วงแคบ (MinPts 5-8)

วิธี P95 และ Xia P95 ที่ค่า MinPts ใกล้เคียงกับวิธี Karami (MinPts 4-8)

วิธี P75 ที่ค่า MinPts เหมือนกับวิธี Karami (MinPts 5-8)

วิธี Xia P75 ที่ค่า MinPts ใกล้เคียงกับวิธี Karami (MinPts 5-9)

วิธี mean Xia mean Xia median และ Xia ที่ค่า MinPts ต่ำแต่ช่วงกว้าง (MinPts 5-10)

วิธี median ที่ค่า MinPts ต่ำแต่ช่วงแคบ (MinPts 5-7, 9)

สำหรับข้อมูลวงแหวนรูปแบบที่ 2 วิธีการที่เหมาะสมคือ

วิธี Daszykowski ที่ค่า MinPts ต่ำ (MinPts 3-4)

วิธี Karami ที่ค่า MinPts ต่ำแต่ช่วงแคบ (MinPts 5-7)

วิธี P95 ที่ค่า MinPts ใกล้เคียงกับวิธี Karami (MinPts 5-8)

วิธี Xia P95 ที่ค่า MinPts ต่ำแต่ช่วงแคบ (MinPts 4-7)

วิธี P75, Xia P75 ที่ค่า MinPts ต่ำแต่ช่วงแคบ (MinPts 5, 7-8)

วิธี mean median Xia mean Xia median และ Xia ที่ค่า MinPts ต่ำแต่ช่วงแคบ (MinPts 8-9)

สำหรับข้อมูลมีลักษณะเป็นมุม วิธีการที่เหมาะสมคือ

วิธี Daszykowski ที่ MinPts 3

วิธี Karami ที่ MinPts 6

วิธี P95 ที่ MinPts 5

วิธี Xia P95 ที่ MinPts 4

และสำหรับข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร วิธีการที่ดีคือวิธี K-means เนื่องจากมีค่า modified purity สูงกว่าวิธีที่เหลือทั้งหมด

เมื่อการเปรียบเทียบภายใต้สถานการณ์จำนวนกลุ่มและขนาดตัวอย่างเท่ากันสำหรับข้อมูลที่มีการแจกแจงปกติ พบว่า

สำหรับจำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 300 โดยที่ความแปรปรวนไม่เท่ากัน วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ที่ค่า MinPts ระดับค่อนข้างสูง (MinPts 25, 30, 35)

สำหรับจำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 300 โดยที่ความแปรปรวนเท่ากัน วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ที่ค่า MinPts ระดับปานกลาง (MinPts 15, 20)

สำหรับจำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 1200 โดยที่ความแปรปรวนไม่เท่ากัน วิธีการที่เหมาะสมคือวิธี P75 และวิธี Xia P75 ที่ค่า MinPts ระดับปานกลางถึงค่อนข้างสูง (MinPts 15, 20, 25, 30)

สำหรับจำนวนกลุ่มเท่ากับ 3 และมีขนาดตัวอย่าง 1200 โดยที่ความแปรปรวนเท่ากัน วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ที่ค่า MinPts ระดับสูง (MinPts 45 เป็นต้นไป)

สำหรับจำนวนกลุ่มเท่ากับ 5 และมีขนาดตัวอย่าง 300 วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ที่ค่า MinPts ระดับปานกลาง (MinPts 12, 15)

สำหรับจำนวนกลุ่มเท่ากับ 5 และมีขนาดตัวอย่าง 1200 วิธีการที่เหมาะสมคือวิธี mean และวิธี Xia mean ตั้งแต่ MinPts 25 ขึ้นไป

นอกจากนี้จากการตรวจสอบผลวิธีการคำนวณค่า MinPts ที่ได้จากวิธี Daszykowski พบว่า สำหรับข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน เมื่อใช้ค่า MinPts ที่ได้จากวิธี Daszykowski มีวิธีที่เหมาะสมได้แก่ P75 และ Xia P75 แต่สำหรับข้อมูลวงแหวนรูปแบบที่ 1 ข้อมูลวงแหวนรูปแบบที่ 2 และข้อมูลมีลักษณะเป็นมุม เมื่อใช้ค่า MinPts ที่ได้จากวิธี Daszykowski พบว่าผลการจัดกลุ่ม DBSCAN ไม่เหมาะสม ส่วนข้อมูลที่มีการแจกแจงปกติ 2 ตัวแปร พบว่าโดยส่วนใหญ่วิธีระยะบุคาพารามิเตอร์ Eps ที่เหมาะสมได้แก่วิธี mean และ Xia mean อย่างไรก็ตามผลการจัดกลุ่ม DBSCAN ก็ยังไม่เหมาะสมเท่ากับวิธี K-means

5.2 ข้อเสนอแนะ

จากงานวิจัยนี้ผู้ที่สนใจอาจเลือกศึกษาเพิ่มเติมในเรื่อง

5.2.1 รูปร่างข้อมูลไม่ปกติในแบบอื่น ๆ นอกเหนือจากข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน ข้อมูลวงแหวนรูปแบบที่ 1 และ 2 ข้อมูลมีลักษณะเป็นมุม

5.2.2 ช่วง MinPts ในการพิจารณา อาจจะมีการขยายช่วงให้กว้างมากขึ้น

5.2.3 เกณฑ์ที่ใช้ในการพิจารณาประสิทธิภาพอาจต้องมีการใช้เกณฑ์อื่น ๆ

5.2.4 ชุดข้อมูลในหลายมิติ



รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. (2552). การวิเคราะห์ข้อมูลหลายตัวแปร (พิมพ์ครั้งที่ 4). กรุงเทพมหานคร: ธรรมสาร.

ชินพัฒน์ แก้วชินพร. (2553). การจำแนกประเภทข้อมูลด้วยเทคนิคต้นไม้ตัดสินใจและการจัดกลุ่ม. ปริญญาโท สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

พงศ์เลิศ สังกะเพศ. (2554). การนำฟังก์ชันความหนาแน่นร่วมกับอัลกอริทึมการค้นหาคำตอบด้วย แรงโน้มถ่วงเพื่อแก้ปัญหาการจัดแบ่งกลุ่มข้อมูล. วิทยานิพนธ์ วิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.

ภาษาต่างประเทศ

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). *OPTICS: ordering points to identify the clustering structure*. Paper presented at the ACM Sigmod Record.

Daszykowski, M., Walczak, B., & Massart, D. L. (2001). Looking for natural patterns in data: Part 1. Density-based approach. *Chemometrics and Intelligent Laboratory Systems*, 56(2), 83-92. doi: [http://dx.doi.org/10.1016/S0169-7439\(01\)00111-3](http://dx.doi.org/10.1016/S0169-7439(01)00111-3)

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. Paper presented at the Kdd.

Karami, A., & Johansson, R. (2014). Choosing DBSCAN Parameters Automatically using Differential Evolution. *International Journal of Computer Applications*, 91(7), 1-11.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press.

- Tan, P.-N., Steinbach, M., & Kumar, V. (2013). *Introduction to Data Mining: Pearson New International Edition*. England: Pearson Education Limited.
- XIA, L.-N., & JING, J.-W. (2009). SA-DBSCAN:A self-adaptive density-based clustering algorithm. *Journal of the Graduate School of the Chinese Academy of Sciences*, 26(4), 530-538.





ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

คำสั่งการวิเคราะห์ข้อมูลด้วยโปรแกรม R

เนื่องจากจำลองข้อมูลในหลายสถานการณ์ ดังนั้น จึงขอยกส่วนของโปรแกรมที่ใช้จำลอง สำหรับข้อมูลข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน (ทำ 1 รอบ) และข้อมูลมีลักษณะเป็นมุม (ทำซ้ำ 500 รอบ)

1 ข้อมูลที่รูปร่างและจำนวนตัวอย่างในแต่ละกลุ่มไม่เท่ากัน

```
library(fpc)
```

```
library(SuppDists)
```

```
library(fitdistrplus)
```

```
library(DescTools)
```

```
data<-read.table("C:/Users/Admin PC/Desktop/clustering/Aggregation.txt",header=FALSE,sep = ",")
```

```
plot(data)
```

```
abline(h=15)
```

```
abline(v=27)
```

```
segments(18,15,18,32)
```

```
segments(c(0,9),c(7,7),c(9,9),c(7,0))
```

```
abline(h=10,ty=2)
```

```
abline(v=11,ty=2)
```

```
G<ifelse(data[,2]>15,ifelse(data[,1]<18,1,ifelse(data[,1]<27,2,3)),ifelse(data[,1]>27,4,ifelse(data[,1]<9&data[,2]<7,5,ifelse(data[,1]>11|data[,2]<10,6,7))))
```

```
plot(data,col=G)
```

```
##function purity##
```

```
#class is real group
```

```
#cluster is calculate group
```

```
purity<-function(class,cluster){
```

```
  sum(apply(table(class[cluster>0],cluster[cluster>0]),2,max))/length(class)
```

```
}
```

```
##function modified purity##
```

```
modified_purity<-function(class,cluster,value_result=F){
```

```
  number_data<-length(class)
```

```
  table_frequency<-table(class[cluster>0],cluster[cluster>0])
```

```
  number_class<-nrow(table_frequency)
```

```
  number_cluster<-max(cluster)
```

```
  tie_position<-c()
```

```
  if(NCOL(table_frequency)>1&NROW(table_frequency)>1){
```

```

if(number_cluster>number_class){
  rank_row<-apply(table_frequency,1,function(r)rank(r,ties.method="max"))
  good_col<-apply(rank_row,2,function(z)which(z>number_cluster-
    number_class))
  table_frequency<-table_frequency[,intersect(1:number_cluster,
    ifelse(is.list(good_col),unlist,as.numeric)(good_col))]
  number_cluster<-ncol(table_frequency)
  col_position<-apply(table_frequency,1,function(y)which(y==max(y)))
  for(i in 1:number_class){
    tie_position[i]<-any(!col_position[[i]]%in%unlist(col_position[-i]))
  }
} else {
  row_position<-apply(table_frequency,2,function(y)which(y==max(y)))
  for(i in 1:number_cluster){
    tie_position[i]<-any(!row_position[[i]]%in%unlist(row_position[-i]))
  }
}
if(all(tie_position)){
  if(number_cluster>number_class){
    max_value<-apply(table_frequency,1,max)
    if(value_result){
      value_event<-cbind(table_frequency,max_value)
      colnames(value_event)[ncol(value_event)]<-"max value"
    }
  } else {
    max_value<-apply(table_frequency,2,max)
    if(value_result){
      value_event<-rbind(table_frequency,max_value)
      rownames(value_event)[nrow(value_event)]<-"max value"
    }
  }
  result<-sum(max_value)/number_data
} else if(number_cluster==number_class){
  col_position<-apply(table_frequency,1,function(y)which(y==max(y)))
  tie_position<-c()
  for(i in 1:number_class){
    tie_position[i]<-any(!col_position[[i]]%in%unlist(col_position[-i]))
  }
  if(all(tie_position)){
    max_value<-apply(table_frequency,1,max)
    if(value_result){

```

```

value_event<-cbind(table_frequency,max_value)
colnames(value_event)[ncol(value_event)]<-"max value"
}
result<-sum(max_value)/number_data
}
}
if(!all(tie_position)){
  if(choose(max(number_class,number_cluster),min(number_class,number_c
    luster))*factorial(min(number_class,number_cluster))>750000){
    rank_row<-apply(table_frequency,1,function(r)rank
      (r,ties.method="max"))
    list_good_col<-list()
    for(i in 1:number_class){
      list_good_col[[i]]<-which(rank_row[i]>number_cluster-
        number_class)
    }
    number_rank<-sapply(list_good_col,length)
    final_form_index_cluster<-list_good_col[[1]]
    for(i in 2:number_class){
      form_index_cluster<-final_form_index_cluster
      final_form_index_cluster<-c()
      for(j in 1:number_rank[i]){
        final_form_index_cluster<-
          rbind(final_form_index_cluster,
            cbind(form_index_cluster,list_good_col[[i]][j])
          )
      }
    }
    final_form_index_cluster<-final_form_index_cluster[apply
      (final_form_index_cluster,1,function(x)!any(duplicated(x))),]
    value_event<-matrix(nrow(final_form_index_cluster),number_class)
    for(i in 1:number_class){
      value_event[i,i]<-
        table_frequency[i,final_form_index_cluster[i]]
    }
  } else {
    possible_event<-CombSet(1:max(number_class,number_cluster),
      min(number_class,number_cluster),ord=T)
    value_event<-matrix(nrow(possible_event),ncol(possible_event))
    if(number_cluster>number_class){
      for(i in 1:number_class){

```



```

value_event[,i]<-
  table_frequency[i,possible_event[,i]]
}
} else {
  for(i in 1:number_cluster){
    value_event[,i]<-
      table_frequency[possible_event[,i],i]
  }
}
}
value_case<-rowSums(value_event)
result<-max(value_case)/number_data
if(value_result){
  value_event<-cbind(value_event,value_case)
  colnames(value_event)[ncol(value_event)]<-"sum event"
}
}
} else {
  result<-max(table_frequency)/number_data
  if(value_result){
    value_event<-table_frequency
  }
}
if(value_result){
  print(value_event)
}
return(result)
}

```

```

minp<-sort(union(c(2:10,seq(15,50,5)),floor(nrow(data)/25)))
lengthminp<-length(minp)
method<-c("Karami","mean","median","P75","P95","Xia","Xia mean","Xia median","Xia P75",
  "Xia P95","Daszykowski")
lengthmethod<-length(method)

```

```

##function calculate best K-means##
grouping<-function(x,g,cond=10,max=Inf){
  grp<-kmeans(x,g)
  eff<-grp$betweenss/grp$totss
  max.eff<-eff
}

```

```

check<-1
i<-0
while(check<cond&i<max){
  temp<-kmeans(x,g)
  eff<-temp$betweenss/temp$totss
  if(eff>max.eff){
    grp<-temp
    max.eff<-eff
    check<-1
    i<-0
  } else if(eff==max.eff){
    check<-check+1
    i<-0
  } else i<-i+1
}
return(grp)
}

##function run (calculate 11 method of eps and run dbscan for various minpts
(c(2:10,seq(15,50,5)),floor(nrow(data$data)/25)) and run K-Means)##
run<-function(data,G){
  Eps<-matrix(,lengthminp,lengthmethod,dimname=list(minp,method))
  pure<-Eps
  modified_pure<-Eps
  number_cluster<-Eps
  D<-dim(data)[2]
  n<-dim(data)[1]
  Ran<-prod(apply(data,2,max)-apply(data,2,min))
  dis<-as.matrix(dist(as.matrix(data), method = "euclidean", diag = T, upper = T))
  s<-apply(dis,2,sort)

  Eps[,1]<-(Ran*(factorial(0.5*D))*minp/(n*(pi^(D/2))))^(1/D)
  Eps[,2:5]<-t(apply(s[,minp,],1,function(x)c(mean(x),median(x),quantile(x,c(0.75,0.95))))))

  Q95dist<-matrix(,lengthminp,30)
  for(i in 1:30){
    newdata<-cbind(runif(n,floor(min(data[,1])),ceiling(max(data[,1])),
      runif(n,floor(min(data[,2])),ceiling(max(data[,2]))))
    dis_newdata<-as.matrix(dist(as.matrix(newdata), method = "euclidean", diag = T,
      upper = T))
    s_dis_newdata<-apply(dis_newdata,2,sort)
  }
}

```

```

      Q95dist[,i]<-apply(s_dis_newdata[minp,],1,function(x)quantile(x,0.95))
    }

    Eps[, "Daszykowski"]<-rowMeans(Q95dist)

    for (i in 1:lengthminp) {
      K<-s[minp[i,],]
      K_no_tie<-K
      level_tie<-which(K%in%K[duplicated(K)])
      K_no_tie[level_tie]<-jitter(K[level_tie])
      param<-fitdist(K_no_tie,"invGauss",start=list(nu=mean(K), lambda=1/mean(1/K-
        1/mean(K))))$est

      Eps[i,6:10]<-c((param["nu"]*sqrt(9*(param["nu"])^2+4*(param["lambda"])^2)-
        3*(param["nu"])^2)/(2*param["lambda"]),param["nu"],qinvGauss(c(0.5,0.75,0.
        95),param["nu"],param["lambda"]))

      for (j in 1:lengthmethod){
        ds<-dbscan(data,eps=Eps[i,j],MinPts=minp[i])
        pure[i,j]<- purity(G,ds$cluster)
        modified_pure[i,j]<- modified_purity(G,ds$cluster)
        number_cluster[i,j]<-max(ds$cluster)
      }
    }

    G_K.Mean<-grouping(data,nlevels(factor(G)))

    G_K.Mean<-match(G_K.Mean$cluster,order(G_K.Mean$centers[,1]))
    K.Mean_purity<-purity(G,G_K.Mean)
    K.Mean_modified_purity<-modified_purity(G,G_K.Mean)
    pure_heatplot<-matrix(,35,50,dimname=list(seq(0.1,6.9,0.2),1:50))
    modified_pure_heatplot<-pure_heatplot
    for(i in 1:35){
      for(j in 2:50){
        ds<-dbscan(data,eps=0.2*i-0.1,MinPts=j)
        pure_heatplot[i,j]<-purity(G,ds$cluster)
        modified_pure_heatplot[i,j]<-modified_purity(G,ds$cluster)
      }
    }
    modified_pure_heatplot[modified_pure_heatplot==-Inf]<-0
    return(list(Eps=Eps,pure=pure,pureKMean=K.Mean_purity,modified_purity=modified_pure

```

```

        ,modifiedpureKMean=K.Mean_modified_purity,pure_heatplot=pure_heatplot,
        modified_pure_heatplot=modified_pure_heatplot,NumberOfCluster=
        number_cluster))
    }

set.seed(587)
result<-run(data,G)
result

number_class<-length(unique(G))
max_pure<-max(result$pure[result$NumberOfCluster==number_class])
max_modified_pure<-max(result$modified_purity[result$NumberOfCluster==number_class])
table_bestpure<-c()
table_bestmodified_pure<-c()
for(i in 1:lengthmethod){
    position.minp<-which(result$pure[,i]>=max_pure)
    if(length(position.minp)>0){
        table_bestpure<-
            rbind(table_bestpure,data.frame(result$NumberOfCluster[position.minp,i],minp[positi
            on.minp],result$Eps[position.minp,i],method[i],result$pure
            [position.minp,i]))
    }
}
names(table_bestpure)<-c("Number Of Cluster","minp","Eps","method","pure")
table_bestpure<-table_bestpure[order(table_bestpure$Eps),]
table_bestpure<-table_bestpure[order(table_bestpure$minp),]
table_bestpure<-table_bestpure[order(table_bestpure$pure,decreasing=T),]
table_bestpure<-table_bestpure[order(table_bestpure[,1]),]
table_bestpure$Eps<-round(table_bestpure$Eps,4)
split(table_bestpure,table_bestpure$method)
table_bestpure

for(i in 1:lengthmethod){
    new.position.minp<-which(result$modified_purity[,i]>=max_modified_pure)
    if(length(new.position.minp)>0){
        table_bestmodified_pure<-
            rbind(table_bestmodified_pure,data.frame(result$NumberOfCluster
            [new.position.minp,i],minp[new.position.minp],result$Eps
            [new.position.minp,i],method[i],result$modified_purity[new.position.minp,i]))
    }
}
}

```

```

names(table_bestmodified_pure)<-c("Number Of Cluster","minp","Eps","method","modified_pure")
table_bestmodified_pure<-table_bestmodified_pure[order(table_bestmodified_pure$Eps),]
table_bestmodified_pure<-table_bestmodified_pure[order(table_bestmodified_pure$minp),]
table_bestmodified_pure<-
  table_bestmodified_pure[order(table_bestmodified_pure$modified_pure,decreasing=T),]
table_bestmodified_pure<-table_bestmodified_pure[order(table_bestmodified_pure[,1]),]
table_bestmodified_pure$Eps<-round(table_bestmodified_pure$Eps,4)
split(table_bestmodified_pure,table_bestmodified_pure$method)
table_bestmodified_pure

write.csv(result$Eps,file="Test Eps aggregation.csv")
write.csv(result$pure,file="Test pure aggregation.csv")
write.csv(result$modified_purity,file="Test modify pure aggregation.csv")

#####
##graph MinPts vs Eps (No modified pure background)##
windows()
plot(c(0,50.4),c(0,11),axes=F,main="modified purity",xlab="MinPts",ylab="Eps",col="white")
axis(1,minp)
axis(2,seq(0,7,0.5))
box()
j<-c(1:6,2:5,7)
line_style<-c(rep(c(1,2),c(6,4)),1)
z<-c(4,1,2,10,7,3,16,17,18,15,6)

for(i in 1:lengthmethod){
  points(minp,result$Eps[,i],col=rainbow(7)[i],pch=z[i])
  lines(minp,result$Eps[,i],col=rainbow(7)[i],lty=line_style[i],lwd=1.8)
}

legend("top",method,lty=line_style,lwd=1.8,pch=z,col=rainbow(7)[j],bty="n",ncol=2)

#####
##graph MinPts vs Eps (modified purity background)##
subEps<-seq(0,7,0.2)
windows()
plot(c(0,50.4),c(0,11),axes=F,main="modified purity",xlab="MinPts",ylab="Eps",col="white")
axis(1,minp)
axis(2,seq(0,7,0.5))
box()
j<-c(1:6,2:5,7)

```

```

line_style<-c(rep(c(1,2),c(6,4)),1)
z<-c(4,1,2,10,7,3,16,17,18,15,6)
for(k in 1:49+0.5){
  pure_allEps<-result$modified_pure_heatplot[,k+0.5]
  bestEps<-which(pure_allEps==max(pure_allEps))
  for(i in 1:35){
    polygon(c(k,k,k+1,k+1,k),subEps[c(i,i+1,i+1,i)],col=grey(pure_allEps[i]),lty=0)
  }
}
for(i in 1:length(method)){
  points(minp,result$Eps[,i],col=rainbow(7)[i],pch=z[i])
  lines(minp,result$Eps[,i],col=rainbow(7)[i],lty=line_style[i],lwd=1.8)
}
legend("top",method,lty=line_style,lwd=1.8,pch=z,col=rainbow(7)[j],bty="n",ncol=2)

#####
##cut region MinPts vs Eps (modified purity background)##
subEps<-seq(0.35,3.4,0.05)
new_modified_pure_heatplot<-matrix(,61,21,dimname=list(seq(0.375,3.375,0.05),1:21))
for(i in 1:61){
  for(j in 2:21){
    ds<-dbscan(data,eps=0.05*i+0.325,MinPts=j)
    new_modified_pure_heatplot[i,j]<-modified_purity(G,ds$cluster)
  }
}
new_modified_pure_heatplot[new_modified_pure_heatplot==-Inf]<-0

j<-c(1:6,2:5,7)
line_style<-c(rep(c(1,2),c(6,4)),1)
z<-c(4,1,2,10,7,3,16,17,18,15,6)
windows()
plot(c(1,20.6),c(0.36,4.6),axes=F,main="modified purity",xlab="MinPts",ylab="Eps",col="white")
axis(1,minp)
axis(2,seq(0,3.4,0.5))
box()

for(k in 1:20+0.5){
  pure_allEps<-new_modified_pure_heatplot[,k+0.5]
  for(i in 1:61){
    polygon(c(k,k,k+1,k+1,k),subEps[c(i,i+1,i+1,i)],col=grey(pure_allEps[i]),
lty=0)

```

```

    }
}

for(i in 1:lengthmethod){
  points(minp,result$Eps[,i],col=rainbow(7)[j[i]],pch=z[i])
  lines(minp,result$Eps[,i],col=rainbow(7)[j[i]],lty=line_style[i],lwd=1.8)
}

polygon(c(0,0,22,22),c(3.4,4.4,4.4,3.4),col="white",lty=0)
polygon(c(0,0,22,22),c(0,0.35,0.35,0),col="white",lty=0)
legend("top",method,lty=line_style,lwd=1.8,pch=z,col=rainbow(7)[j],bty="n",ncol=2)
box()

#####
##graph MinPts vs modified purity##
windows()
plot(c(0.2,50),c(0,1.6),axes=F,xlab="MinPts",ylab="modified purity",col="white")
axis(1,minp)
axis(2,seq(0,1,0.2))
box()

j<-c(1:6,NA,2:5,7)
line_style<-c(rep(c(1,NA,2),c(6,1,4)),1)
z<-c(4,1,2,10,7,3,NA,16,17,18,15,6)

for(i in 1:lengthmethod){
  i.<-i+ifelse(i>6,1,0)
  points(minp,result$modified_purity[,i],col=rainbow(7)[j[i.]],pch=z[i.])
  lines(minp,result$modified_purity[,i],col=rainbow(7)[j[i.]],lty=line_style[i.])
}

lines(range(minp),rep(result$modifiedpureKMean,2),lty=2)

legend("top",c(method[1:6],NA,method[7:11],"K-Means"),lty=c(line_style,2),pch=c(z,NA),
      col=c(rainbow(7)[j],1),lwd=1,bty="n",ncol=2)

#####
##graph original data vs K-Means##
windows()
plot(data,col=ifelse(G==1,2,ifelse(G==2,5,ifelse(G==3,6,ifelse(G==4,7,ifelse(G==6,4,ifelse(G==5,1,3)
))))))

```

```
G_K.Mean<-grouping(data,nlevels(factor(G)))  
G_K.Mean<-match(G_K.Mean$cluster,order(G_K.Mean$centers[,1]))  
windows()  
plot(data,col=G_K.Mean)
```



2 ข้อมูลวงแหวนรูปแบบที่ 1

```

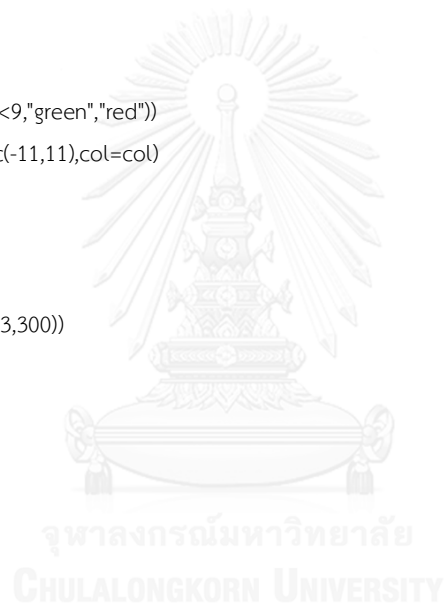
tetra<-seq(0,2*pi,len=21)
tetra<-tetra[-1]
x<-2*cos(tetra)
y<-2*sin(tetra)
g<-rep(2,20)

for(i in c(3:4,5,5:7,5,9:11)){
  tetra<-seq(0,2*pi,len=1+i*10)
  tetra<-tetra[-1]
  x<-c(x,i*cos(tetra))
  y<-c(y,i*sin(tetra))
  g<-c(g,rep(i,i*10))
}
col<-ifelse(g<5,"blue",ifelse(g<9,"green","red"))
plot(x,y,xlim=c(-11,11),ylim=c(-11,11),col=col)

data<-cbind(x,y)

G<-c(rep(1,90),rep(2,195),rep(3,300))
G

```



3 ข้อมูลวงแหวนรูปแบบที่ 2

```

tetra<-seq(0,pi,len=12)
x_up<-2*cos(tetra)
y_up<-1.25*sin(tetra)
x<-c(x_up,x_up)
y<-c(y_up+0.75,-y_up-0.75)
G<-c(rep(2,12),rep(-2,12))

for(i in c(3:4,9:11)){
  tetra<-seq(0,pi,len=i*6)
  x_up<-i*cos(tetra)
  y_up<-(-i-0.75)*sin(tetra)
  x<-c(x_up,x_up)
  y<-c(y_up+0.75,-y_up-0.75)
  G<-c(G,rep(i,i*6),rep(-i,i*6))
}

for(i in 5.5:7.5){
  tetra<-seq(0,2*pi,len=1+i*12)
  tetra<-tetra[-1]
  x<-c(x,i*cos(tetra))
  y<-c(y,i*sin(tetra))
  G<-c(G,rep(i,i*12))
}

col<-ifelse(G>0,ifelse(G<5,"blue",ifelse(G<9,"green","red")),ifelse(G>-5,"orange","purple"))
plot(x,y,xlim=c(-11,11),ylim=c(-11,11),col=col)
data<-cbind(x,y)
G<-col

```



4 ข้อมูลมีลักษณะเป็นมุม

```
##function generate corner data##
fn_gen<-function(n,l=1,h=1,x=0,y=0,G=NULL,seed=NA,where="center"){
  input<-list(n,l,h,x,y)
  number_param<-sapply(input,length)
  max_number_param<-max(number_param)
  input[number_param<max_number_param]<-
    lapply(input[number_param<max_number_param],
           function(Z)rep_len(Z,max_number_param))
  if(where%in%c("topright","bottomright","right")){
    input[[4]]<-input[[4]]-input[[2]]
  } else if(where%in%c("top","bottom","center")){
    input[[4]]<-input[[4]]-input[[2]]/2
  } if(where%in%c("topleft","topright","top")){
    input[[5]]<-input[[5]]-input[[3]]
  } else if(where%in%c("left","right","center")){
    input[[5]]<-input[[5]]-input[[3]]/2
  }
  para<-cbind(input[[1]],input[[2]],input[[3]],input[[4]],input[[5]])
  if(is.null(G)){
    G<-1:max_number_param
  } else G<-rep_len(G,max_number_param)
  X<-Y<-c()
  if(!is.na(seed))set.seed(seed)
  for(i in 1:max_number_param){
    X<-c(X,runif(para[i,1],0,para[i,2])+para[i,4])
    Y<-c(Y,runif(para[i,1],0,para[i,3])+para[i,5])
  }
  return(list(data=data.frame(X,Y),G=rep(G,para[,1])))
}

data<-fn_gen(150,c(12,5),c(5,12),c(-12.5,-5.5,0.5,0.5),
            c(5.5,12.5,5.5,12.5,rep(-0.5,4)),rep(1:4,each=2),where="topleft")
plot(data$data,col=data$G)

library(fpc)
library(SuppDists)
library(fitdistrplus)
library(DescTools)

##function purity##
#class is real group
#cluster is calculate group
```

```

purity<-function(class,cluster){
  sum(apply(table(class[cluster>0],cluster[cluster>0]),2,max))/length(class)
}

##function modified purity##
modified_purity<-function(class,cluster,value_result=F){
  number_data<-length(class)
  table_frequency<-table(class[cluster>0],cluster[cluster>0])
  number_class<-nrow(table_frequency)
  number_cluster<-max(cluster)
  tie_position<-c()
  if(NCOL(table_frequency)>1&NROW(table_frequency)>1){
    if(number_cluster>number_class){
      rank_row<-apply(table_frequency,1,function(r)rank(r,ties.method="max"))
      good_col<-apply(rank_row,2,function(z)which(z>number_cluster-
        number_class))
      table_frequency<-
        table_frequency[,intersect(1:number_cluster,ifelse(is.list(good_col),
          unlist(as.numeric)(good_col)))]
      number_cluster<-ncol(table_frequency)
      col_position<-apply(table_frequency,1,function(y)which(y==max(y)))
      for(i in 1:number_class){
        tie_position[i]<-any(!col_position[[i]]%in%unlist(col_position[-i]))
      }
    } else {
      row_position<-apply(table_frequency,2,function(y)which(y==max(y)))
      for(i in 1:number_cluster){
        tie_position[i]<-any(!row_position[[i]]%in%unlist(row_position[-i]))
      }
    }
  }
  if(all(tie_position)){
    if(number_cluster>number_class){
      max_value<-apply(table_frequency,1,max)
      if(value_result){
        value_event<-cbind(table_frequency,max_value)
        colnames(value_event)[ncol(value_event)]<-"max value"
      }
    } else {
      max_value<-apply(table_frequency,2,max)
      if(value_result){
        value_event<-rbind(table_frequency,max_value)

```

```

        rownames(value_event)[nrow(value_event)]<-"max value"
    }
}
result<-sum(max_value)/number_data
} else if(number_cluster==number_class){
col_position<-apply(table_frequency,1,function(y)which(y==max(y)))
tie_position<-c()
for(i in 1:number_class){
    tie_position[i]<-any(!col_position[[i]]%in%unlist(col_position[-i]))
}
if(all(tie_position)){
    max_value<-apply(table_frequency,1,max)
    if(value_result){
        value_event<-cbind(table_frequency,max_value)
        colnames(value_event)[ncol(value_event)]<-"max value"
    }
    result<-sum(max_value)/number_data
}
}
if(!all(tie_position)){
    if(choose(max(number_class,number_cluster),min(number_class,
        number_cluster))*factorial(min(number_class,number_cluster))
        >750000){
        rank_row<-
            apply(table_frequency,1,function(r)rank(r,ties.method="max
                "))
        list_good_col<-list()
        for(i in 1:number_class){
            list_good_col[[i]]<-which(rank_row[,i]>number_cluster-
                number_class)
        }
        number_rank<-sapply(list_good_col,length)
        final_form_index_cluster<-list_good_col[[1]]
        for(i in 2:number_class){
            form_index_cluster<-final_form_index_cluster
            final_form_index_cluster<-c()
            for(j in 1:number_rank[i]){
                final_form_index_cluster<-
                    rbind(final_form_index_cluster,
                        cbind(form_index_cluster,
                            list_good_col[[i]][j]))
            }
        }
    }
}

```

```

    }
  }
  final_form_index_cluster<-
    final_form_index_cluster[apply(final_form_index_cluster,
    1,function(x)!any(duplicated(x))),]
  value_event<-
    matrix(nrow(final_form_index_cluster),number_class)
  for(i in 1:number_class){
    value_event[,i]<-
      table_frequency[i,final_form_index_cluster[,i]]
  }
} else {
  possible_event<-
    CombSet(1:max(number_class,number_cluster),min(number_
    r_class,number_cluster),ord=T)
  value_event<-matrix(nrow(possible_event),ncol(possible_event))
  if(number_cluster>number_class){
    for(i in 1:number_class){
      value_event[,i]<-
        table_frequency[i,possible_event[,i]]
    }
  } else {
    for(i in 1:number_cluster){
      value_event[,i]<-
        table_frequency[possible_event[,i],i]
    }
  }
}
value_case<-rowSums(value_event)
result<-max(value_case)/number_data
if(value_result){
  value_event<-cbind(value_event,value_case)
  colnames(value_event)[ncol(value_event)]<-"sum event"
}
} else {
  result<-max(table_frequency)/number_data
  if(value_result){
    value_event<-table_frequency
  }
}

```

```

    }
    if(value_result){
        print(value_event)
    }
    return(result)
}

minp<-sort(union(c(2:10,seq(15,50,5)),floor(nrow(data$data)/25)))
lengthminp<-length(minp)
method<-c("Karami","mean","median","P75","P95","Xia","Xia mean","Xia median","Xia P75",
          "Xia P95","Daszykowski")
lengthmethod<-length(method)

##function calculate best K-means##
grouping<-function(x,g,cond=10,max=Inf){
    grp<-kmeans(x,g)
    eff<-grp$betweenss/grp$totss
    max.eff<-eff
    check<-1
    i<-0
    while(check<cond&i<max){
        temp<-kmeans(x,g)
        eff<-temp$betweenss/temp$totss
        if(eff>max.eff){
            grp<-temp
            max.eff<-eff
            check<-1
            i<-0
        } else if(eff==max.eff){
            check<-check+1
            i<-0
        } else i<-i+1
    }
    return(grp)
}

##function run (calculate 11 method of eps and run dbscan for various minpts
(c(2:10,seq(15,50,5)),floor(nrow(data$data)/25)) and run K-Means)##
run<-function(data,G){
    Eps<-matrix(,lengthminp,lengthmethod,dimname=list(minp,method))
    pure<-Eps

```

```

modified_pure<-Eps
number_cluster<-Eps
D<-dim(data)[2]
n<-dim(data)[1]
Ran<-prod(apply(data,2,max)-apply(data,2,min))
dis<-as.matrix(dist(as.matrix(data), method = "euclidean", diag = T, upper = T))
s<-apply(dis,2,sort)

Eps[,1]<-(Ran*(factorial(0.5*D))*minp/(n*(pi^(D/2))))^(1/D)
Eps[,2:5]<-t(apply(s[minp,],1,function(x)c(mean(x),median(x),quantile(x,c(0.75,0.95))))))

Q95dist<-matrix(,lengthminp,30)
for(i in 1:30){
  newdata<-cbind(runif(n,floor(min(data[,1])),ceiling(max(data[,1]))),
    runif(n,floor(min(data[,2])),ceiling(max(data[,2]))))
  dis_newdata<-as.matrix(dist(as.matrix(newdata), method = "euclidean", diag = T,
    upper = T))
  s_dis_newdata<-apply(dis_newdata,2,sort)
  Q95dist[,i]<-apply(s_dis_newdata[minp,],1,function(x)quantile(x,0.95))
}

Eps[,"Daszykowski"]<-rowMeans(Q95dist)

for (i in 1:lengthminp) {
  K<-s[minp[i,]
  K_no_tie<-K
  level_tie<-which(K%in%K[duplicated(K)])
  K_no_tie[level_tie]<-jitter(K[level_tie])
  param<-fitdist(K_no_tie,"invGauss",start=list(nu=mean(K),lambda=1/mean(1/K-
    1/mean(K))))$est
  Eps[i,6:10]<-c((param["nu"]*sqrt(9*(param["nu"])^2+4*(param["lambda"])^2)-
    3*(param["nu"]^2)/(2*param["lambda"]),param["nu"],qinvGauss(c(0.5,0.75,0.
    95),param["nu"],param["lambda"]))

  for (j in 1:lengthmethod){
    ds<-dbscan(data,eps=Eps[i,j],MinPts=minp[i])
    pure[i,j]<- purity(G,ds$cluster)
    modified_pure[i,j]<- modified_purity(G,ds$cluster)
    number_cluster[i,j]<-max(ds$cluster)
  }
}

```



```

G_K.Mean<-grouping(data,nlevels(factor(G)))

G_K.Mean<-match(G_K.Mean$cluster,order(G_K.Mean$centers[,1]))
K.Mean_purity<-purity(G,G_K.Mean)
K.Mean_modified_purity<-modified_purity(G,G_K.Mean)
return(list(Eps=Eps,pure=pure,pureKMean=K.Mean_purity,modified_purity=modified_pure
,modifiedpureKMean=K.Mean_modified_purity,NumberOfCluster=number_cluster))
}

##run 500 rounds##
table.value<-as.data.frame(matrix(,500,11,dimname=list(c(),method)))
EpsValue<-list()
length(EpsValue)<-lengthminp
for(j in 1:lengthminp){
  EpsValue[[j]]<-table.value
}
names(EpsValue)<-minp
pureValue<-EpsValue
modifiedpureValue<-EpsValue
NumberOfClusterValue<-EpsValue

pureKMean<-c()
modifiedpureKMean<-c()
set.seed(347623)
for(i in 1:500){
  data<-fn_gen(150,c(12,5),c(5,12),c(-12.5,-5.5,0.5,0.5),
c(5.5,12.5,5.5,12.5,rep(-0.5,4)),rep(1:4,each=2),where="topleft")
  result<-run(data$data,data$G)
  pureKMean[i]<-result$pureKMean
  modifiedpureKMean[i]<-result$modifiedpureKMean
  for(j in 1:lengthminp){
    EpsValue[[j]][i,<-result$Eps[j,]
    pureValue[[j]][i,<-result$pure[j,]
    modifiedpureValue[[j]][i,<-result$modified_purity[j,]
    NumberOfClusterValue[[j]][i,<-result$NumberOfCluster[j,]
  }
  pie(c(i,500-i),c(i,500-i),radius=1,clockwise=T)
}
}

```

```
#####
pureValue
pureKMean
EpsValue

for (i in 1:lengthminp){
  write.csv(pureValue[[i]],file=paste0("TEST 500TIMES pure corner minpts= ",minp[i],".csv"))
  write.csv(modifiedpureValue[[i]],file=paste0("TEST 500TIMES modified pure corner minpts=
    ",minp[i],".csv"))
  write.csv(EpsValue[[i]],file=paste0("TEST 500TIMES eps corner minpts= ",minp[i],".csv"))
}

pure<-t(sapply(pureValue,colMeans))
modified_pure<-t(sapply(modifiedpureValue,colMeans))
Eps<-t(sapply(EpsValue,colMeans))
NumberOfCluster<-t(sapply(NumberOfClusterValue,colMeans))

write.csv(pure,file="TEST 500TIMES pure corner.csv")
write.csv(modified_pure,file="TEST 500TIMES modified pure corner.csv")
write.csv(Eps,file="TEST 500TIMES eps corner.csv")
rm(result)

##table of pure
number_class<-length(unique(data$G))
range(NumberOfCluster)
v<-ceiling(min(NumberOfCluster[NumberOfCluster>=number_class]))
max_pure<-max(pure[NumberOfCluster<=v])
max_modified_pure<-max(modified_pure[NumberOfCluster<=v])
table_bestpure<-c()
table_bestmodified_pure<-c()
for(i in 1:lengthmethod){
  position.minp<-which(pure[,i]>=max_pure)
  if(length(position.minp)>0){
    table_bestpure<-
      rbind(table_bestpure,data.frame(NumberOfCluster[position.minp,i],
        minp[position.minp],Eps[position.minp,i],method[i],pure[position.minp,i]))
  }
}

names(table_bestpure)<-c("Number Of Cluster","minp","Eps","method","pure")
table_bestpure<-table_bestpure[order(table_bestpure$Eps),]
table_bestpure<-table_bestpure[order(table_bestpure$minp),]
```

```

table_bestpure<-table_bestpure[order(table_bestpure$pure,decreasing=T),]
table_bestpure<-table_bestpure[order(table_bestpure[,1]),]
table_bestpure$Eps<-round(table_bestpure$Eps,4)
split(table_bestpure,table_bestpure$method)
table_bestpure

##table of modified purity
for(i in 1:lengthmethod){
  new.position.minp<-which(modified_pure[,i]>=max_modified_pure)
  if(length(new.position.minp)>0){
    table_bestmodified_pure<-
      rbind(table_bestmodified_pure,data.frame(NumberOfCluster[new.position.minp,i],mi
np[new.position.minp],Eps[new.position.minp,i],method[i],modified_pure[new.positio
n.minp,i]))
  }
}
names(table_bestmodified_pure)<-c("Number Of Cluster","minp","Eps","method","modified_pure")
table_bestmodified_pure<-table_bestmodified_pure[order(table_bestmodified_pure$Eps),]
table_bestmodified_pure<-table_bestmodified_pure[order(table_bestmodified_pure$minp),]
table_bestmodified_pure<-
  table_bestmodified_pure[order(table_bestmodified_pure$modified_pure,decreasing=T),]
table_bestmodified_pure<-table_bestmodified_pure[order(table_bestmodified_pure[,1]),]
table_bestmodified_pure$Eps<-round(table_bestmodified_pure$Eps,4)
split(table_bestmodified_pure,table_bestmodified_pure $method)
table_bestmodified_pure

mean(modifiedpureKMean)

#####
##graph MinPts vs Eps##
j<-c(1:6,2:5,7)
line_style<-c(rep(1:2,c(6,4)),1)
z<-c(4,1,2,10,7,3,16,17,18,15,6)

windows()
plot(c(0,50.4),c(0,7),axes=F,xlab="MinPts",ylab="Eps",col="white")
axis(1,minp)
axis(2,seq(0,4.5,0.5))
box()

for(i in 1:lengthmethod){

```

```

        points(minp,Eps[,i],col=rainbow(7)[j[i]],pch=z[i])
        lines(minp,Eps[,i],col=rainbow(7)[j[i]],lty=line_style[i],lwd=1.8)
    }
    legend("top",method,lty=line_style,lwd=1.8,pch=z,col=rainbow(7)[j],bty="n",ncol=2)

#####
##graph MinPts vs modified purity##
windows()
plot(c(0.2,50),c(0,1.5),axes=F,xlab="MinPts",ylab="modified purity",col="white")
axis(1,minp)
axis(2,seq(0,1,0.2))
box()

j<-c(1:6,NA,2:5,7)
line_style<-c(rep(c(1,NA,2),c(6,1,4)),1)
z<-c(4,1,2,10,7,3,NA,16,17,18,15,6)

for(i in 1:length(method)){
    i.<-i+ifelse(i>6,1,0)
    points(minp,modified_pure[,i],col=rainbow(7)[j[i.]],pch=z[i.])
    lines(minp,modified_pure[,i],col=rainbow(7)[j[i.]],lty=line_style[i.])
}
lines(range(minp),rep(mean(modifiedpureKMean),2),lty=2)

legend("top",c(method[1:6],NA,method[7:11],"K-Means"),lty=c(line_style,2),pch=c(z,NA),
       col=c(rainbow(7)[j],1),lwd=1,bty="n",ncol=2)

#####
##graph original data vs K-Means##
windows()
plot(data$data,col=ifelse(data$G==1,3,ifelse(data$G==2,5,ifelse(data$G==4,1,4))))
G_K.Mean<-grouping(data$data,nlevels(factor(data$G)))
G_K.Mean<-match(G_K.Mean$cluster,order(G_K.Mean$centers[,1]))
windows()
plot(data$data,col=G_K.Mean)

```

5 ข้อมูลที่มีการแจกแจงปกติ

กรณีที่ 4.5.1.1 ขนาดตัวอย่าง $N=300$ กลุ่มละ 50, 100, 150 ตามลำดับ

$$\text{โดยที่ } \mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mu_2 = \begin{bmatrix} 6 \\ 0 \end{bmatrix}, \mu_3 = \begin{bmatrix} 0 \\ 6 \end{bmatrix}$$

$$\text{และมีความแปรปรวนไม่เท่ากัน } \Sigma_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \Sigma_3 = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

```
gen_normal<-function(n,mean1,mean2,sd1,sd2,seed=NA){
  input<-list(n,mean1,mean2,sd1,sd2)
  number_param<-sapply(input,length)
  max_number_param<-max(number_param)
  input[number_param<max_number_param]<-lapply(input[number_param<max_number_param],
    function(Z)rep_len(Z,max_number_param))
  para<-cbind(input[[1]],input[[2]],input[[3]],input[[4]],input[[5]])
  G<-1:max_number_param
  X<-Y<-c()
  if(!is.na(seed))set.seed(seed)
  for(i in 1:max_number_param){
    X<-c(X,rnorm(para[i,1],para[i,2],para[i,4]))
    Y<-c(Y,rnorm(para[i,1],para[i,3],para[i,5]))
  }
  return(list(data=data.frame(X,Y),G=rep(G,para[,1])))
}
```

```
data<-gen_normal(c(50,100,150),c(0,6,0),c(0,0,6),sqrt(c(1,2,3)),sqrt(c(1,2,3)))
plot(data$data, col=data$G)
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาววรรณัญ โกวิทวิชกานนท์ เกิดวันจันทร์ที่ 10 มิถุนายน พ.ศ. 2534 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2555 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2556

