การเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอน

นางสาวกัณฐิกา พรมมา

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาสถิติ ภาควิชาสถิติ
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย
บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย
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 Graduate School.

COMPARISON OF TWO-STAGE CLUSTERING ALGORITHMS

Miss Kanthika Promma

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 2013

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอน
โดย	นางสาวกัณฐิกา พรมมา
สาขาวิชา	สถิติ
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	อาจารย์ ดร.อัครินทร์ ไพบูลย์พานิช

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะพาณิชยศาสตร์และการบัญชี
(รองศาสตราจารย์ ดร.พสุ เดชะรินทร์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(อาจารย์ ดร.อัครินทร์ ไพบูลย์พานิช)

..... กรรมการ
(อาจารย์ ดร.วิฐรา พึ่งพาพงศ์)

..... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.บุญสิทธิ์ ยิ้มวาสนา)

กัณฐิกา พรมมา : การเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอน. (COMPARISON OF TWO-STAGE CLUSTERING ALGORITHMS) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : อ.ดร. อัครินทร์ ไพบูลย์พานิช , 127หน้า.

การวิจัยในครั้งนี้ มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมการจัดกลุ่มข้อมูลแบบ 2 ขั้นตอน 4 วิธีด้วยกัน ได้แก่ วิธีที่ 1 SOM ร่วมกับ K-Means วิธีที่ 2 SOM ร่วมกับ Fuzzy C-Means วิธีที่ 3 Ward ร่วมกับ K-Means และวิธีที่ 4 Ward ร่วมกับ Fuzzy C-means โดยทำการศึกษาสำหรับข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) จำนวนตัวแปรที่ใช้ในการศึกษาครั้งนี้ คือ 2, 4 และ 6 ตัวแปร ลักษณะความสัมพันธ์ของตัวแปร ศึกษาทั้งในกรณีตัวแปรอิสระกัน และ ตัวแปรมีความสัมพันธ์กัน ข้อมูลที่ใช้ในการศึกษามาจากการจำลองข้อมูลโดยมีขนาดตัวอย่าง 25, 100 และ 300 จำนวนกลุ่มที่จำลอง คือ 2, 4 และ 6 กลุ่มสำหรับแต่ละขนาดตัวอย่าง โดยพิจารณาทั้งข้อมูลแต่ละกลุ่มไม่มีการซ้อนทับกัน (Nonoverlapping Clusters) และ ข้อมูลแต่ละกลุ่มมีการซ้อนทับกัน 40% (40% Overlapping Clusters)

จากผลการศึกษาโดยเปรียบเทียบค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate พบว่า การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี Ward ร่วมกับ Fuzzy C-Means ให้ประสิทธิภาพการจัดกลุ่มดีกว่าวิธี SOM ร่วมกับ K-Means และ Ward ร่วมกับ K-Means

ภาควิชา สถิติ ลายมือชื่อนิสิต .....

สาขาวิชา สถิติ ลายมือ ชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2556

5481510826 : MAJOR STATISTICS

KEYWORDS: Two-Stage Clustering Algorithms, K-Means Clustering Algorithm, Fuzzy C-Means Clustering Algorithm, Kohonen' s Self-Organizing maps (SOM), Ward' s Minimum Variance Method

KANTHIKA PROMMA : COMPARISON OF TWO-STAGE CLUSTERING ALGORITHMS. ADVISOR : AKARIN PHAIBULPANICH, Ph.D., 127 pp.

The objective of this study is to compare four two-stage clustering algorithms: (1) SOM with K-Means, (2) SOM with Fuzzy C-Means, (3) Ward with K-Means and (4) Ward with Fuzzy C-Means. The simulated data using in this study is the multivariate normal distribution with number of variables p=2, 4 and 6 variables with uncorrelated and correlated variables. Data were simulated with sample sizes n=25, 100, 300 and number of clusters k=2, 4, 6. Clusters were also simulated considering both nonoverlapping and 40% overlapping clusters.

Based on the simulation results, by comparing values of Wilk' s Lambda, root mean square standard deviation (RMSSTD) and misclassification rate, SOM with Fuzzy C-Means and Ward with Fuzzy C-Means are better than SOM with K-Means and Ward with K-Means.

Department : Statistics Student's Signature……..……..……

Field of Study : Statistics Advisor's Signature……..……..

Academic Year : 2013

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สามารถสำเร็จลุล่วงไปได้ด้วยดี ด้วยความช่วยเหลือและเอาใจใส่อย่างดีของอาจารย์ ดร. อัครินทร์ ไพบูลย์พานิช อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณอาจารย์เป็นอย่างสูง ที่กรุณาให้คำปรึกษา อบรมสั่งสอน และให้ข้อคิดเห็นต่างๆ ตลอดจนให้ความช่วยเหลือและคำแนะนำเพื่อปรับปรุงแก้ไข จนกระทั่งวิทยานิพนธ์เสร็จสมบูรณ์

ผู้วิจัยขอกราบขอบพระคุณรองศาสตราจารย์ ดร.กัลยา วานิชย์บัญชา ประธานกรรมการสอบวิทยานิพนธ์ อาจารย์ ดร.วิฐรา พึ่งพาพงศ์ และอาจารย์ ดร. บุญสิทธิ์ ยิ้มวาสนา กรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณครอบครัว ที่ช่วยส่งเสริม สนับสนุน และให้กำลังใจเสมอมาจนสำเร็จการศึกษา รวมทั้งขอขอบคุณเพื่อน ๆ ทุกคน ที่คอยช่วยให้กำลังใจผู้วิจัยมาโดยตลอด

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญภาพ ..... ฎ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการศึกษา 3

1.3 สมมติฐานในการศึกษา 4

1.4 ขอบเขตของการศึกษา 4

1.5 วิธีดำเนินการศึกษา 5

1.6 ประโยชน์ที่คาดว่าจะได้รับ 6

บทที่ 2 ทฤษฎีและตัวสถิติที่เกี่ยวข้อง 7

2.1 อัลกอริทึมการจัดกลุ่มแบบลำดับขั้น 7

2.2 อัลกอริทึมการจัดกลุ่มแบบเค-มีน 13

2.3 อัลกอริทึมการจัดกลุ่มแบบฟัซซี่ซีมีน 20

2.4 ระบบโครงข่ายประสาทเทียม 27

2.5 การวัดประสิทธิภาพการจัดกลุ่มข้อมูล 36

บทที่ 3 วิธีดำเนินการศึกษา 38

3.1 แผนการดำเนินการวิจัย 38

3.2 ขั้นตอนในการดำเนินการวิจัย 41

3.3 ขั้นตอนการทำงานของโปรแกรม 42

หน้า

บทที่ 4 ผลการวิเคราะห์ข้อมูล 44

บทที่ 5 สรุปผลการศึกษาและข้อเสนอแนะ 111
5.1 สรุปผลการศึกษา 111
5.2 ข้อเสนอแนะ 119

รายการอ้างอิง 120
ภาคผนวก 121
ประวัติผู้เขียนวิทยานิพนธ์ 127

สารบัญตาราง

ตารางที่ หน้า

2.1.1 แสดงค่า ESS ของการรวมกลุ่มเป็น 4 กลุ่มทั้งหมดที่เป็นไปได้ 11
2.1.2 แสดงค่า ESS ของการรวมกลุ่มเป็น 3 กลุ่มทั้งหมดที่เป็นไปได้ 12
2.1.3 แสดงค่า ESS ของการรวมกลุ่มเป็น 2 กลุ่มทั้งหมดที่เป็นไปได้ 12
2.2.1 ข้อมูลตัวอย่างที่ 2.2 17
2.2.2 จุดศูนย์กลางเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 2.2 17
2.2.3 ค่าเฉลี่ยของข้อมูลแต่ละกลุ่ม เมื่อข้อมูลมีการเปลี่ยนแปลงกลุ่ม 19
2.2.4 ระยะห่างยุคลิดกำลังสองระหว่างข้อมูลกับค่าเฉลี่ยของข้อมูลแต่ละกลุ่ม 19
2.3.1 ข้อมูลตัวอย่างที่ 2.3 22
2.3.2 จุดตัวแทนเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 2.3 22
2.3.2 จุดตัวแทนเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 2.3 22
2.3.3 จุดตัวแทนที่ได้รับการปรับค่าแล้วของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 2.3 26
2.3.4 ค่าความเป็นสมาชิกของข้อมูลแต่ละตัวกับตัวแทนกลุ่ม 26
2.4.1 ข้อมูลตัวอย่างที่ 2.4 30
2.4.2 ตัวอย่างกลุ่มข้อมูลที่ได้จากการจัดกลุ่มโดยใช้ SOM อัลกอริทึม 35
4.1 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม 45
4.2 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่ม ไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม 48
4.3 แสดงค่า Misclassification Rate ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม 51
4.4 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม 56
4.5 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่ม ซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม 59
4.6 แสดงค่า Misclassification Rate ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม 62

ตารางที่ หน้า

4.7 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม 67
4.8 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม 70
4.9 แสดงค่า Misclassification Rate ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม 73
4.10 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม 78
4.11 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่ม ซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม 81
4.12 แสดงค่า Misclassification Rate ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม 84
4.13 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม 89
4.14 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่ม ไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม 92
4.15 แสดงค่า Misclassification Rate ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม 95
4.16 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม 100
4.17 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่ม ซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม 103
4.18 แสดงค่า Misclassification Rateของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูล ที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม 106

สารบัญภาพ

ภาพที่		หน้า
2.1	การจัดกลุ่มแบบลำดับขั้นของข้อมูลการแสดงออกของยีน	7
2.2	ตัวอย่างการสร้างและจัดกลุ่มแบบลำดับขั้นด้วยเดนโดรแกรม	8
2.3	ตัวอย่างการสร้างกลุ่มแบบเกาะกลุ่มและแตกกลุ่ม	9
2.4	การกำหนดจุดข้อมูลเพื่อใช้ในการจัดกลุ่มข้อมูลแบบเค-มีน	14
2.5	ตัวอย่างการจัดกลุ่มข้อมูล โดยพิจารณาค่าระยะห่างระหว่างข้อมูลกับจุดศูนย์กลางของกลุ่มข้อมูลที่สุ่มได้ในครั้งแรก	15
2.6	ตัวอย่างการปรับจุดศูนย์กลางของข้อมูลแต่ละกลุ่มไปยังตำแหน่งที่ครอบคลุมข้อมูลบริเวณใกล้เคียงภายในกลุ่มเดียวกันให้ได้มากที่สุด โดยการหาค่าเฉลี่ยของข้อมูลภายในกลุ่ม	15
2.7	การจัดกลุ่มที่เสร็จสมบูรณ์ โดยจุดศูนย์กลางที่ได้จะอยู่ในตำแหน่งที่เป็นตัวแทนของข้อมูลแต่ละกลุ่ม	16
2.8	การทำงานของระบบโครงข่ายประสาทเทียมแบบ SOM	27
2.9	การทำงานของระบบโครงข่ายประสาทเทียมแบบ SOM	28
2.10	ระบบโครงข่ายประสาทเทียมแบบมีชั้นซ่อน	28
2.11	Input และ Output ของโครงข่ายประสาทเทียม	31
3.1	ข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน (Nonoverlapping Clusters)	39
3.2	ข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน 40% (40% Overlapping Clusters)	40

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในการวิเคราะห์ข้อมูล เทคนิคการจัดกลุ่มข้อมูล (Clustering) ถือเป็นเทคนิคที่นำมาใช้อย่างแพร่หลาย โดยจะจัดข้อมูลที่มีลักษณะคล้ายคลึงกันไว้ในกลุ่มเดียวกัน และข้อมูลที่มีลักษณะแตกต่างกันก็จะถูกจัดไว้ต่างกลุ่มกัน ในปัจจุบันเทคนิคการจัดกลุ่มข้อมูลถูกนำมาใช้ประโยชน์ในด้านต่างๆ เช่น ด้านการแพทย์ ด้านสังคมศาสตร์ ด้านการตลาด และด้านการศึกษา วัตถุประสงค์ของการจัดกลุ่มข้อมูลก็แตกต่างกันออกไปตามแต่ละสาขา เช่น ในด้านการแพทย์อาจจัดแบ่งคนไข้ตามระดับอาการหรือความรุนแรงของโรค เพื่อวางแผนในการรักษาตามระดับอาการนั้นๆ หรือในด้านการตลาด ก็อาจจัดแบ่งลูกค้าตามพฤติกรรมการบริโภคสินค้า เพื่อวางแผนทางด้านการตลาดและเพิ่มประสิทธิภาพในการขายต่อไป เป็นต้น (กัลยา, 2552)

เทคนิคในการจัดกลุ่มข้อมูลที่นำมาใช้วิเคราะห์ข้อมูลในปัจจุบันมีอยู่หลากหลายเทคนิควิธี ทั้งเทคนิควิธีที่เป็นที่รู้จักและถูกนำมาใช้ในการจัดกลุ่มข้อมูลอย่างแพร่หลาย เช่น อัลกอริทึมการจัดกลุ่มแบบวอร์ด (Ward' s Minimum Variance Method) เป็นวิธีการจัดกลุ่มแบบลำดับขั้นประเภทการเกาะกลุ่ม(Agglomerative Hierarchical Clustering) ซึ่งมีข้อดีในแง่ที่ไม่ต้องกำหนดจำนวนกลุ่มและตัวแทนกลุ่มเริ่มต้น แต่อัลกอริทึมการจัดกลุ่มแบบวอร์ดนี้ไม่เหมาะกับข้อมูลขนาดใหญ่ อาจส่งผลต่อประสิทธิภาพในแง่ของเวลา ส่วนอัลกอริทึมการจัดกลุ่มแบบเค-มีน (K-Means Clustering Algorithm) ซึ่งมีข้อดีในแง่ของการจัดกลุ่มข้อมูลได้รวดเร็ว ง่ายและนิยมนำไปใช้งาน แต่การจัดกลุ่มข้อมูลแบบเค-มีนจะมีปัญหากับข้อมูลที่มี outlier เนื่องจากถ้าหากมีข้อมูลสำคัญกระจายอยู่นอกกลุ่ม อาจทำให้การจัดกลุ่มมีความคลาดเคลื่อนได้ นอกจากนี้การกำหนดจำนวนกลุ่มและตัวแทนกลุ่มเริ่มต้น อาจส่งผลต่อประสิทธิภาพการจัดกลุ่มทั้งในเชิงเวลาและความถูกต้อง และในปัจจุบันได้มีการพัฒนาสร้างเทคนิควิธีการจัดกลุ่มใหม่ๆ เช่น อัลกอริทึมการจัดระบบด้วยตัวเองของโคโฮเนน (Kohonen' s Self-Organizing Maps : SOM) อัลกอริทึมจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C-Means Algorithm) โดยมีการนำไปประยุกต์ใช้ให้เกิดประโยชน์ในด้านต่างๆ เช่น การแบ่งขอบเขตภาพ (Image Segmentation) การรู้จำตัวพิมพ์อักษร (Character Recognition) เป็นต้น ซึ่งทั้งสองเทคนิควิธีดังกล่าวก็มีทั้งข้อดีและข้อด้อยที่ส่งผลต่อประสิทธิภาพในการจัดกลุ่มที่

แตกต่างกันออกไป จากการที่ผู้วิจัยได้ศึกษางานวิจัยที่ผ่านมาพบว่า มีการรวมอัลกอริทึมที่กล่าวมาข้างต้นเพื่อใช้เป็นทางเลือกในการจัดกลุ่มข้อมูล โดยงานวิจัยที่ผู้วิจัยได้ทำการศึกษามีดังนี้

Laerhoven (2001) เสนอการรวมอัลกอริทึมระหว่าง SOM และ K-Means เพื่อประยุกต์ใช้ในการจัดกลุ่ม (Clustering) จากตัว sensor ข้อมูลของผู้ใช้โทรศัพท์ว่าขณะใช้โทรศัพท์แล้วเดินผ่านตัว sensor ที่ติดตั้งอยู่ ณ จุดๆหนึ่ง ผลที่ได้จากงานวิจัยนี้ คือ สามารถจัดกลุ่มพฤติกรรมของผู้ใช้โทรศัพท์มือถือว่าขณะใช้มือถืออยู่นั้น กำลัง เดิน วิ่ง นั่ง หรือยืน ซึ่งข้อมูลที่ได้จากการจัดกลุ่มนี้นำไปใช้ประโยชน์โดยจะเก็บข้อมูลไว้และตอบกลับไปยังผู้ที่โทรเข้ามาหาผู้ใช้

Kuo และคณะ (2002) นำเสนอการรวมอัลกอริทึมระหว่าง SOM และ K-Means เพื่อนำมาใช้กับการทำ Market Segmentation โดยทำการเปรียบเทียบ 3 อัลกอริทึม คือ อัลกอริทึมแบบ 2 ขั้นตอน (Conventional two-stage method) ซึ่งคือ อัลกอริทึม Agglomerative Hierarchical กับ K-Means, อัลกอริทึม SOM และอัลกอริทึมที่นำเสนอคือ SOM กับ K-Means โดยใช้ค่า Wilk's Lambda เป็นตัววัดผลการวิจัย ผลการทดลองพบว่า SOM กับ K-Means มีประสิทธิภาพในการจัดกลุ่มดีกว่า agglomerative Hierarchical กับ K-Means

เลวันและก่อตระกูล (2003) นำเสนอวิธีการรู้จำตัวพิมพ์อักษรภาษาไทยด้วยเทคนิคการจัดกลุ่มข้อมูลผสมระหว่างเทคนิคฟัซซี่ซีมีน (Fuzzy C-Means) และเคมีน(K-Means) เพื่อเพิ่มประสิทธิภาพการรู้จำตัวอักษรภาษาไทย ทั้งในเชิงความเร็วและความถูกต้อง จากผลการทดสอบกับภาพตัวพิมพ์อักษรภาษาไทย 920 ตัว รวมทั้งภาพที่มีสัญญาณรบกวน พบว่าวิธีการรู้จำตัวอักษรที่นำเสนอมีผลการรู้จำถูกต้องสูงถึง 99.35 เปอร์เซ็นต์ และใช้เวลาในการรู้จำตัวอักษรเฉลี่ย 0.136 วินาทีต่อตัวอักษร สำหรับภาพตัวพิมพ์อักษรภาษาไทยที่มีสัญญาณรบกวนให้ผลการรู้จำสูงถึง 94.32 เปอร์เซ็นต์

Weng และ Liu (2004) ใช้วิธีการผสมผสานอัลกอริทึมระหว่าง SOM และ K-Means เพื่อนำมาประยุกต์ใช้ในเรื่องของการทำระบบการแนะนำลูกค้าของการตลาด แบบหนึ่งต่อหนึ่ง โดยข้อมูลที่ใช้ทำการวิจัย คือ ข้อมูลของลูกค้าที่ซื้อสินค้าและข้อมูลของสินค้าที่ถูกสั่งซื้อ หรือประเภทสินค้าและบริการที่ลูกค้าสนใจ โดยวิธีการทดลอง คือ ผู้วิจัยใช้คุณลักษณะ (Feature) ของสินค้าและลูกค้าในการจัดกลุ่ม โดยในขั้นตอนที่ 1 ใช้อัลกอริทึม SOM จะได้ข้อมูลจำนวนกลุ่ม จากนั้นนำข้อมูลที่ได้ คือจำนวนกลุ่มไปเป็นข้อมูลนำเข้าของอัลกอริทึมในขั้นตอนที่ 2 คือ K-Means และใช้ K-Means จัดกลุ่มข้อมูลออกมา สุดท้ายก็จะได้ข้อมูลที่ถูกจัดกลุ่มแล้ว

อรนุช ชัยหมื่น(2005) นำหลักการของการทำเหมืองข้อมูล (Data mining) มาใช้ในการจัดกลุ่มลูกค้าที่สั่งซื้อสินค้าหัตถกรรมของไทย โดยวิธีการที่ใช้ คือ วิธีการจัดกลุ่มแบบ 2 ขั้นตอน เพื่อให้เกิดประสิทธิภาพในการจัดกลุ่มสูงสุด งานวิจัยได้ทำการเปรียบเทียบประสิทธิภาพในการจัดกลุ่ม 2 วิธีการ คือ SOM กับK-Means เปรียบเทียบกับ Hierarchical Clustering กับ K-Means โดยทำการเปรียบเทียบจากค่าส่วนเบี่ยงเบนมาตรฐานของแต่ละกลุ่ม ที่แต่ละวิธีจัดกลุ่มได้ ผลการทดลอง การเปรียบเทียบประสิทธิภาพจากค่าส่วนเบี่ยงเบนมาตรฐานของทั้ง 2 วิธีการ พบว่าวิธีการที่ 1 SOM กับ K-Means จัดกลุ่มได้ดีกว่าวิธีการที่ 2 Hierarchical Clustering กับ K-Means และพบว่าเมื่อชุดข้อมูลทดลองมีขนาดใหญ่ขึ้น และมีการจัดกลุ่มที่มากขึ้น จะทำให้ค่าส่วนเบี่ยงเบนมาตรฐานของแต่ละกลุ่มและเวลาในการประมวลผลเพิ่มมากขึ้น

จากงานวิจัยที่กล่าวมาทั้งหมดข้างต้น จะพบว่าเทคนิคในการจัดกลุ่มข้อมูลแต่ละเทคนิควิธีต่างก็มีทั้งข้อดีและข้อด้อยที่ส่งผลต่อประสิทธิภาพในการจัดกลุ่มที่แตกต่างกัน ผู้วิจัยจึงมีความสนใจที่จะศึกษาและทำการเปรียบเทียบประสิทธิภาพของอัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอน (Two-Stage Clustering Algorithms) โดยการรวมอัลกอริทึมการจัดกลุ่มแบบขั้นตอนเดียวที่กล่าวมาแล้วข้างต้นเพื่อดูการทำงานร่วมกันของแต่ละอัลกอริทึม โดยในการศึกษาครั้งนี้จะเปรียบเทียบอัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอนทั้งหมด 4 แบบ คือ แบบที่ 1 SOM ร่วมกับ K-Means แบบที่ 2 SOM ร่วมกับ Fuzzy C-Means แบบที่ 3 Ward' s Method ร่วมกับ K-Means และแบบที่ 4 Ward' s Method ร่วมกับ Fuzzy C-Means เพื่อให้เห็นถึงประสิทธิภาพ ข้อดีและข้อเสียของอัลกอริทึมการจัดกลุ่มแต่ละแบบ เพื่อเป็นอีกทางเลือกหนึ่งในการนำไปใช้จัดกลุ่มข้อมูล ให้ผู้ที่สนใจสามารถเอาไปประยุกต์ใช้กับการจัดกลุ่มข้อมูลในชีวิตจริงได้

1.2 วัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาและเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลด้วยอัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอน (Two-Stage Clustering Algorithms) ทั้งหมด 4 แบบซึ่งคือ แบบที่ 1 SOM ร่วมกับกับ K-Means แบบที่ 2 SOM ร่วมกับ Fuzzy C-Means แบบที่ 3 Ward' s Method ร่วมกับ K-Means และแบบที่ 4 Ward' s Method ร่วมกับ Fuzzy C-Means

2. เพื่อเป็นอีกทางเลือกหนึ่งในการเอาไปประยุกต์ใช้กับการจัดกลุ่มข้อมูลในชีวิตจริงด้วยอัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอน (Two-Stage Clustering Algorithms)

1.3 สมมติฐานในการศึกษา

ภายใต้สถานการณ์ที่แตกต่างกัน นั่นคือ จำนวนตัวแปร ขนาดตัวอย่าง ลักษณะความสัมพันธ์ของตัวแปร ลักษณะการซ้อนทับกันและไม่ซ้อนทับกันของข้อมูล จำนวนกลุ่ม และวิธีการที่ใช้ในการจัดกลุ่ม มีผลต่อประสิทธิภาพในการจัดกลุ่มแตกต่างกัน

1.4 ขอบเขตของการศึกษา

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลสำหรับการวิเคราะห์กลุ่ม ซึ่งมีขอบเขตการศึกษาดังต่อไปนี้

1. จำนวนตัวแปรที่ใช้ศึกษาคือ 2, 4 และ 6 ตัวแปร

2. ขนาดตัวอย่าง 25, 100 และ 300 แทนข้อมูลขนาดเล็ก, ขนาดกลาง และขนาดใหญ่ตามลำดับ

3. จำนวนกลุ่มที่จำลองคือ 2, 4 และ 6 กลุ่ม

4. รูปแบบการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution)

5. ลักษณะข้อมูล

    5.1 ข้อมูลในแต่ละกลุ่มไม่ซ้อนทับกัน (Nonoverlapping Clusters)

    5.2 ข้อมูลในแต่ละกลุ่มซ้อนทับกัน (40% Overlapping Clusters)

6. อัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอน (Two-Stage Clustering Algorithms) ที่จะศึกษาได้แก่

    6.1 แบบที่ 1 SOM ร่วมกับ K-Means

    6.2 แบบที่ 2 SOM ร่วมกับ Fuzzy C-Means

    6.3 แบบที่ 3 Ward' s Method ร่วมกับ K-Means

    6.4 แบบที่ 4 Ward' s Method ร่วมกับ Fuzzy C-Means

7. รูปแบบความสัมพันธ์ของตัวแปร

    7.1 ตัวแปรอิสระต่อกัน

    7.2 ตัวแปรมีความสัมพันธ์กัน โดย $\rho = Corr(X_i, X_j) \sim Unif(-1,1)$

    7.3 ตัวแปรมีความสัมพันธ์กัน โดย $\rho = Corr(X_i, X_j) \sim Unif(0,1)$

8. เกณฑ์ในการวัดประสิทธิภาพ

8.1 ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD)

8.2 Wilk' s Lambda

8.3 Misclassification Rate

9. จำนวนการทำซ้ำคือ 500 รอบ ในแต่ละสถานการณ์ของการทดลอง

1.5 วิธีดำเนินการศึกษา

1. จำลองข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) โดยจำนวนตัวแปรที่ศึกษา คือ 2, 4 และ 6 ตัวแปร และกำหนดลักษณะความสัมพันธ์ของตัวแปร 3 ลักษณะคือ ตัวแปรอิสระต่อกัน ตัวแปรมีความสัมพันธ์กัน โดยค่าความสัมพันธ์สุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ในช่วง (-1,1) และตัวแปรมีความสัมพันธ์กัน โดยค่าความสัมพันธ์สุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ในช่วง (0,1)

2. กำหนดขนาดตัวอย่างเท่ากับ 25, 100 และ 300

3. จำลองข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปรตามลักษณะที่กำหนดในขั้นตอนที่ 1 และขั้นตอนที่ 2 โดยจำลองข้อมูลเป็น 2, 4 และ 6 กลุ่ม และทำซ้ำจำนวน 500 ครั้งในแต่ละชุดข้อมูล

4. ทำการจัดกลุ่มข้อมูลที่จำลองขึ้น โดยใช้อัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอน (Two- Stage Clustering Algorithms) ทั้ง 4 แบบ ดังนี้

4.1 แบบที่ 1 SOM ร่วมกับ K-Means โดยในขั้นตอนที่ 1 ใช้ SOM ในการกำหนดจำนวน กลุ่มและตัวแทนกลุ่มเริ่มต้น และในขั้นตอนที่ 2 ใช้ K-Means จัดกลุ่มข้อมูล

4.2 แบบที่ 2 SOM ร่วมกับ Fuzzy C-Means โดยในขั้นตอนที่ 1 ใช้ SOM ในการจำนวน กลุ่มและตัวแทนกลุ่มเริ่มต้น และในขั้นตอนที่ 2 ใช้ Fuzzy C-Means จัดกลุ่มข้อมูล

4.3 แบบที่ 3 Ward' s Method ร่วมกับ K-Means โดยในขั้นตอนที่ 1 ใช้ Ward' s Method ในการกำหนดจำนวนกลุ่มและตัวแทนกลุ่มเริ่มต้น และในขั้นตอนที่ 2 ใช้ K-Means จัดกลุ่มข้อมูล

4.4 แบบที่ 4 Ward' s Method ร่วมกับ Fuzzy C-Means โดยในขั้นตอนที่ 1 ใช้ Ward' s Method ในการกำหนดจำนวนกลุ่มและตัวแทนกลุ่มเริ่มต้น และในขั้นตอนที่ 2 ใช้ K-Means จัดกลุ่มข้อมูล

5. วัดประสิทธิภาพการจัดกลุ่มข้อมูลตามเกณฑ์ที่ได้ระบุไว้

6. เปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูล

7. สรุปผลที่ได้จากการเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูล ว่าวิธีการจัดกลุ่มแบบใดให้ผลการจัดกลุ่มดีที่สุด

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. ทำให้ทราบถึงข้อดีและข้อด้อยของการจัดกลุ่มด้วยวิธีการต่างๆ และสามารถเลือกวิธีการในการจัดกลุ่มมาใช้ให้เหมาะสมกับชุดข้อมูลแต่ละชนิดได้

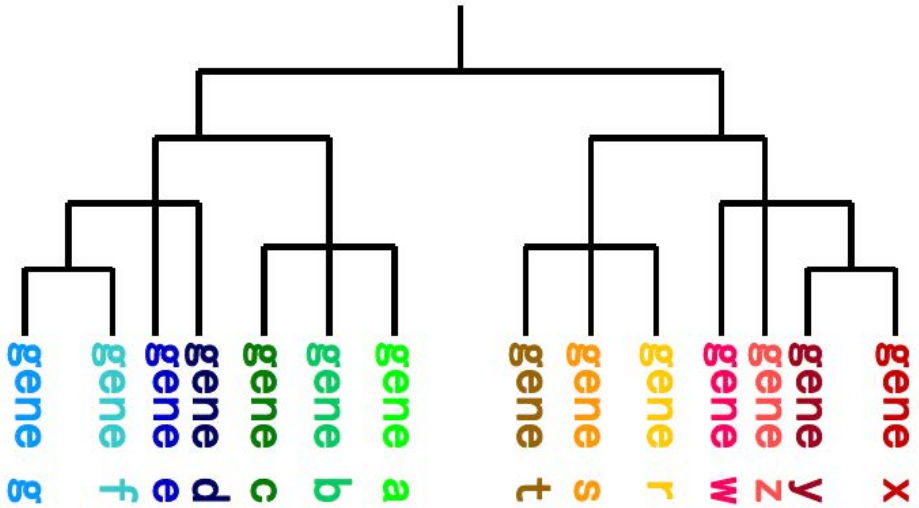
2. ในการศึกษาอัลกอริทึมจัดกลุ่มแบบ 2 ขั้นตอน (Two-Stage Clustering Algorithm) เป็นการนำเอาข้อดีและข้อด้อยของวิธีการจัดกลุ่มที่มีอยู่แล้วมาผสมผสานกัน เพื่อให้เกิดประสิทธิภาพในการจัดกลุ่มที่ดีขึ้น และเพื่อเป็นอีกทางเลือกหนึ่งในการนำเทคนิควิธีการจัดกลุ่มแต่ละวิธีไปใช้ให้เกิดประโยชน์กับข้อมูลในชีวิตจริง

บทที่ 2

ทฤษฎีและตัวสถิติที่เกี่ยวข้อง

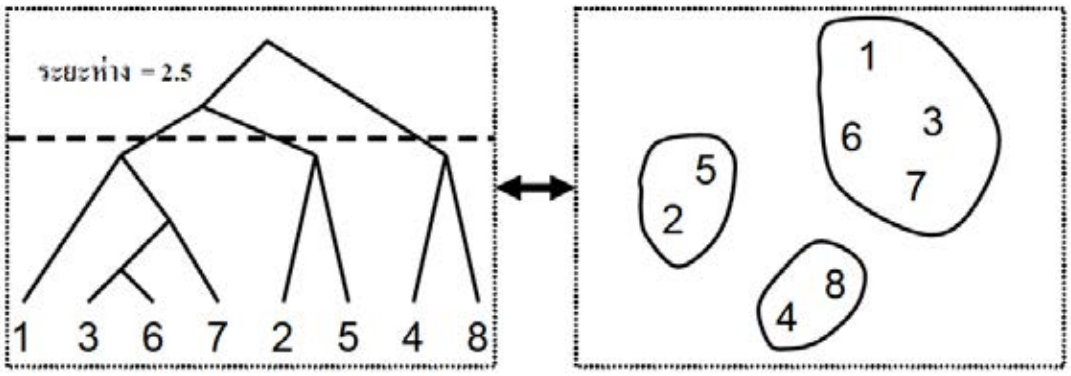
2.1 อัลกอริทึมการจัดกลุ่มแบบลำดับขั้น (Hierarchical Clustering Algorithm)

อัลกอริทึมจัดกลุ่มแบบลำดับขั้น (Johnson, 1967) เป็นเทคนิควิธีที่นิยมมากที่สุด เนื่องจากคำนวณได้ง่าย ซึ่งสามารถนำไปใช้ในการจัดกลุ่มข้อมูลที่มีคุณลักษณะซ้อน (Nested) เช่น ข้อมูลการแสดงออกของยีน (Gene Expression) (Eisen*et al*., 1998) แสดงในภาพที่ 2.1



ภาพที่ 2.1 การจัดกลุ่มแบบลำดับขั้นของข้อมูลการแสดงออกของยีน
ที่มา : Eisen*et al*. (1998)

การจัดกลุ่มแบบลำดับขั้นสามารถแสดงในรูปแผนภูมิต้นไม้ที่เรียกว่า เดนโดรแกรม (Dendrogram) โดยกำหนดจำนวนกลุ่มได้จากการลากเส้นตัดผ่านเดนโดรแกรม ด้วยค่าระยะห่าง หรือค่าความคล้าย



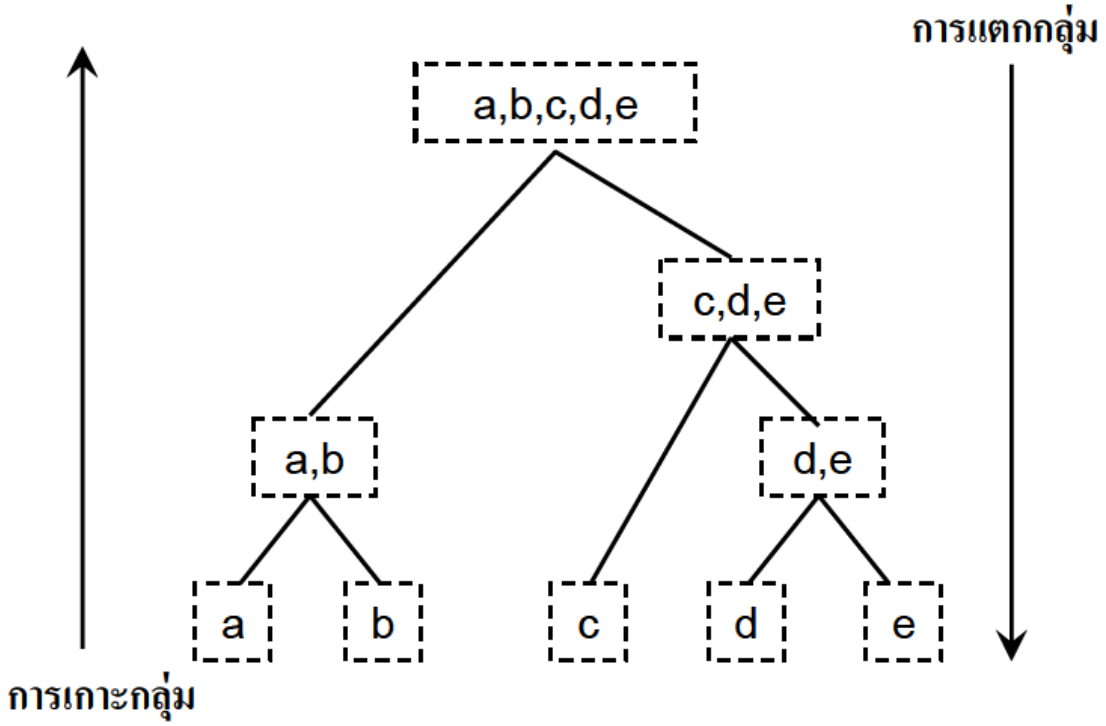
ภาพที่ 2.2 ตัวอย่างการสร้างและจัดกลุ่มแบบลำดับขั้นด้วยเดนโดรแกรม
ที่มา : ปิยธิดา, 2552

จากภาพที่ 2.2 ทางด้านซ้าย เมื่อกำหนดระยะห่างเท่ากับ 2.5 ลากเส้นตัดเดนโดรแกรม จะสามารถจัดกลุ่มข้อมูลได้เป็น 3 กลุ่ม ดังแสดงในภาพที่ 2.2 ทางด้านขวา

การจัดกลุ่มแบบลำดับขั้นสามารถแบ่งออกเป็น 2 ประเภท คือ

1.1 การเกาะกลุ่ม (Agglomerative) มีหลักการทำงาน คือ เริ่มจากการจัดกลุ่มข้อมูลเป็นจำนวน n กลุ่ม จากข้อมูลทั้งหมด n ตัว ให้เป็นบัพภายนอก (External node) หลังจากนั้นทำการรวมข้อมูลที่มีค่าระยะห่างระหว่าง 2 กลุ่มน้อยที่สุด โดยรวมข้อมูลครั้งละ 2 กลุ่ม ทำซ้ำจนกระทั่งสุดท้ายรวมข้อมูลเหลือเพียง 1 กลุ่ม ซึ่งเป็นบัพราก (Root node)

1.2 การแตกกลุ่ม (Divisive) มีหลักการในการทำงานคล้ายวิธีของการเกาะกลุ่ม แต่มีความแตกต่างกันที่วิธีการแตกกลุ่มนี้ จะทำงานในทิศตรงข้ามกัน กล่าวคือ การแตกกลุ่มจะเริ่มจากบัพรากไปสู่บัพภายนอกหรือจากบนลงล่างนั่นเอง ซึ่งวิธีนี้ไม่เป็นที่นิยม เนื่องจากยุ่งยากและใช้เวลาในการคำนวณมากกว่าวิธีของการเกาะกลุ่ม (Lebart and Rajman, 2000)



ภาพที่ 2.3 ตัวอย่างการสร้างกลุ่มแบบเกาะกลุ่มและแตกกลุ่ม
ที่มา : ปิยธิดา, 2552

สำหรับขั้นตอนวิธีการคำนวณของการจัดกลุ่มแบบลำดับขั้น โดยการเกาะกลุ่มมีวิธีการดังนี้

1) แบ่งข้อมูลออกเป็น n กลุ่ม โดยในแต่ละกลุ่มมีสมาชิกเพียง 1 ตัว

2) รวมกลุ่มข้อมูล 2 กลุ่ม ให้อยู่ในกลุ่มเดียวกัน โดยพิจารณาจากค่าระยะห่าง และใช้หลักเกณฑ์ในการรวมกลุ่มแบบ Ward' s Minimum Variance Method

3) พิจารณาว่าควรรวมกลุ่มข้อมูลกลุ่มที่ 3 เข้ากับข้อมูล 2 กลุ่มแรกหรือรวมเข้ากับข้อมูลกลุ่มใหม่อีกกลุ่มหนึ่ง โดยการพิจารณาจากค่าระยะห่างและใช้หลักเกณฑ์ในการรวมกลุ่มเช่นเดียวกับขั้นตอนที่ 2

4) ทำซ้ำขั้นตอนที่ 3ไปเรื่อยๆ จนกระทั่งข้อมูลทุกกลุ่มอยู่ในกลุ่มเดียวกัน นั่นคือ สุดท้ายเหลือข้อมูลเพียง 1 กลุ่ม และกลุ่มข้อมูลใดที่ถูกจัดกลุ่มแล้วจะไม่มีการเปลี่ยนแปลงอีก

5) สร้างเดนโดรแกรม

6) กำหนดค่าระยะห่าง แล้วนำมาลากเส้นตัดผ่านเดนโดรแกรมเพื่อจัดกลุ่มข้อมูล

หลักเกณฑ์ในการรวมกลุ่มสำหรับวิธีการเกาะกลุ่มที่นำมาศึกษาในครั้งนี้ คือ

หลักเกณฑ์ในการรวมกลุ่มแบบ Ward' s Minimum Variance Methodวิธีการจัดกลุ่มแบบวอร์ดซึ่งบางครั้งเรียกว่าวิธีแปรปรวนต่ำสุด เป็นวิธีที่พิจารณาผลบวกของค่าคลาดเคลื่อนกำลังสอง (Error Sum Square : ESS) โดยจะหากลุ่ม 2 กลุ่มที่เมื่อรวมกันแล้วทำให้ผลบวกของค่าคลาดเคลื่อนกำลังสองของกลุ่มใหม่ที่รวมกันแล้วมีค่าต่ำสุด เมื่อเทียบกับเมื่อนำไปรวมกับกลุ่มอื่นๆ (กัลยา,2552) โดยมีขั้นตอนดังนี้

1) กำหนดให้มี n กลุ่ม โดย 1 กลุ่มมี 1 หน่วย ทำให้ $ESS_i = 0; i = 1, 2, ..., n$

2) ลดเหลือ $n-1$กลุ่ม โดยรวม 2 กลุ่มเข้าด้วยกันที่ทำให้ ESS ของกลุ่มที่รวมแล้วมีค่าต่ำสุด ทำเช่นนี้ไปเรื่อยๆ ในแต่ละขั้นจะลดลง 1 กลุ่ม จนกระทั่งเหลือ 1 กลุ่ม

$$ESS = \sum_{j=1}^{k}[\sum_{i=1}^{n_j} X_{ij}^2 - \frac{1}{n_j}(\sum_{i=1}^{n_j} X_{ij})^2]$$

โดยที่ X_{ij} คือ ค่าของข้อมูลคนที่ i ที่อยู่ในกลุ่มที่ j

k คือ จำนวนกลุ่มในแต่ละขั้น

n_j คือ จำนวนหน่วยหรือคนในกลุ่มที่ j

ข้อเด่นของเทคนิคการจัดกลุ่มแบบวอร์ด

1) ไม่ต้องกำหนดจำนวนกลุ่มเริ่มต้น

ข้อด้อยของเทคนิคการจัดกลุ่มแบบวอร์ด

1) ไม่เหมาะกับข้อมูลขนาดใหญ่ เพราะใช้เวลาในการคำนวณมาก

2) ได้รับผลกระทบจากค่าผิดปกติ (Outlier) ได้ง่าย

ตัวอย่างที่ 2.1 การจัดกลุ่มโดยวิธีการจัดกลุ่มแบบลำดับขั้นประเภทการเกาะกลุ่ม แสดงดังนี้

สมมติมีข้อมูลของคน 5 คนคือ A, B, C, D และ E โดยมีรายละเอียดดังนี้

คนที่	ค่าข้อมูล
A	1
B	2
C	3
D	4
E	5

ขั้นที่ 1 ให้ 1 กลุ่ม = 1 คน ดังนั้นจึงมี 5 กลุ่ม $ESS_i = 0; i = 1,2,...,5$

ขั้นที่ 2 ทำการลดจำนวนกลุ่มให้เหลือ 4 กลุ่ม โดยพิจารณาค่า ESS ของการรวมกลุ่ม 2 กลุ่มเข้าด้วยกัน ดังนั้นกลุ่มที่จัดออกมา กลุ่มหนึ่งจะมี 2 คน ที่เหลืออีก 3 กลุ่ม จะมีกลุ่มละ 1 คน คำนวณ ESS ของการนำ 2 กลุ่มมารวมกันได้ดังตารางที่ 2.1.1

ตารางที่ 2.1.1 แสดงค่า ESS ของการรวมกลุ่มเป็น 4 กลุ่มทั้งหมดที่เป็นไปได้

	กลุ่มที่				ESS
	1	2	3	4	
1	A,B	C	D	E	2.5
2	A,C	B	C	E	2
3	A,D	B	C	E	4.5
4	A,E	B	C	D	8
5	B,C	A	C	E	0.5
6	B,D	A	C	E	2
7	B,E	A	C	D	4.5
8	C,D	A	B	E	0.5
9	C,E	A	B	D	2
10	D,E	A	B	C	0.5

จากตารางที่ 2.1.1 เมื่อพิจารณา ESS พบว่าค่า ESS ของการรวมกลุ่ม ในรูปแบบที่ 5, 8 และ 10 มีค่าต่ำสุด คือ 0.5 ดังนั้นการรวมกลุ่ม 4 กลุ่มที่เป็นไปได้ อาจเป็นรูปแบบที่ 5 หรือรูปแบบที่ 8 หรือรูปแบบที่ 10 ก็ได้ โดยรูปแบบที่ 5 รวม B และ Cให้อยู่ในกลุ่มเดียวกัน หรือรูปแบบที่ 8 รวม Cและ D ให้อยู่ในกลุ่มเดียวกัน หรือรูปแบบที่ 10 รวม D และ E ให้อยู่ในกลุ่มเดียวกัน ในที่นี้เลือกรูปแบบที่ 5 คือ รวม B และ C ให้อยู่ในกลุ่มเดียวกันดังนั้น 4 กลุ่มที่รวมได้ จะเป็นดังนี้ { A }, { B,C }, { D }, { E }

ขั้นที่ 3 ทำการลดจำนวนกลุ่มให้เหลือ 3 กลุ่ม โดยรวมกลุ่มครั้งละ 2 กลุ่มจากกลุ่มทั้งหมด 4 กลุ่ม ในขั้นตอนที่ 2 คำนวณหาค่า *ESS* ของการรวมกลุ่มได้ดังตารางที่ 2.1.2

ตารางที่ 2.1.2 แสดงค่า *ESS* ของการรวมกลุ่มเป็น 3 กลุ่มทั้งหมดที่เป็นไปได้

	กลุ่มที่			*ESS*
	1	2	3	
1	A,B,C	D	E	2
2	B,C,D	A	E	2
3	B,C,E	A	D	4.67
4	B,C	A,D	E	5
5	B,C	A,E	D	8.5
6	B,C	D,E	A	1

จากตารางที่ 2.1.2 พบว่าค่า *ESS* ของการรวมกลุ่มในรูปแบบที่ 6 มีค่าต่ำที่สุด คือ 1 ดังนั้นจึงเหลือ 3 กลุ่ม ดังนี้ { A }, { B,C }, { D,E }

ขั้นที่ 4 ทำการลดจำนวนกลุ่มให้เหลือ 2 กลุ่ม โดยรวมกลุ่มครั้งละ 2 กลุ่มจากกลุ่มทั้งหมด 3 กลุ่ม ในขั้นตอนที่ 3 คำนวณหาค่า *ESS* ของการรวมกลุ่มได้ดังตารางที่ 2.1.3

ตารางที่ 2.1.3 แสดงค่า *ESS* ของการรวมกลุ่มเป็น 2 กลุ่มทั้งหมดที่เป็นไปได้

	กลุ่มที่		*ESS*
	1	2	
1	A,B,C	D,E	2.5
2	A,D,E	B,C	9.17
3	B,C,D,E	A	5

จากตารางที่ 2.1.3 พบว่าค่า *ESS* ของการรวมกลุ่มในรูปแบบที่ 1 มีค่าต่ำที่สุด คือ 2.5 ดังนั้นจึงเหลือ 2 กลุ่ม ดังนี้ { A,B,C }, { D,E }

ขั้นที่ 5 คำนวณค่า ESS ของ 2 กลุ่มที่ได้ในขั้นตอนที่ 4 ดังนั้น $ESS_{ABCDE} = 10$

2.2 **อัลกอริทึมการจัดกลุ่มแบบเค-มีน** (K-Means Clustering Algorithm)

เทคนิควิธีการจัดกลุ่มข้อมูลโดยอัลกอริทึมจัดกลุ่มแบบเค-มีน (K-Means Clustering Algorithm) นั้นมีการพัฒนาและนำเสนอโดย Mac Queen ในปี 1967 ซึ่งอัลกอริทึมการจัดกลุ่มแบบเค-มีนนี้ สมาชิกแต่ละตัวภายในกลุ่มจะมีระยะใกล้จุดศูนย์กลางหรือตัวแทนของกลุ่ม (Centroid) นั้นมากที่สุด ขั้นตอนการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมการจัดกลุ่มแบบเค-มีน ประกอบด้วย การกำหนดจำนวนกลุ่มเริ่มต้น การสุ่มตัวแทนกลุ่ม การจัดข้อมูลแต่ละตัวเข้ากลุ่ม และสุดท้ายคือ การปรับปรุงตัวแทนกลุ่มในแต่ละกลุ่มขั้นตอนวิธีการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมจัดกลุ่มแบบเค-มีน มีขั้นตอนดังนี้

1) กำหนดจำนวนกลุ่มที่ต้องการ โดยกำหนดให้ k เท่ากับจำนวนกลุ่มที่ต้องการแบ่งกลุ่ม

2) สุ่มจุดศูนย์กลางของแต่ละกลุ่ม

3) คำนวณค่าระยะห่างของข้อมูลทุกตัวกับตัวแทนของกลุ่มในแต่ละกลุ่ม ว่าข้อมูลตัวนั้นสมควรที่จะจัดให้อยู่ในกลุ่มใด โดยพิจารณาจากค่าระยะห่างที่น้อยที่สุด

4) ปรับปรุงจุดศูนย์กลางของข้อมูลแต่ละกลุ่มใหม่ โดยคำนวณค่าเฉลี่ยของข้อมูลภายในกลุ่ม ($\bar{x}_j$) เพื่อใช้เป็นตัวแทนใหม่ของกลุ่ม

ค่าเฉลี่ยของข้อมูลภายในกลุ่ม ($\bar{x}_j$)คำนวณได้ดังนี้

$$\bar{x}_j = \frac{1}{n_i}\sum_{i=1}^{n_i} x_{ij}; i = 1,2,...,n_i; j = 1,2,...,k$$

โดยที่ n_i คือ จำนวนข้อมูลในกลุ่มแต่ละครั้งของการแบ่งกลุ่ม

x_{ij} คือ ข้อมูลในหน่วยที่ i กลุ่มที่ j

$\bar{x}_j$ คือ ค่าเฉลี่ยของข้อมูลในกลุ่มที่ j

k คือ จำนวนกลุ่มทั้งหมด

5) ทำขั้นตอนที่ 3 และ 4 ซ้ำ จนกระทั่งสมาชิกทุกตัวในกลุ่มแต่ละกลุ่มจะไม่มีการเปลี่ยนแปลงไปยังกลุ่มอื่นๆ

ข้อเด่นของอัลกอริทึมจัดกลุ่มแบบเค-มีน

1) ง่ายและนิยมนำไปใช้งาน

2) จัดกลุ่มข้อมูลได้รวดเร็ว

ข้อด้อยของอัลกอริทึมจัดกลุ่มแบบเค-มีน

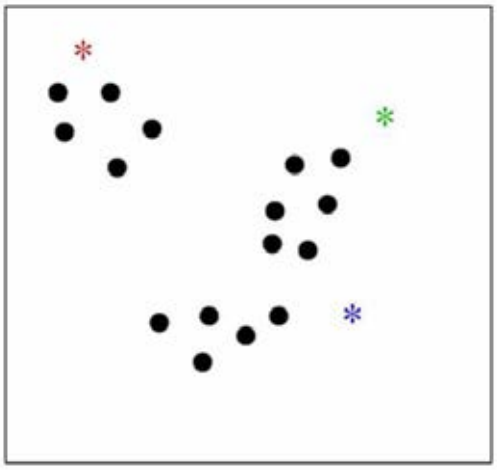
1) การกำหนดจำนวนกลุ่มและตัวแทนกลุ่มเริ่มต้น มีผลต่อประสิทธิภาพการจัดกลุ่มทั้งในเชิงเวลาและความถูกต้อง

2) ขั้นตอนวิธีการจัดกลุ่มข้อมูลแบบเค-มีนจะมีปัญหากับข้อมูลที่มี outlier

3) ข้อมูลมีโอกาสเป็นสมาชิกเพียงกลุ่มใดกลุ่มหนึ่งเท่านั้น

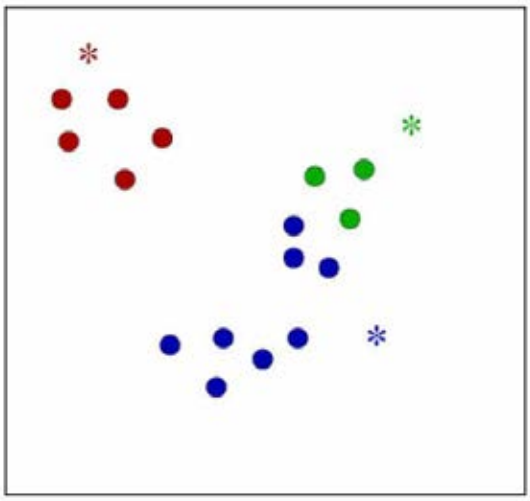
4) ผลลัพธ์ของการจัดกลุ่มที่ได้เป็นแบบ local optimum

ตัวอย่างการจัดกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเค-มีน เมื่อแบ่งข้อมูลเป็น 3 กลุ่ม แสดงดังภาพที่ 4-7



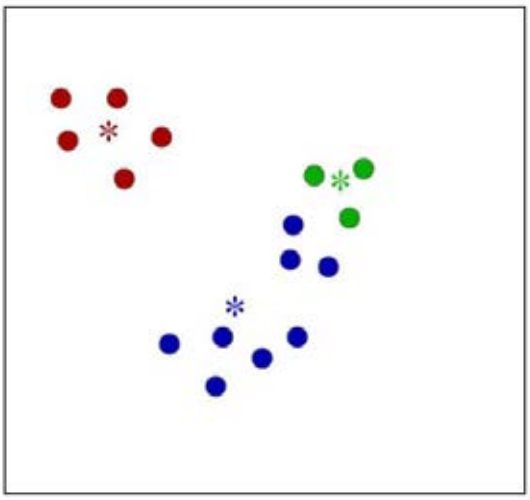
ภาพที่ 2.4 การกำหนดจุดข้อมูลเพื่อใช้ในการจัดกลุ่มข้อมูลแบบเค-มีน
ที่มา : Vipin Kumar, Parallel Issues in Data Mining, VECPAR 2002

จากภาพที่ 2.4 สัญลักษณ์วงกลมสีดำคือข้อมูลใดๆ สัญลักษณ์ดอกจันสีแดง สีเขียว และสีน้ำเงิน คือจุดข้อมูลที่สุ่มได้และใช้แทนจุดศูนย์กลางของกลุ่มข้อมูล ซึ่งในกรณีนี้กำหนดจำนวนกลุ่มเป็น 3 กลุ่ม (k = 3)



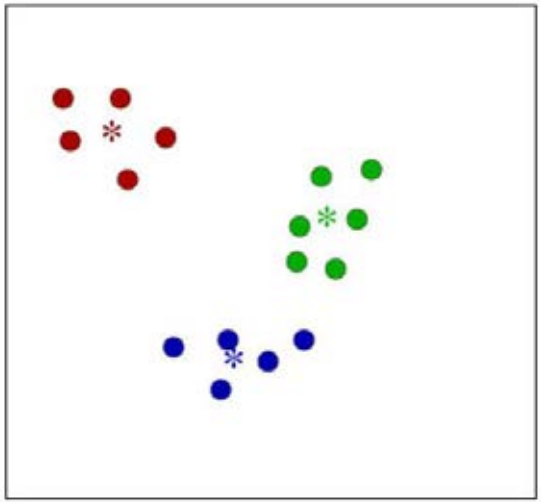
ภาพที่ 2.5 ตัวอย่างการจัดกลุ่มข้อมูล โดยพิจารณาค่าระยะห่างระหว่างข้อมูลกับจุดศูนย์กลางของกลุ่มข้อมูลที่สุ่มได้ในครั้งแรก
ที่มา : Vipin Kumar, Parallel Issues in Data Mining, VECPAR 2002

จากภาพที่ 2.5 สัญลักษณ์วงกลมสีแดง สีเขียว และสีน้ำเงิน คือข้อมูลใดๆ ที่ถูกจัดกลุ่มโดยพิจารณาค่าระยะห่างระหว่างข้อมูลกับจุดศูนย์กลางของกลุ่มข้อมูลที่สุ่มได้ในครั้งแรก สัญลักษณ์ดอกจันสีแดง สีเขียว และสีน้ำเงิน คือ จุดศูนย์กลางของกลุ่มข้อมูลที่สุ่มได้ในครั้งแรก



ภาพที่ 2.6 ตัวอย่างการปรับจุดศูนย์กลางของข้อมูลแต่ละกลุ่มไปยังตำแหน่งที่ครอบคลุมข้อมูลบริเวณใกล้เคียงภายในกลุ่มเดียวกันให้ได้มากที่สุด โดยการหาค่าเฉลี่ยของข้อมูลภายในกลุ่ม
ที่มา : Vipin Kumar, Parallel Issues in Data Mining, VECPAR 2002

จากภาพที่ 2.6 สัญลักษณ์วงกลมสีแดง สีเขียว และสีน้ำเงิน คือข้อมูลใดๆ สัญลักษณ์ดอกจันสีแดง สีเขียว และสีน้ำเงิน คือ จุดศูนย์กลางของกลุ่มข้อมูลที่ได้รับการปรับให้อยู่ในตำแหน่งที่ครอบคลุมข้อมูลในกลุ่มเดียวกันมากขึ้น



ภาพที่ 2.7 การจัดกลุ่มที่เสร็จสมบูรณ์ โดยจุดศูนย์กลางที่ได้จะอยู่ในตำแหน่งที่เป็นตัวแทนของข้อมูลแต่ละกลุ่ม
ที่มา : Vipin Kumar, Parallel Issues in Data Mining, VECPAR 2002

จากภาพที่ 2.7 สัญลักษณ์วงกลมสีแดง สีเขียว และสีน้ำเงิน คือข้อมูลใดๆ สัญลักษณ์ดอกจันสีแดง สีเขียว และสีนำเงิน คือจุดศูนย์กลางของข้อมูลที่เป็นตัวแทนที่เหมาะสมที่สุดของข้อมูลในแต่ละกลุ่ม

ตัวอย่างที่ 2.2 การจัดกลุ่มโดยอัลกอริทึมจัดกลุ่มแบบเคมีน แสดงดังนี้

สมมติมีข้อมูลอยู่ 4 รายการ คือ A, B, C และ D โดยแต่ละชุดข้อมูลมีตัวแปร 2 ตัวแปร คือ X_1 และ X_2 กำหนดให้มีการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม (k = 2) ชุดข้อมูลมีรายละเอียดดังแสดงในตารางที่ 2.2.1

ตารางที่ 2.2.1 ข้อมูลตัวอย่างที่ 2.2 : A, B, C และ D และมีตัวแปรจำนวน 2 ตัวแปร

รายการข้อมูล	ตัวแปร	
	X_1	X_2
A	-5	3
B	1	-1
C	-1	2
D	3	2

1) เนื่องจากต้องการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม จึงสุ่มเลือกจุดศูนย์กลางข้อมูลมา 2 จุด เพื่อใช้เป็นตัวแทนของแต่ละกลุ่ม ในที่นี้จะจัดให้ A กับ B อยู่กลุ่มเดียวกัน และ C กับ D อยู่ในกลุ่มเดียวกันแสดงในตารางที่ 2.2.2

ตารางที่ 2.2.2 จุดศูนย์กลางเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 2.2

กลุ่ม	จุดกลางกลุ่ม	
	$\bar{X}_1$	$\bar{X}_2$
(AB)	$\frac{(-5)+1}{2}=-2$	$\frac{3+(-1)}{2}=1$
(CD)	$\frac{(-1)+3}{2}=1$	$\frac{2+2}{2}=2$

ในที่นี้จะให้R1 แทน จุดกลางกลุ่มของ A กับ B และ R2 แทน จุดกลางกลุ่มของ C กับ D

2) คำนวณค่าระยะห่างแบบยุคลิดกำลังสองระหว่างข้อมูลกับตัวแทนกลุ่ม (D_{ij}) ที่สุ่มเลือกในขั้นตอนที่ 1 ในที่นี้มี 2 ตัวแปร (p=2)

$$D_{ij} = \sum_{p=1}^{2} (X_{pi} - X_{pj})^2$$

คำนวณค่า D_{ij} ของข้อมูลตัวอย่าง ค่าที่ได้จะเป็นค่าระยะห่างระหว่างจุดข้อมูลแต่ละตัวกับตัวแทนกลุ่ม เมื่อพิจารณาค่าระยะห่างที่จุดข้อมูล A คือ D(A,R1) และ D(A,R2) จะเลือกให้ข้อมูลไปอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่มีค่าระยะห่างที่คำนวณได้น้อยที่สุด

$$D(A,R1) = (-5-(-2))^2 + (3-1)^2 = 13$$

$$D(A,R2) = (-5-1)^2 + (3-2)^2 = 37$$

จะเห็นว่า ค่าระยะห่างระหว่างจุดข้อมูล A กับจุดกลางกลุ่มR1 (D(A,R1)) มีค่าน้อยกว่าระยะห่างระหว่างจุดข้อมูล A กับจุดกลางกลุ่มR2 (D(A,R2)) ดังนั้น A จึงควรอยู่กลุ่มเดียวกับ R1

3) คำนวณค่าระยะห่างของตัวแทนกลุ่มกับข้อมูลที่เหลือจนครบทั้งหมด

$$D(B,R1) = (1-(-2))^2 + (-1-1)^2 = 13,\ D(B,R2) = (1-1)^2 + (-1-2)^2 = 9$$

$$D(C,R1) = ((-1)-(-2))^2 + (2-1)^2 = 2,\ D(C,R2) = (-1-1)^2 + (2-2)^2 = 4$$

$$D(D,R1) = (3-(-2))^2 + (2-1)^2 = 26,\ D(D,R2) = (3-1)^2 + (2-2)^2 = 4$$

จะเห็นว่า ระยะห่างระหว่างข้อมูล B กับจุดกลางกลุ่มR2 มีค่าน้อยกว่าระยะห่างระหว่าง B กับจุดกลางกลุ่มR1 ดังนั้น B จึงควรอยู่ในกลุ่มเดียวกับ R2 ในทำนองเดียวกันเมื่อพิจารณาค่าระยะห่างของ D กับจุดกลางกลุ่ม R2 พบว่าD ควรอยู่ในกลุ่มเดียวกับ R2 เช่นกัน นั่นคือ Bและ D จะอยุ่ในกลุ่มเดียวกับ R2 ส่วน ระยะห่างระหว่างข้อมูล C กับจุดกลางกลุ่ม R1 มีค่าน้อยกว่าระยะห่างระหว่าง C กับจุดกลางกลุ่ม R2 ดังนั้น C จึงควรอยู่ในกลุ่มเดียวกับ R1

4) เมื่อข้อมูลทุกตัวได้รับการกำหนดกลุ่มแล้ว จะทำการปรับเปลี่ยนจุดศูนย์กลางของแต่ละกลุ่ม โดยใช้ค่าเฉลี่ยของสมาชิกภายในกลุ่ม ดังแสดงในตารางที่ 2.2.3

ตารางที่ 2.2.3 ค่าเฉลี่ยของข้อมูลแต่ละกลุ่ม เมื่อข้อมูลมีการเปลี่ยนแปลงกลุ่ม

กลุ่ม	จุดกลางกลุ่ม	
	$\bar{X}_1$	$\bar{X}_2$
(AC)	$\frac{(-5)+(-1)}{2}=-3$	$\frac{3+2}{2}=2.5$
(BD)	$\frac{1+3}{2}=2$	$\frac{-1+2}{2}=0.5$

5) นำค่าเฉลี่ยของข้อมูลที่ได้มาใช้เป็นตัวแทนกลุ่ม แล้วทำซ้ำขั้นตอนที่ 2-4 จนกว่าข้อมูลจะไม่มีการเปลี่ยนแปลงกลุ่ม โดยสุดท้ายแล้วจะได้ผลดังแสดงในตารางที่ 2.2.4

ตารางที่ 2.2.4 ระยะห่างยุคลิดกำลังสองระหว่างข้อมูลกับค่าเฉลี่ยของข้อมูลแต่ละกลุ่ม

กลุ่ม	ค่าเฉลี่ยของข้อมูลในกลุ่ม			
	A	B	C	D
(AC)	4.25	28.25	4.25	36.25
(BD)	55.25	3.25	11.25	3.25

จากตารางที่ 2.2.4 พบว่าไม่มีการเปลี่ยนแปลงกลุ่มแล้วจึงหยุดดำเนินการ และสรุปได้ว่าสามารถจัดกลุ่มข้อมูลออกเป็น 2 กลุ่ม ดังนี้ กลุ่มที่ 1 คือ A กับ C และกลุ่มที่ 2 คือ Bกับ D

2.3 อัลกอริทึมการจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C-Means clustering algorithm)

อัลกอริทึมการจัดกลุ่มแบบฟัซซี่ซีมีน (Fuzzy C-Means algorithm) นำเสนอโดย Dunn (1973) และมีการนำมาพัฒนาต่อโดย Bezdek ในปี ค.ศ. 1981 จนกระทั่งปัจจุบันได้มีการพัฒนาอัลกอริทึมการจัดกลุ่มข้อมูลแบบฟัซซี่ซีมีนมาประยุกต์ใช้ในงานด้านต่างๆ เช่น การรู้จำแบบ (Pattern Recognition) การประมวลผลภาพ (Image Processing) รวมถึงการทำเหมืองข้อมูล เป็นต้น

หลักการทำงานของขั้นตอนวิธีการจัดกลุ่มแบบฟัซซี่ซีมีน พัฒนามาจากการจัดกลุ่มข้อมูลแบบดั้งเดิมที่จัดจำแนกกลุ่มออกเป็นกลุ่มย่อยๆ ที่แต่ละกลุ่มแบ่งแยกกันอย่างชัดเจน มาเป็นกลุ่มข้อมูลที่มีบางส่วนร่วมกัน ข้อมูลมีโอกาสถูกจัดเข้ากลุ่มได้มากกว่าหนึ่งกลุ่ม โดยการทำงานจะมีการกำหนดให้ข้อมูลแต่ละตัวมีค่าความเป็นสมาชิกกับตัวแทนของกลุ่มแต่ละกลุ่ม แต่ข้อมูลจะถูกจัดเข้ากลุ่มที่มีค่าความเป็นสมาชิกกับตัวแทนของกลุ่มสูงสุด

การทำงานประกอบด้วย การกำหนดจำนวนกลุ่มเริ่มต้นในชุดข้อมูล กำหนดจำนวนระดับความเป็นสมาชิกในกลุ่ม โดยข้อมูลทุกตัวจะถูกนำมาคำนวณค่าความเป็นสมาชิก จากนั้นเป็นการหาตัวแทนกลุ่มแต่ละกลุ่มที่มีลักษณะสำคัญของกลุ่มครบถ้วน ทำการจัดข้อมูลเข้ากลุ่ม ปรับค่าความเป็นสมาชิกของสมาชิกทุกตัวในแต่ละกลุ่ม ทำการปรับค่าตัวแทน จนกระทั่งตัวแทนปรับค่าเข้าสู่ศูนย์กลางของกลุ่มได้และค่าดัชนีที่วัดความผิดพลาดของการจัดกลุ่มมีค่าน้อยลงจนถึงจุดที่แสดงว่าการจัดกลุ่มข้อมูลน่าจะถูกต้อง ขั้นตอนวิธีการจัดกลุ่มข้อมูลแบบฟัซซี่ซีมีน (Albayrak and Amasyali, 2003) มีขั้นตอนดังนี้

1) กำหนดจำนวนกลุ่มที่ต้องการจัดเป็น C กลุ่ม เลือกตัวแทนของกลุ่มเริ่มต้น โดยการสุ่ม และกำหนดค่าระดับความเป็นสมาชิก (m) ในกลุ่ม

2) คำนวณความเป็นสมาชิก U ของข้อมูลทุกตัว โดยใช้สมการ

$$U_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{(m-1)}}}$$

เมื่อ d_{ij} คือ การคำนวณระยะห่างระหว่างตัวแทนของกลุ่มที่ i กับ ข้อมูล j ตามสูตรของ Euclidean Distance

$$d_{ij} = \left\| c_i - x_j \right\|$$

โดยที่ c_i คือ ตัวแทนของกลุ่มที่ i

x_j คือ ข้อมูลตัวที่ j

m คือ ค่าความเป็นสมาชิก (fuzziness index) โดยที่ $m \in (1, \infty)$

3) คำนวณค่าดัชนีวัดความผิดพลาด objective function และจัดข้อมูลเข้ากลุ่ม

$$J(U, c_1, c_2, ...c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j=1}^{n} U_{ij}{}^{m} \left\| c_i - x_j \right\|^2$$

โดยที่ c_i คือ ตัวแทนของกลุ่มที่ i

x_j คือ ข้อมูลตัวที่ j

m คือ ค่าความเป็นสมาชิก (fuzziness index) โดยที่ $m \in (1, \infty)$

U_{ij} คือ ค่าความเป็นสมาชิกของข้อมูลตัวที่ j ของกลุ่มที่ i

4) คำนวณตัวแทนของกลุ่มใหม่โดยใช้สูตร และปรับค่าความเป็นสมาชิกของสมาชิกทุกตัวในกลุ่ม

$$C_i = \frac{\sum_{i=1}^{n} U_{ij}{}^{m} x_i}{\sum_{j=1}^{n} U_{ij}{}^{m}}$$

โดยที่ C_i คือ ตัวแทนของกลุ่มที่ i

x_j คือ ข้อมูลตัวที่ j

m คือ ค่าความเป็นสมาชิก (fuzziness index) โดยที่ $m \in (1, \infty)$

U_{ij} คือ ค่าความเป็นสมาชิกของข้อมูลตัวที่ j ของกลุ่มที่ i

5) ทำซ้ำตั้งแต่ขั้นตอนที่ 2 จนกระทั่งตัวแทนปรับค่าเข้าสู่ศูนย์กลางของกลุ่มได้ และค่าดัชนีที่วัดความผิดพลาดของการจัดกลุ่มมีค่าเป็นที่น่าพอใจ

ข้อเด่นของอัลกอริทึมการจัดกลุ่มแบบฟัซซี่ซีมีน

1) สามารถจัดจำแนกกลุ่มข้อมูลที่มีความคล้ายกันมากได้

2) การจัดกลุ่มมีความน่าเชื่อถือเพราะจัดข้อมูลเข้ากลุ่มตามระดับความเป็นสมาชิกสูงสุดกับตัวแทนกลุ่ม

3) ผลลัพธ์ของการจัดกลุ่มที่ได้ไม่มีปัญหาแบบ local minimum

ข้อด้อยของอัลกอริทึมการจัดกลุ่มแบบฟัซซี่ซีมีน

1) ความซับซ้อนของการประมวลผลสูงและใช้เนื้อที่หน่วยความจำมาก

2) การกำหนดจำนวนกลุ่มและตัวแทนกลุ่มเริ่มต้น มีผลต่อประสิทธิภาพการจัดกลุ่มทั้งในเชิงเวลาและความถูกต้อง

ตัวอย่างที่ 2.3 การจัดกลุ่มโดยอัลกอริทึมการจัดกลุ่มแบบฟัซซี่ซีมีน แสดงดังนี้

สมมติมีข้อมูลอยู่ 4 รายการ คือ A, B, C และ D โดยแต่ละชุดข้อมูลมีตัวแปร 2 ตัวแปร คือ X_1 และ X_2 กำหนดให้มีการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม (k = 2) ชุดข้อมูลมีรายละเอียดดังแสดงในตารางที่ 2.3.1

ตารางที่ 2.3.1 ข้อมูลตัวอย่างที่ 2.3 : A, B, C และ D และมีตัวแปรจำนวน 2 ตัวแปร

รายการข้อมูล	ตัวแปร	
	X_1	X_2
A	-5	3
B	1	-1
C	-1	2
D	3	2

1) เนื่องจากต้องการจัดกลุ่มข้อมูลเป็น 2 กลุ่ม จึงสุ่มเลือกตัวแทนจุดข้อมูลมา 2 จุด เพื่อใช้เป็นตัวแทนของแต่ละกลุ่ม โดยตัวแทนที่สุ่มเลือกมาแสดงดังในตารางที่ 2.3.2

ตารางที่ 2.3.2 จุดตัวแทนเริ่มต้นของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 2.3

กลุ่ม	ตัวแทนกลุ่ม	
	X_1	X_2
1	-2.5	2
2	1.7	1.3

ในที่นี้จะใช้ C1 แทน จุดตัวแทนกลุ่มที่ 1 และ C2 แทน จุดตัวแทนกลุ่มที่ 2

2) กำหนดค่าความเป็นสมาชิก (fuzziness index)m=2 และคำนวณความเป็นสมาชิก U ของข้อมูลทุกตัวกับตัวแทนกลุ่มที่สุ่มได้ในขั้นตอนที่ 1 ในที่นี้ใช้ค่าความเป็นสมาชิก m=2 และจำนวนกลุ่มที่ต้องการ คือ k=2 ดังนั้นสมการที่ใช้เป็นดังนี้

$$U_{ij} = \frac{1}{\sum_{k=1}^{2}\left(\frac{d_{ij}}{d_{kj}}\right)^2}$$

คำนวณความเป็นสมาชิกของข้อมูลแต่ละตัวกับตัวแทนกลุ่มที่สุ่มได้จะเลือกให้ข้อมูลที่มีค่าความเป็นสมาชิกกับตัวแทนกลุ่มที่มากที่สุดไปอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มนั้นๆ

$$U_{1A} = \frac{1}{\left(\frac{d_{1A}}{d_{1A}}\right)^2 + \left(\frac{d_{1A}}{d_{2A}}\right)^2} = \frac{1}{\left(\frac{\sqrt{7.25}}{\sqrt{7.25}}\right)^2 + \left(\frac{\sqrt{7.25}}{\sqrt{47.78}}\right)^2} = 0.87$$

$$U_{2A} = \frac{1}{\left(\frac{d_{2A}}{d_{1A}}\right)^2 + \left(\frac{d_{2A}}{d_{2A}}\right)^2} = \frac{1}{\left(\frac{\sqrt{47.78}}{\sqrt{7.25}}\right)^2 + \left(\frac{\sqrt{47.78}}{\sqrt{47.78}}\right)^2} = 0.13$$

โดย d_{1A} คือ ระยะห่างแบบยุคลิดระหว่างจุดข้อมูล A กับตัวแทนกลุ่มที่ 1 (C1)

d_{2A} คือ ระยะห่างแบบยุคลิดระหว่างจุดข้อมูล A กับตัวแทนกลุ่มที่ 2 (C2)

$$d_{1A} = \sqrt{(-2.5-(-5))^2 + (2-3)^2} = \sqrt{7.25}$$

$$d_{2A} = \sqrt{(1.7-(-5))^2 + (1.3-3)^2} = \sqrt{47.78}$$

จะเห็นว่า ค่าความเป็นสมาชิกระหว่างจุดข้อมูล A กับ ตัวแทนกลุ่มที่ 1 (U_{1A}) มีค่ามากกว่าค่าความเป็นสมาชิกระหว่างจุดข้อมูล A กับ ตัวแทนกลุ่มที่ 2 (U_{2A}) ดังนั้น A จึงควรอยู่กลุ่มเดียวกับตัวแทนกลุ่มที่ 1 (C1)

3) คำนวณค่าความเป็นสมาชิกของตัวแทนกลุ่มกับข้อมูลที่เหลือจนครบทั้งหมด

$$U_{1B} = \frac{1}{\left(\frac{d_{1B}}{d_{1B}}\right)^2 + \left(\frac{d_{1B}}{d_{2B}}\right)^2} = \frac{1}{\left(\frac{\sqrt{21.25}}{\sqrt{21.25}}\right)^2 + \left(\frac{\sqrt{21.25}}{\sqrt{5.78}}\right)^2} = 0.2$$

$$U_{2B} = \frac{1}{\left(\frac{d_{2B}}{d_{1B}}\right)^2 + \left(\frac{d_{2B}}{d_{2B}}\right)^2} = \frac{1}{\left(\frac{\sqrt{5.78}}{\sqrt{21.25}}\right)^2 + \left(\frac{\sqrt{5.78}}{\sqrt{5.78}}\right)^2} = 0.8$$

โดย d_{1B} คือ ระยะห่างแบบยุคลิดระหว่างจุดข้อมูล B กับตัวแทนกลุ่มที่ 1 (C1)
d_{2B} คือ ระยะห่างแบบยุคลิดระหว่างจุดข้อมูล B กับตัวแทนกลุ่มที่ 2 (C2)

$$d_{1B} = \sqrt{(-2.5-1)^2 + (2-(-1))^2} = \sqrt{21.25}$$

$$d_{2B} = \sqrt{(1.7-1)^2 + (1.3-(-1))^2} = \sqrt{5.78}$$

$$U_{1C} = \frac{1}{\left(\frac{d_{1C}}{d_{1C}}\right)^2 + \left(\frac{d_{1C}}{d_{2C}}\right)^2} = \frac{1}{\left(\frac{\sqrt{2.25}}{\sqrt{2.25}}\right)^2 + \left(\frac{\sqrt{2.25}}{\sqrt{7.78}}\right)^2} = 0.77$$

$$U_{2C} = \frac{1}{\left(\frac{d_{2C}}{d_{1C}}\right)^2 + \left(\frac{d_{2C}}{d_{2C}}\right)^2} = \frac{1}{\left(\frac{\sqrt{7.78}}{\sqrt{2.25}}\right)^2 + \left(\frac{\sqrt{7.78}}{\sqrt{7.78}}\right)^2} = 0.23$$

โดย d_{1C} คือ ระยะห่างแบบยุคลิดระหว่างจุดข้อมูล Cกับตัวแทนกลุ่มที่ 1 (C1)
d_{2C} คือ ระยะห่างแบบยุคลิดระหว่างจุดข้อมูล Cกับตัวแทนกลุ่มที่ 2 (C2)

$$d_{1C} = \sqrt{(-2.5-(-1))^2 + (2-2)^2} = \sqrt{2.25}$$

$$d_{2C} = \sqrt{(1.7-(-1))^2 + (1.3-2)^2} = \sqrt{7.78}$$

$$U_{1D} = \frac{1}{\left(\frac{d_{1D}}{d_{1D}}\right)^2 + \left(\frac{d_{1D}}{d_{2D}}\right)^2} = \frac{1}{\left(\frac{\sqrt{30.25}}{\sqrt{30.25}}\right)^2 + \left(\frac{\sqrt{30.25}}{\sqrt{2.18}}\right)^2} = 0.07$$

$$U_{2D} = \frac{1}{\left(\frac{d_{2D}}{d_{1D}}\right)^2 + \left(\frac{d_{2D}}{d_{2D}}\right)^2} = \frac{1}{\left(\frac{\sqrt{2.18}}{\sqrt{30.25}}\right)^2 + \left(\frac{\sqrt{2.18}}{\sqrt{2.18}}\right)^2} = 0.93$$

โดย d_{1D} คือ ระยะห่างแบบยุคลิดระหว่างจุดข้อมูล D กับตัวแทนกลุ่มที่ 1 (C1)

d_{2D} คือ ระยะห่างแบบยุคลิดระหว่างจุดข้อมูล D กับตัวแทนกลุ่มที่ 2 (C2)

$$d_{1D} = \sqrt{(-2.5-3)^2 + (2-2)^2} = \sqrt{30.25}$$

$$d_{2D} = \sqrt{(1.7-3)^2 + (1.3-2)^2} = \sqrt{2.18}$$

จะเห็นว่า ค่าความเป็นสมาชิกระหว่างข้อมูล B กับ ตัวแทนกลุ่มที่ 2 มีค่ามากกว่าค่าความเป็นสมาชิกระหว่าง B กับ ตัวแทนกลุ่มที่ 1 ดังนั้น B จึงควรอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่ 2 ในทำนองเดียวกันเมื่อพิจารณาค่าความเป็นสมาชิกของ D กับตัวแทนกลุ่มทั้ง 2 กลุ่ม พบว่า D ควรอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่ 2 เช่นกัน นั่นคือ B และ D จะอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่ 2 ในขณะที่ ค่าความเป็นสมาชิกระหว่างข้อมูล C กับ ตัวแทนกลุ่มที่ 1 มีค่ามากกว่าค่าความเป็นสมาชิกระหว่าง C กับ ตัวแทนกลุ่มที่ 2 ดังนั้น C จึงควรอยู่ในกลุ่มเดียวกับตัวแทนกลุ่มที่ 1 นั่นคือ A และ C จะอยู่กลุ่มเดียวกับตัวแทนกลุ่มที่ 1

4) เมื่อข้อมูลทุกตัวได้รับการจัดกลุ่มแล้ว จะทำการปรับเปลี่ยนตัวแทนของแต่ละกลุ่ม โดยใช้สูตรดังต่อไปนี้

$$C_i = \frac{\sum_{i=1}^{n} U_{ij}{}^m x_i}{\sum_{j=1}^{n} U_{ij}{}^m}$$

ในที่นี้ m=2 และ n=4 ดังนั้นตัวแทนกลุ่มจะถูกปรับด้วยสูตร

$$C_i = \frac{\sum_{i=1}^{4} U_{ij}^{\ 2} x_i}{\sum_{j=1}^{4} U_{ij}^{\ 2}} = \frac{U_{iA}^{\ 2} x_i + U_{iB}^{\ 2} x_i + U_{iC}^{\ 2} x_i + U_{iD}^{\ 2} x_i}{U_{iA}^{\ 2} + U_{iB}^{\ 2} + U_{iC}^{\ 2} + U_{iD}^{\ 2}}; i = 1, 2$$

ดังนั้นตัวแทนกลุ่มใหม่ที่ได้รับการปรับค่าแล้ว แสดงดังตารางที่ 2.3.3

ตารางที่ 2.3.3 จุดตัวแทนที่ได้รับการปรับค่าแล้วของข้อมูลแต่ละกลุ่มในตัวอย่างที่ 2.3

กลุ่ม	ตัวแทนกลุ่ม	
	X_1	X_2
1	-3	2.5
2	2	0.8

5) นำตัวแทนกลุ่มที่ได้รับการปรับค่าแล้วจากขั้นตอนที่ 4 ทำซ้ำขั้นตอนที่ 2-4 จนกว่าข้อมูลจะไม่มีการเปลี่ยนแปลงกลุ่ม โดยสุดท้ายหลังจากคำนวณด้วยโปรแกรม R100 รอบแล้วจะได้ผลลัพธ์ดังแสดงในตารางที่ 2.3.4

ตารางที่ 2.3.4 ค่าความเป็นสมาชิกของข้อมูลแต่ละตัวกับตัวแทนกลุ่ม

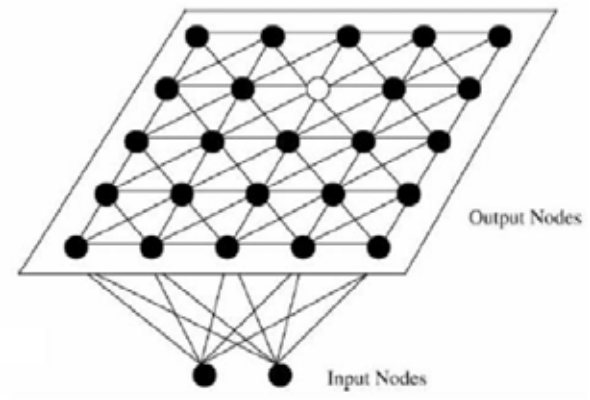
กลุ่ม	ค่าความเป็นสมาชิก			
	A	B	C	D
1	1	0.07	0.37	0.06
2	0	0.93	0.63	0.94

จากตารางที่ 2.3.4 พบว่าไม่มีการเปลี่ยนแปลงกลุ่มแล้วจึงหยุดดำเนินการ และสรุปได้ว่าสามารถจัดกลุ่มข้อมูลออกเป็น 2 กลุ่ม ดังนี้ กลุ่มที่ 1 คือ A และกลุ่มที่ 2 คือ (BCD)

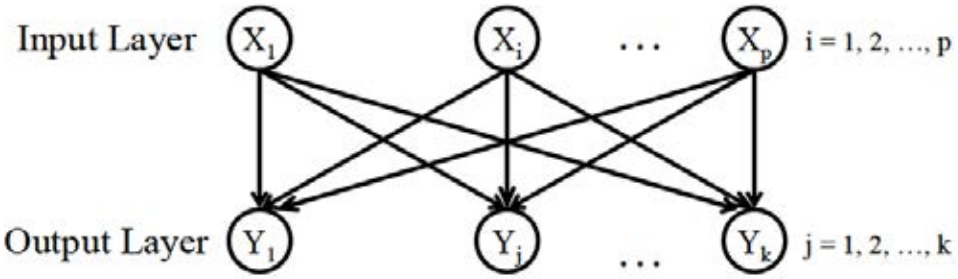
2.4 ระบบโครงข่ายประสาทเทียม (Neural Networks Algorithm)

ระบบโครงข่ายประสาทเทียม (Neural Networks Algorithm) เป็นเทคนิคที่นิยมใช้ในสาขาปัญญาประดิษฐ์ (Artificial Intelligent) ซึ่งมีความสามารถในการเรียนรู้ที่คล้ายคลึงกับระบบประสาทของมนุษย์ เช่น ในด้านการพยากรณ์ ขั้นตอนของการนำโครงข่ายประสาทเทียมมาใช้สำหรับการพยากรณ์ ก็มีลักษณะเช่นเดียวกับวิธีการพยากรณ์อื่นๆ นั่นคือ อาศัยข้อมูลนำเข้าเพื่อสร้างแบบจำลองในการพยากรณ์ข้อมูลในอนาคต และมีความสามารถในการรวม การวิเคราะห์พื้นฐานและเทคนิคเพื่อสร้างแบบจำลอง โดยที่โครงข่ายประสาทเทียมจะพยายามลดจำนวนของการทำนายที่ผิดพลาดให้ต่ำที่สุดซึ่งเป็นเหตุผลหลักที่มีการนำมาใช้ในการทำนายข้อมูลทางธุรกิจ (Wilson, 1994)

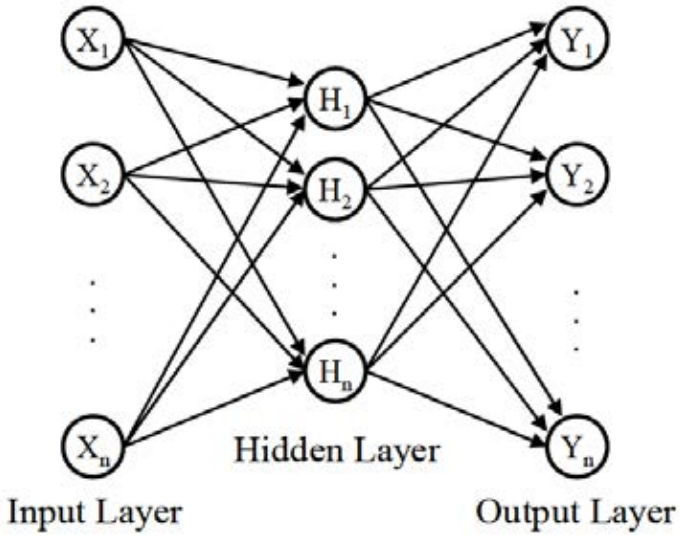
อัลกอริทึมของโครงข่ายประสาทเทียม ที่นิยมใช้ในการจัดกลุ่ม (Clustering) คือ Kohonen' s Self-Organizing Maps (SOM) (Kohonen, 1988) ซึ่งเป็น Neural Networks Algorithmที่มีลักษณะการเรียนรู้แบบเครือข่าย 2 ชั้น คือ ชั้นนำเข้าข้อมูล (Input Layer) และ ชั้นผลลัพธ์ (Output Layer) โดยจะจัดข้อมูลนำเข้าหลายมิติ ให้อยู่ในรูปแบบของข้อมูลสองมิติ ดังภาพต่อไปนี้



ภาพที่ 2.8 การทำงานของระบบโครงข่ายประสาทเทียมแบบ SOM
ที่มา : Jain and Dubes (1988)



ภาพที่ 2.9 การทำงานของระบบโครงข่ายประสาทเทียมแบบ SOM
ที่มา : ปิยธิดา, 2552



ภาพที่ 2.10 ระบบโครงข่ายประสาทเทียมแบบมีชั้นซ่อน
ที่มา : ปิยธิดา, 2552

แนวคิดของ Kohonen' s Self-Organizing Maps (SOM) คือ การทำซ้ำของข้อมูล (Jain and Dubes, 1988) เพื่อหาค่าของน้ำหนักของข้อมูลที่มีอยู่ทั้งหมดตามจำนวนกลุ่มที่ต้องการ มีวิธีการคิดดังนี้

1) สุ่มค่าน้ำหนักเริ่มต้น (W)ให้กับข้อมูล และกำหนดค่าผิดพลาดที่ยอมรับได้ (α) โดยให้ค่า α มีค่าอยู่ในช่วง 0 ถึง 1

2) นำค่าน้ำหนักของข้อมูลที่สุ่มได้ มาคำนวณค่าระยะห่างระหว่างข้อมูลกับค่าน้ำหนักด้วยสูตร Square Euclidean Distance คือ

$$D(j) = \sum_{i=1}^{p} \left(w_{ij} - x_i \right)^2 ; j = 1, 2, ..., k$$

โดยที่ w_{ij} คือ น้ำหนักของข้อมูลกลุ่มที่ j ตัวแปรที่ i

x_i คือ ค่าของข้อมูลในตัวแปรที่ i

$D(j)$ คือ ระยะทางระหว่างข้อมูลกับค่าน้ำหนัก

p คือ จำนวนตัวแปร

3) เปรียบเทียบกับ $D(j)$ ของแต่ละกลุ่มข้อมูลว่าค่าของกลุ่มไหนมีค่าน้อยที่สุด

4) นำข้อมูลของกลุ่มที่มีค่าระยะห่างน้อยที่สุดมาคำนวณหาน้ำหนักใหม่ โดยใช้สูตรดังนี้

$$w_{ij}(new) = w_{ij}(old) + \alpha(x_i - w_{ij}(old))$$

โดยที่ x_i คือ ค่าของข้อมูลในตัวแปรที่ i

$w_{ij}(new)$ คือ ค่าน้ำหนักใหม่ของข้อมูลกลุ่มที่ j ตัวแปรที่ i

$w_{ij}(old)$ คือ ค่าน้ำหนักเก่าของข้อมูลกลุ่มที่ j ตัวแปรที่ i

α คือ ค่าความผิดพลาดที่ยอมรับได้มีค่าอยู่ในช่วง 0 ถึง 1

5) ทำการลดค่า α ลง

6) ทำซ้ำตั้งแต่ขั้นตอนที่ 2 ใหม่จนกว่าจะไม่มีการเปลี่ยนแปลงค่าน้ำหนัก

ค่าน้ำหนักของข้อมูล (W) สามารถเขียนในรูปเมทริกซ์ได้ดังนี้

$$W = \begin{array}{c} \begin{matrix} W_1 & W_2 & \cdots & W_k \end{matrix} \\ \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & \cdots & w_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ w_{p1} & w_{p2} & \cdots & w_{pk} \end{bmatrix}_{p \times k} \end{array}$$

ข้อเด่นของเทคนิคการจัดกลุ่มแบบ SOM

1) สามารถจัดการกับข้อมูลจำนวนมากได้

2) มีการคำนึงถึงค่าน้ำหนักของตัวแปรที่ใช้จัดกลุ่ม

ข้อด้อยของเทคนิคการจัดกลุ่มแบบ SOM

1) ใช้เวลาในการคำนวณสูง

2) การใช้ค่าน้ำหนักที่ต่างกันก็จะให้ผลแตกต่างกัน

ตัวอย่างที่ 2.4 การจัดกลุ่มข้อมูลโดยใช้วิธีจัดระบบด้วยตัวเองของโคโฮเนน แสดงดังนี้

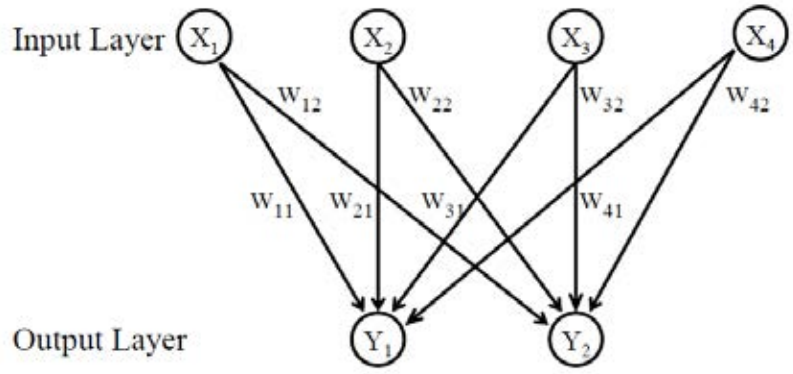
สมมติมีข้อมูลอยู่ 4 รายการ คือ A, B, C และ D โดยแต่ละชุดข้อมูลมี 4 ตัวแปร คือ X_1, X_2, X_3 และ X_4 กำหนดให้มีการแบ่งข้อมูลเป็น 2 กลุ่ม (k = 2)

ตารางที่ 2.4.1 ข้อมูลตัวอย่างที่ 2.4 : ประกอบด้วยตัวแปร X_1, X_2, X_3 และ X_4

รายการข้อมูล	ตัวแปร			
	X_1	X_2	X_3	X_4
A	1	1	0	0
B	0	0	0	1
C	1	0	0	0
D	0	0	1	1

ทำการลดค่า α ลงในทุกๆรอบ รอบละ ½ ของค่า α เดิม โดยกำหนดค่า α ในรอบแรกเท่ากับ 0.6 คือ $\alpha(0) = 0.5$

จากค่าที่กำหนดดังตัวอย่าง สามารถเขียนในรูปของโครงข่ายประสาทเทียมที่เป็น Input และ Output ได้ดังต่อไปนี้



ภาพที่ 2.11 Input และ Output ของโครงข่ายประสาทเทียม
ที่มา : ปิยธิดา, 2552

1) สมมติสุ่มน้ำหนัก (w) ให้แก่ข้อมูลที่จะทำการจัดกลุ่ม เมื่อข้อมูลตัวอย่างมี 4 ตัวแปร และจัดข้อมูลออกเป็น 2 กลุ่ม ได้ดังเมทริกซ์ต่อไปนี้

$$\begin{matrix} W_1 & W_2 \end{matrix}$$
$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \\ 0.5 & 0.7 \\ 0.9 & 0.3 \end{bmatrix}$$

ค่าน้ำหนักที่สุ่มมาได้สามารถเขียนในรูปเวกเตอร์ W_1 และ W_2 ได้ดังนี้

$$W_1 = (w_{11}, w_{21}, w_{31}, w_{41}) = (0.2, 0.6, 0.5, 0.9)$$

$$W_2 = (w_{21}, w_{22}, w_{32}, w_{42}) = (0.8, 0.4, 0.7, 0.3)$$

2) ทำการคำนวณระยะห่างระหว่างข้อมูลกับน้ำหนัก (D_j) ตามสมการ ดังต่อไปนี้

$$D_{ij} = (w_{1j} - x_1)^2 + (w_{2j} - x_2)^2 + (w_{3j} - x_3)^2 + (w_{4j} - x_4)^2$$

นำข้อมูล A จากตารางที่ คือ $(x_1, x_2, x_3, x_4) = (1,1,0,0)$ มาทำการคำนวณค่าระยะห่างยุคลิดกำลังสอง (D_j) ได้ดังนี้

$$D(W_1, A) = (0.2-1)^2 + (0.6-1)^2 + (0.5-0)^2 + (0.9-0)^2 = 1.86$$

$$D(W_2, A) = (0.8-1)^2 + (0.4-1)^2 + (0.7-0)^2 + (0.3-0)^2 = 0.98$$

เมื่อได้ค่า $D(W_1, A)$ และ $D(W_2, A)$ แล้วเลือกระยะห่างที่มีค่าน้อยที่สุด เพื่อคำนวณหาค่าน้ำหนักใหม่ จากผลที่ได้แสดงว่าควรเลือก W_2 มาคำนวณน้ำหนักใหม่ โดยใช้สมการ คือ

$$w_{ij}(new) = w_{ij}(old) + \alpha(x_i - w_{ij}(old))$$

คำนวณค่าได้ดังต่อไปนี้

$$w_{12} = 0.8 + 0.5(1-0.8) = 0.9$$

$$w_{22} = 0.4 + 0.5(1-0.4) = 0.7$$

$$w_{32} = 0.7 + 0.5(0-0.7) = 0.35$$

$$w_{42} = 0.3 + 0.5(0-0.3) = 0.15$$

สามารถเขียนน้ำหนักใหม่ในรูปเมทริกซ์ได้ดังต่อไปนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.9 \\ 0.6 & 0.7 \\ 0.5 & 0.35 \\ 0.9 & 0.15 \end{bmatrix}$$

นำข้อมูล B จากตารางที่ คือ $(x_1, x_2, x_3, x_4) = (0,0,0,1)$ มาคำนวณซ้ำ ด้วยวิธีเดียวกับข้อมูล A ซึ่งจะได้ผลดังต่อไปนี้

$D(W_1, B) = 0.66$ และ $D(W_2, B) = 2.145$

เลือก W_1 เนื่องจาก $D(W_1, B)$ มีค่าน้อยกว่า $D(W_2, B)$ เพื่อทำการคำนวณหาค่าน้ำหนักใหม่ ซึ่งจะได้เมทริกซ์ต่อไปนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.1 & 0.9 \\ 0.3 & 0.7 \\ 0.25 & 0.35 \\ 0.95 & 0.15 \end{bmatrix}$$

นำข้อมูล C จากตารางที่ คือ $(x_1, x_2, x_3, x_4) = (1,0,0,0)$ มาทำการคำนวณซ้ำ ด้วยวิธีเดียวกับข้อมูล A ซึ่งจะได้ผลดังต่อไปนี้

$D(W_1, C) = 1.865$ และ $D(W_2, C) = 0.645$

เลือก W_2 เนื่องจาก $D(W_2, C)$ มีค่าน้อยกว่า $D(W_1, C)$ เพื่อทำการคำนวณหาค่าน้ำหนักใหม่ ซึ่งจะได้เมทริกซ์ต่อไปนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.1 & 0.95 \\ 0.3 & 0.35 \\ 0.25 & 0.175 \\ 0.95 & 0.075 \end{bmatrix}$$

นำข้อมูล D จากตารางที่ คือ $(x_1, x_2, x_3, x_4) = (0,0,1,1)$ มาทำการคำนวณซ้ำ ด้วยวิธีเดียวกับข้อมูล A ซึ่งจะได้ผลดังต่อไปนี้

$D(W_1, D) = 0.665$ และ $D(W_2, D) = 2.56$

เลือก W_1 เนื่องจาก $D(W_1, D)$ มีค่าน้อยกว่า $D(W_2, D)$ เพื่อทำการคำนวณหาค่าน้ำหนักใหม่ ซึ่งจะได้เมทริกซ์ต่อไปนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.05 & 0.95 \\ 0.15 & 0.35 \\ 0.625 & 0.175 \\ 0.975 & 0.075 \end{bmatrix}$$

3) ปรับลดค่า α ลงครึ่งหนึ่งดังต่อไปนี้

$$\alpha(1) = \alpha(0) / 2 = 0.5 / 2 = 0.25$$

4) หลังจากลดค่า α แล้วทำการคำนวณซ้ำขั้นตอนที่ 2 และ 3 จนกระทั่งค่าน้ำหนัก (w) ไม่มีการเปลี่ยนแปลง ซึ่งหลังจากคำนวณครบ 100 รอบ ได้ผลดังนี้

$$\begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \end{bmatrix} = \begin{bmatrix} 0.25 & 0.77 \\ 0.19 & 0.35 \\ 0.43 & 0.17 \\ 0.76 & 0.24 \end{bmatrix}$$

กลุ่มที่ 1 กลุ่มที่ 2

จากเมทริกซ์น้ำหนัก (w) ที่ได้นี้ จะเห็นว่าสามารถจัดข้อมูลได้เป็น 2 กลุ่มอย่างชัดเจน โดยนำค่าน้ำหนักที่ได้มาหาค่าระยะห่างกับข้อมูลเพื่อทำการจัดกลุ่ม ซึ่งแสดงตัวอย่างการคำนวณค่าระยะห่างระหว่างน้ำหนักกับข้อมูล A คือ $(x_1, x_2, x_3, x_4) = (1,1,0,0)$ ได้ดังต่อไปนี้

$$D_{A1} = (0.25-1)^2 + (0.19-1)^2 + (0.43-0)^2 + (0.76-0)^2 = 1.98$$

$$D_{A2} = (0.77-1)^2 + (0.35-1)^2 + (0.17-0)^2 + (0.24-0)^2 = 0.56$$

เนื่องจากค่า D_{A2} มีค่าน้อยกว่า D_{A1} ดังนั้นจึงจัดข้อมูล A อยู่ในกลุ่มที่ 2 และคำนวณค่าระยะห่าง (D_j) เพื่อจัดกลุ่มให้กับข้อมูล B, C และ D ซึ่งผลที่ได้แสดงในตารางที่ 2.4.2

ตารางที่ 2.4.2 ตัวอย่างกลุ่มข้อมูลที่ได้จากการจัดกลุ่มโดยใช้ SOM อัลกอริทึม

ข้อมูล	X_1	X_2	X_3	X_4	กลุ่ม
A	1	1	0	0	2
B	0	0	0	1	1
C	1	0	0	0	2
D	0	0	1	1	1

จากข้อมูลในตัวอย่างที่ 2.4 สรุปได้ว่าสามารถจัดกลุ่มข้อมูลออกเป็น 2 กลุ่มดังนี้ กลุ่มที่ 1 คือ (AC) และ กลุ่มที่ 2 คือ (BD)

2.5 การวัดประสิทธิภาพการจัดกลุ่มข้อมูล

2.5.1 การวัดค่าความแตกต่างของข้อมูลภายในกลุ่ม (Root Mean Square Standard Deviation: RMSSTD)

โดยค่าความแตกต่างของข้อมูลภายในกลุ่มนั้นแสดงให้เห็นถึงประสิทธิภาพการจัดกลุ่มว่ามีมากน้อยเพียงใด ถ้าค่าความแตกต่างภายในกลุ่มน้อย แสดงถึงการจัดกลุ่มที่ดี ข้อมูลภายในกลุ่มมีความใกล้เคียงกันมาก (Halkidi*et al.*, 2002)

สามารถคำนวณค่าความแตกต่างภายในกลุ่มได้ดังนี้

$$RMSSTD = \sqrt{\frac{\sum_{i=1}^{k}\sum_{j=1}^{p}(x_{ij} - \bar{x}_{ij})^2}{\sum_{i=1}^{k}\sum_{j=1}^{p}(n_{ij} - 1)}}$$

โดย k คือ จำนวนกลุ่มที่แบ่งได้ทั้งหมด

p คือ จำนวนตัวแปรอิสระทั้งหมดภายในชุดข้อมูล

x_{ij} คือ ข้อมูลกลุ่มที่ i ตัวแปรอิสระที่ j

$\bar{x}_{ij}$ คือ ค่าเฉลี่ยของข้อมูลกลุ่มที่ i ตัวแปรอิสระที่ j

n_{ij} คือ จำนวนข้อมูลในกลุ่มที่ i ตัวแปรอิสระที่ j

2.5.2 ค่า Wilk' s Lambda

เป็นอัตราส่วนระหว่างความแปรปรวนภายในกลุ่ม (Sum of Squares Within-Groups : SS_w) กับ ความแปรปรวนรวม (Sum of Squares Total : SS_t) เป็นค่าที่แสดงถึงประสิทธิภาพในการจัดกลุ่ม โดยถ้า Wilk' s Lambda มีค่าน้อย แสดงถึงการจัดกลุ่มที่ดี

Wilk' s Lambda = SS_w / SS_t

เมื่อ $SS_t = \sum_{i=1}^{k}\sum_{j=1}^{p}(x_{ij} - \bar{x}_j)^2$ และ $SS_w = \sum_{i=1}^{k}\sum_{j=1}^{p}(x_{ij} - \bar{x}_{ij})^2$

โดยที่ SS_t คือ ผลรวมของผลต่างกำลังสองของข้อมูลทั้งหมด

SS_w คือ ผลรวมของผลต่างกำลังสองทุกข้อมูลภายในกลุ่ม

k คือ จำนวนกลุ่มที่แบ่งได้ทั้งหมด

p คือ จำนวนตัวแปรอิสระทั้งหมด

x_{ij} คือ ข้อมูลกลุ่มที่ i ตัวแปรอิสระที่ j

$\bar{x}_j$ คือ ค่าเฉลี่ยข้อมูลในตัวแปรอิสระที่ j

$\bar{x}_{ij}$ คือ ค่าเฉลี่ยของข้อมูลกลุ่มที่ i ตัวแปรอิสระที่ j

n_j คือ จำนวนข้อมูลในตัวแปรอิสระที่ j

n_{ij} คือ จำนวนข้อมูลในกลุ่มที่ i ตัวแปรอิสระที่ j

2.5.3 ค่า Misclassification Rate

เป็นค่าที่บอกถึงความผิดพลาดในการจัดกลุ่ม โดยการวิจัยในครั้งนี้ผู้วิจัยได้ทำการจำลองข้อมูลในแต่ละขนาดตัวอย่างขึ้นมาเป็นจำนวน 2, 4 และ 6 กลุ่ม ดังนั้นเราจะถือว่ากลุ่มที่เราจำลองขึ้นมาเป็นกลุ่มที่แท้จริง และจะใช้อัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอนจัดกลุ่มข้อมูลออกมา กลุ่มที่ถูกจัดออกมาจากอัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอนทั้ง 4 แบบดังที่ได้กล่าวมาจะถูกนำมาตรวจสอบกับกลุ่มที่แท้จริงที่เราจำลองไว้ตั้งแต่เริ่มต้น โดยถ้าค่า Misclassification Rate มีค่าน้อย จะแสดงถึงการจัดกลุ่มที่ดี

Misclassification Rate = จำนวนข้อมูลที่ทำนายผิด / จำนวนข้อมูลทั้งหมด

บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้เป็นการเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลแบบ 2 ขั้นตอน ซึ่งได้เสนอไว้ 4 วิธี ด้วยกันคือ วิธีที่ 1 SOM ร่วมกับ K-Means วิธีที่ 2 SOM ร่วมกับ Fuzzy C-Means วิธีที่ 3 Ward ร่วมกับ K-Means และวิธีที่ 4 Ward ร่วมกับ Fuzzy C-Means โดยทำการศึกษาสำหรับข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) จำนวนตัวแปรที่ใช้ในการศึกษาครั้งนี้ คือ 2, 4 และ 6 ตัวแปร ลักษณะความสัมพันธ์ของตัวแปรที่ศึกษามี 2 แบบ คือ ตัวแปรอิสระกัน และ ตัวแปรมีความสัมพันธ์กัน ขนาดตัวอย่างที่ศึกษาคือ 25, 100 และ 300 โดยผู้วิจัยจะจำลองข้อมูลเป็นกลุ่มสำหรับแต่ละขนาดตัวอย่าง จำนวนกลุ่มที่จำลองมีขนาดเป็น 2 กลุ่ม, 4 กลุ่ม และ 6 กลุ่ม โดยทำการศึกษาทั้งกรณีที่ข้อมูลแต่ละกลุ่มไม่มีการซ้อนทับกัน และ ข้อมูลแต่ละกลุ่มมีการซ้อนทับกัน 40% ในการเปรียบเทียบประสิทธิภาพของการจัดกลุ่มข้อมูลแบบ 2 ขั้นตอน จะพิจารณาประสิทธิภาพการจัดกลุ่มด้วยค่า Wilk' s Lambda, ค่า RMSSTD และค่า Misclassification Rate สำหรับการวิจัยครั้งนี้ใช้โปรแกรม R เวอร์ชัน 3.0.1 ซึ่งมีแผนการจำลองข้อมูลและขั้นตอนในการวิจัยดังนี้

3.1 แผนการดำเนินการวิจัย

3.1.1ลักษณะการแจกแจงของข้อมูลคือ การแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) โดยจำนวนตัวแปรที่ใช้ในการศึกษาเท่ากับ 2, 4 และ 6 ตัวแปร

3.1.2 ขนาดตัวอย่างแบ่งเป็น 3 ระดับ คือ 25, 100 และ 300

3.1.3 ลักษณะความสัมพันธ์ของตัวแปรที่ศึกษามี 2 แบบ คือ ตัวแปรอิสระต่อกัน และ

ตัวแปรมีความสัมพันธ์กัน ค่าความสัมพันธ์ของตัวแปรแต่ละคู่ถูกสุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) โดยค่าที่สุ่มอยู่ในช่วง (-1,1) และ (0,1)

3.1.4 ในการศึกษาครั้งนี้จะจำลองข้อมูลเป็น 2 ลักษณะ คือ ข้อมูลแต่ละกลุ่มไม่มีการซ้อนทับกัน (Nonoverlapping Clusters) และข้อมูลแต่ละกลุ่มมีการซ้อนทับกัน 40% (40% Overlapping Cluster) (Milligan and Cooper, 1980)

กรณีที่ข้อมูลแต่ละกลุ่มไม่มีการซ้อนทับกัน (Nonoverlapping Clusters)

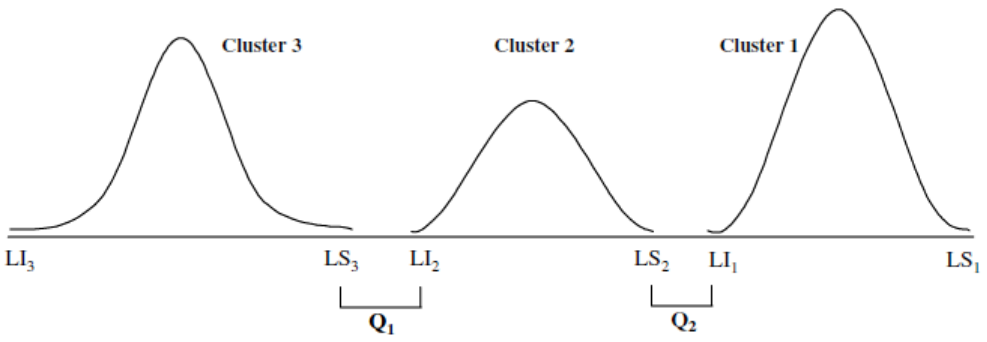
เนื่องจากในการวิจัยครั้งนี้จะจำลองข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) เป็น 2 กลุ่ม, 4 กลุ่ม และ 6 กลุ่มตามแต่ละขนาดตัวอย่าง ดังนั้นข้อมูลแต่ละกลุ่มจะประกอบด้วยเวกเตอร์ค่าเฉลี่ย (Mean Vector : μ_i) และเมทริกซ์ความแปรปรวนร่วม (Covariance Matrix : $\sum_i ; i = 1, 2, ..., k$) โดยขอบเขตของการจำลองข้อมูลในลักษณะไม่ซ้อนทับกัน จะกำหนดดังต่อไปนี้

1) ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ของแต่ละตัวแปรสุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) โดยค่าที่สุ่มอยู่ในช่วง (10,40)

2) พิสัย (Range) สำหรับแต่ละกลุ่ม กำหนดเป็น 3 เท่าของค่าส่วนเบี่ยงเบนมาตรฐาน

3) เวกเตอร์ค่าเฉลี่ย (Mean Vector) คือค่ากลางของพิสัย

4) ขอบเขตของการไม่ซ้อนทับกัน คือ 1.5 เท่าของส่วนเบี่ยงเบนมาตรฐานที่ห่างจากเวกเตอร์ค่าเฉลี่ย (Mean Vector) บวกกับค่าองศาสุ่มของการแยกกัน (Random Degree of Separation : Q_i) โดย Q_i เป็นค่าที่สุ่มจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ในช่วง (0.25, 0.75)



ภาพที่ 3.1 ข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน (Nonoverlapping Clusters)

ที่มา : S.A. Mingoti, J.O. Lima, 2006

ตัวอย่างการจำลองข้อมูลโดยที่ข้อมูลแต่ละกลุ่มไม่ซ้อนทับกัน ในกรณีนี้จะจำลองข้อมูลเป็น 2 กลุ่ม โดยกำหนดเวกเตอร์ค่าเฉลี่ยสำหรับกลุ่มที่ 1 เป็นเวกเตอร์ศูนย์ นั้นคือ $\tilde{\mu}_1 = \tilde{0}$

จะได้ว่า $\tilde{\mu}_2 = 1.5sd_1 + Q + 1.5sd_2$

กรณีที่ข้อมูลแต่ละกลุ่มมีการซ้อนทับกัน 40% (40% Overlapping Clusters)

ขอบเขตของการจำลองข้อมูลในลักษณะซ้อนทับกัน จะกำหนดดังต่อไปนี้

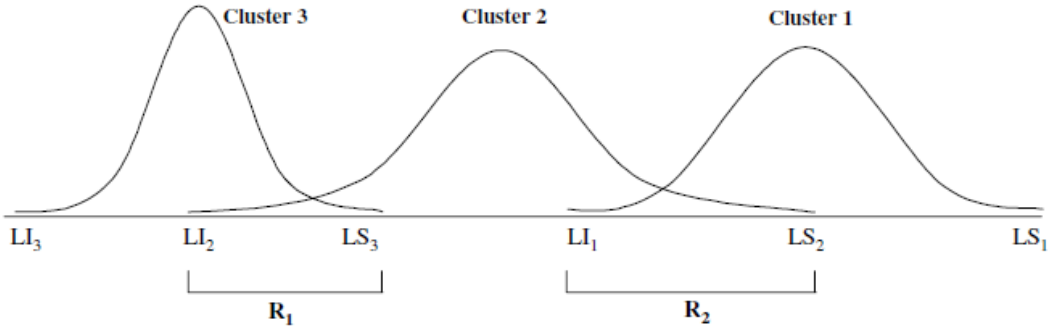
1) กำหนด LI_i และ LI_j เป็นลิมิตต่ำสุด (Lower Limits) ของกลุ่มที่ i และกลุ่มที่ j

$$LI_j = (1-m)range_i + LI_i$$

โดย m เป็นค่าของการซ้อนทับกัน (Intersection) ระหว่างกลุ่มที่ i และกลุ่มที่ j และกำหนด $m = 0.4$

2) กำหนด ความยาวช่วงของการซ้อนทับกันเป็น R_i โดย

$$R_i = m(range_i); i = 1, 2, ..., (k-1)$$



ภาพที่ 3.2 ข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน (Overlapping Clusters)

ที่มา : S.A. Mingoti, J.O. Lima, 2006

ตัวอย่างการจำลองข้อมูลโดยที่ข้อมูลแต่ละกลุ่มซ้อนทับกัน ในกรณีนี้จะจำลองข้อมูลเป็น 2 กลุ่ม โดยกำหนดเวกเตอร์ค่าเฉลี่ยสำหรับกลุ่มที่ 1 เป็นเวกเตอร์ศูนย์ นั้นคือ $\tilde{\mu}_1 = \tilde{0}$ และกำหนดค่าของการซ้อนทับกัน (Intersection) ระหว่างกลุ่มที่ 1 กับกลุ่มที่ 2 เป็น 0.4 นั่นคือ $m = 0.4$ จะได้ว่า

$$\tilde{\mu}_2 - \tilde{\mu}_1 = 1.5sd_1 - R_1 + 1.5sd_2$$

$$\tilde{\mu}_2 - \tilde{0} = 1.5sd_1 - m(range_1) + 1.5sd_2 = 1.5sd_1 - (0.4)(3sd_1) + 1.5sd_2$$

$$\tilde{\mu}_2 = 0.3sd_1 + 1.5sd_2$$

3.1.5 จำลองจำนวนกลุ่มสำหรับแต่ละขนาดตัวอย่างเป็น 2, 4 และ 6 กลุ่ม

3.1.6 เปรียบเทียบวิธีการจัดกลุ่มข้อมูลแบบ 2 ขั้นตอนจาก 4 วิธี คือ วิธีที่ 1SOM ร่วมกับ K-Means วิธีที่ 2 SOM ร่วมกับ Fuzzy C-Means วิธีที่ 3 Ward ร่วมกับ K-Means และวิธีที่ 4 Ward ร่วมกับ Fuzzy C-Means

- SOM ร่วมกับ K-Means ในขั้นตอนที่ 1 ใช้ SOM ในการกำหนดจำนวนกลุ่มเริ่มต้น เพื่อเป็นข้อมูลนำเข้าให้กับขั้นตอนที่ 2 และในขั้นตอนที่ 2 ใช้ K-Means จัดกลุ่มข้อมูล

- SOM ร่วมกับ Fuzzy C-Meansในขั้นตอนที่ 1 ใช้ SOM ในการกำหนดจำนวนกลุ่มเริ่มต้น เพื่อเป็นข้อมูลนำเข้าให้กับขั้นตอนที่ 2 และในขั้นตอนที่ 2 ใช้ Fuzzy C-Means จัดกลุ่มข้อมูล

- Ward ร่วมกับ K-Meansในขั้นตอนที่ 1 ใช้ Ward ในการกำหนดจำนวนกลุ่มเริ่มต้น เพื่อเป็นข้อมูลนำเข้าให้กับขั้นตอนที่ 2 โดยเราจะเลือกลำดับขั้น (Level) จากการพิจารณาค่าความแตกต่างของข้อมูลภายในกลุ่มนั้น (RMSSTD) และในขั้นตอนที่ 2 ใช้ K-Means จัดกลุ่มข้อมูล

- Ward ร่วมกับ Fuzzy C-Meansในขั้นตอนที่ 1 ใช้ Ward ในการกำหนดจำนวนกลุ่มเริ่มต้น เพื่อเป็นข้อมูลนำเข้าให้กับขั้นตอนที่ 2 โดยเราจะเลือกลำดับขั้น (Level) จากการพิจารณาค่าแตกต่างของข้อมูลภายในกลุ่มนั้น (RMSSTD) และในขั้นตอนที่ 2 ใช้ Fuzzy C-Means จัดกลุ่มข้อมูล

3.1.7 เกณฑ์ที่ใช้พิจารณาการจัดกลุ่มข้อมูลแบบ 2 ขั้นตอน ประกอบด้วย 3 เกณฑ์ คือ ค่า Wilk' s Lambda ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD)และค่า Misclassification Rate

3.1.8 การจำลองข้อมูลในแต่ละสถานการณ์จะกระทำซ้ำ 500 รอบ

3.2 ขั้นตอนในการดำเนินการวิจัย

3.2.1 จำลองข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) โดยกำหนดค่าเฉลี่ยและค่าความแปรปรวนร่วม

3.2.2 จำลองข้อมูลตามลักษณะความสัมพันธ์ของตัวแปร 2 แบบ คือ ตัวแปรอิสระกันและตัวแปรมีความสัมพันธ์กัน โดยจำนวนตัวแปรที่ใช้ในการศึกษา คือ 2, 4 และ 6 ตัวแปร

3.2.3 จำลองข้อมูลตามขนาดตัวอย่าง โดยในแต่ละขนาดตัวอย่างจะมีการจำลองข้อมูลเป็นกลุ่ม จำนวนกลุ่มที่ใช้ในการจำลอง คือ 2, 4และ 6 กลุ่ม

3.2.4 จำลองข้อมูลในแต่ละกลุ่มตามลักษณะที่กำหนด คือ ข้อมูลในแต่ละกลุ่มไม่มีการซ้อนทับกัน (Nonoverlapping Clusters) และ ข้อมูลในแต่ละกลุ่มมีการซ้อนทับกัน 40% (40% Overlapping Clusters)

3.2.5 ในแต่ละสถานการณ์การทดลอง จำลองข้อมูลโดยกระทำซ้ำ 500 รอบ

3.2.6 ทำการจัดกลุ่มข้อมูลที่จำลองขึ้น

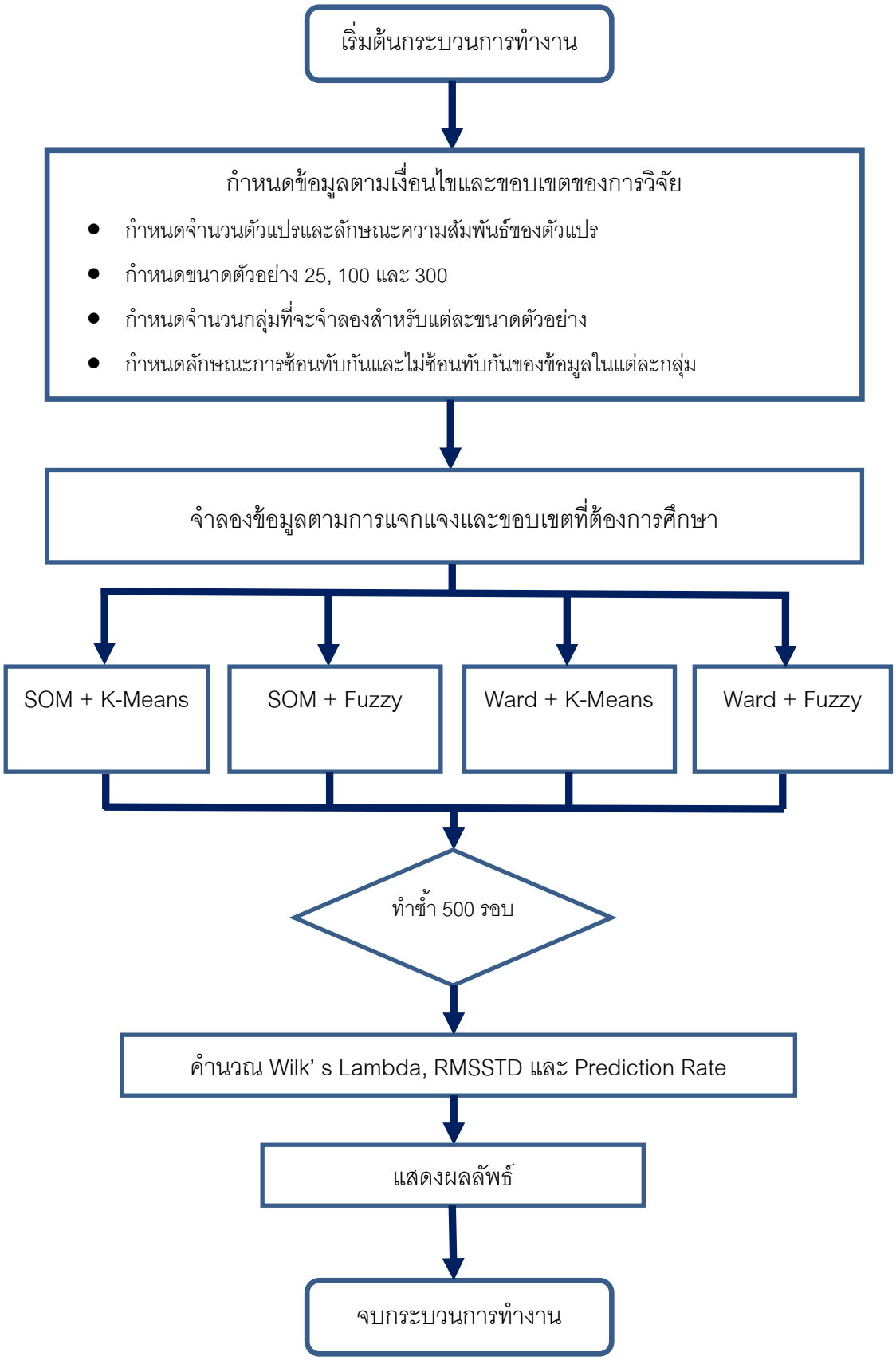
3.2.7 วัดประสิทธิภาพการจัดกลุ่มข้อมูลตามเกณฑ์ที่ได้ระบุไว้

3.2.8 เปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูล

3.2.9 สรุปผลที่ได้จากการเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูล ว่าวิธีการจัดกลุ่มแบบใดให้ผลการจัดกลุ่มดีที่สุด

3.3 ขั้นตอนการทำงานของโปรแกรม

โปรแกรมที่ใช้ในการศึกษาการวิจัยในครั้งนี้เขียนด้วยโปรแกรม R เวอร์ชัน 3.0.1 ซึ่งในแต่ละสถานการณ์ของการทดลองจะกระทำซ้ำ 500 รอบ สามารถแสดงขั้นตอนการทำงานของโปรแกรมได้ดังนี้


เริ่มต้นกระบวนการทำงาน
กำหนดข้อมูลตามเงื่อนไขและขอบเขตของการวิจัย
กำหนดจำนวนตัวแปรและลักษณะความสัมพันธ์ของตัวแปร
กำหนดขนาดตัวอย่าง 25, 100 และ 300
กำหนดจำนวนกลุ่มที่จะจำลองสำหรับแต่ละขนาดตัวอย่าง
กำหนดลักษณะการซ้อนทับกันและไม่ซ้อนทับกันของข้อมูลในแต่ละกลุ่ม
จำลองข้อมูลตามการแจกแจงและขอบเขตที่ต้องการศึกษา
SOM + K-Means
SOM + Fuzzy
Ward + K-Means
Ward + Fuzzy
ทำซ้ำ 500 รอบ
คำนวณ Wilk' s Lambda, RMSSTD และ Prediction Rate
แสดงผลลัพธ์
จบกระบวนการทำงาน

บทที่ 4

ผลการวิเคราะห์ข้อมูล

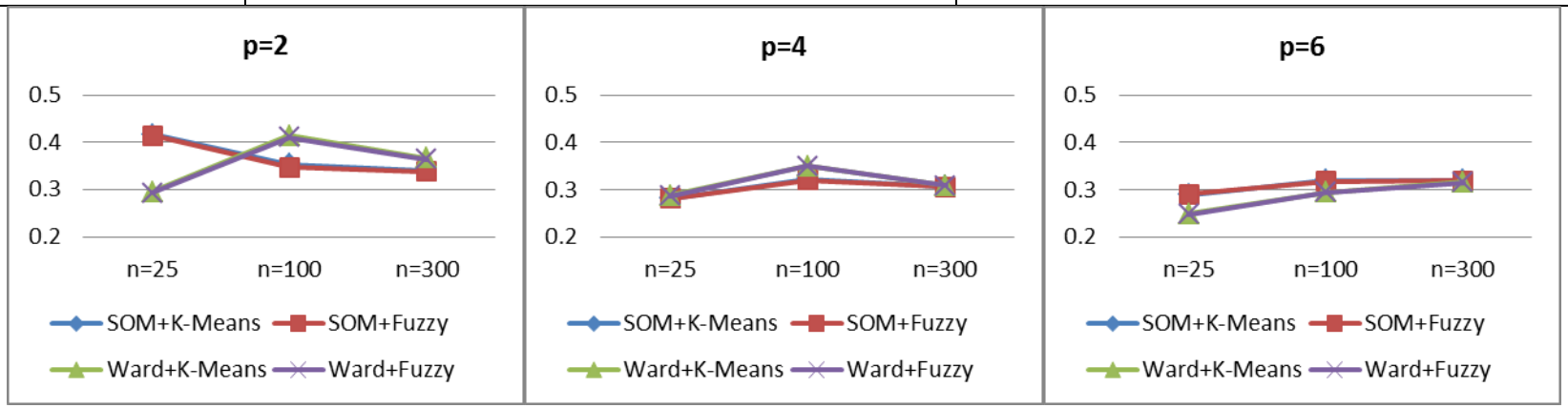
จากการศึกษาเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอน ทั้ง 4 วิธี ได้แก่ วิธีที่ 1 SOM ร่วมกับ K-Means วิธีที่ 2 SOM ร่วมกับ Fuzzy C-Means วิธีที่ 3 Ward ร่วมกับ K-Means และวิธีที่ 4 Ward ร่วมกับ Fuzzy C-Means โดยทำการศึกษาสำหรับข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) จำนวนตัวแปรที่ใช้ในการศึกษาครั้งนี้ คือ 2, 4 และ 6 ตัวแปร ลักษณะความสัมพันธ์ของตัวแปรที่ศึกษามี 2 แบบ คือ ตัวแปรอิสระกัน และ ตัวแปรมีความสัมพันธ์กัน ขนาดตัวอย่างที่ศึกษาคือ 25, 100 และ 300 โดยผู้วิจัยจะจำลองข้อมูลเป็นกลุ่มสำหรับแต่ละขนาดตัวอย่าง จำนวนกลุ่มที่จำลองมีขนาดเป็น 2 กลุ่ม, 4 กลุ่ม และ 6 กลุ่ม โดยทำการศึกษาทั้งกรณีที่ข้อมูลแต่ละกลุ่มไม่มีการซ้อนทับกัน และ ข้อมูลแต่ละกลุ่มมีการซ้อนทับกัน 40% ในการเปรียบเทียบประสิทธิภาพของการจัดกลุ่มข้อมูลแบบ 2 ขั้นตอน จะพิจารณาประสิทธิภาพการจัดกลุ่มด้วยค่า Wilk' s Lambda, ค่า RMSSTD และค่า Misclassification Rate

ในการนำเสนอผลงานวิจัยจะแสดงในรูปแบบของตารางและกราฟ โดยมีสัญลักษณ์ที่ใช้แทนความหมายต่างๆ ดังนี้

SOM+K	แทน การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ K-Means
SOM+F	แทน การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means
Ward+K	แทน การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี Ward ร่วมกับ K-Means
Ward+F	แทน การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี Ward ร่วมกับ Fuzzy C-Means
Case 1	แทน กรณีที่ตัวแปรอิสระต่อกัน
Case 2	แทน กรณีที่ตัวแปรมีความสัมพันธ์กันในช่วง (-1,1)
Case 3	แทน กรณีที่ตัวแปรมีความสัมพันธ์กันในช่วง (0,1)

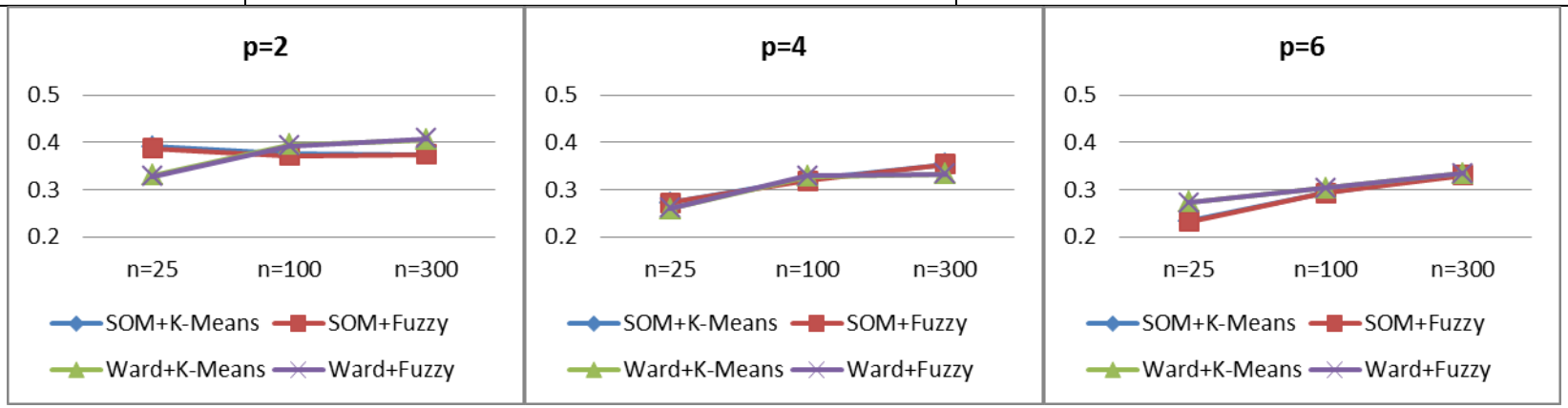
ตารางที่ 4.1 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.4257	0.4179	0.4146	0.3097	0.2952	0.2929	0.0725	0.0782	0.0736	0.0840	0.0866	0.0807
		100	0.3603	0.3534	0.3479	0.4394	0.4146	0.4115	0.0480	0.0502	0.0441	0.0532	0.0448	0.0430
		300	0.3578	0.3390	0.3379	0.3733	0.3663	0.3642	0.0215	0.0283	0.0295	0.0411	0.0316	0.0330
	4	25	0.2985	0.2814	0.2808	0.3097	0.2879	0.2871	0.0155	0.0344	0.0278	0.0322	0.0372	0.0459
		100	0.3278	0.3222	0.3196	0.3627	0.3501	0.3495	0.0200	0.0191	0.0144	0.0149	0.0180	0.0177
		300	0.3115	0.3078	0.3060	0.3151	0.3080	0.3082	0.0060	0.0126	0.0106	0.0131	0.0113	0.0111
	6	25	0.2940	0.2893	0.2899	0.2485	0.2493	0.2481	0.0119	0.0164	0.0201	0.0046	0.0201	0.0187
		100	0.3254	0.3196	0.3176	0.3033	0.2943	0.2938	0.0096	0.0081	0.0121	0.0098	0.0066	0.0079
		300	0.3219	0.3196	0.3193	0.3199	0.3158	0.3147	0.0015	0.0042	0.0034	0.0047	0.0033	0.0113



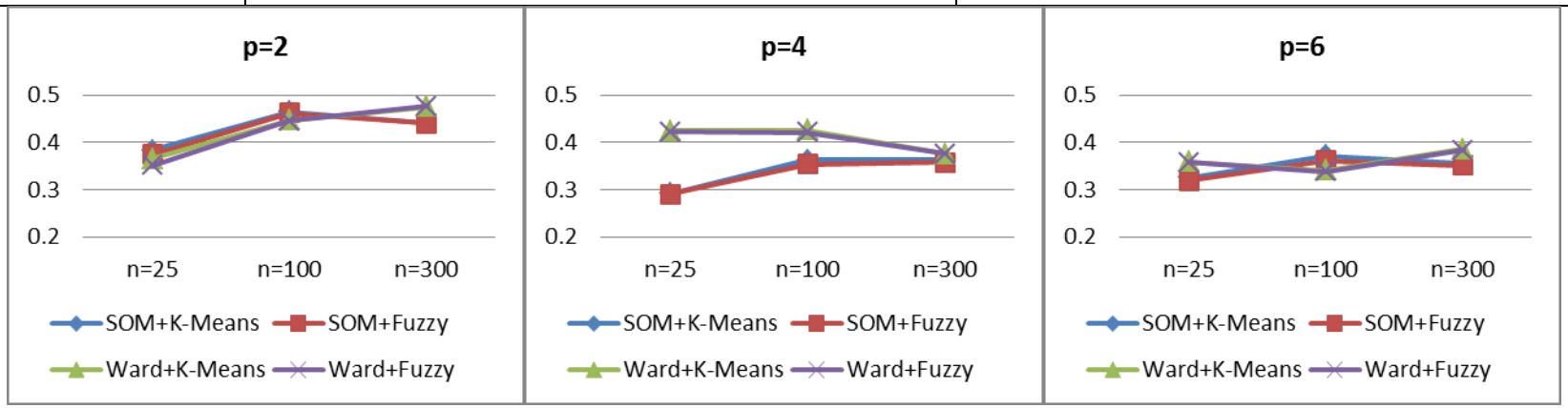
ตารางที่ 4.1 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.4120	0.3918	0.3881	0.3414	0.3304	0.3276	0.0787	0.0847	0.0864	0.0872	0.0850	0.0879
		100	0.3841	0.3769	0.3723	0.4008	0.3951	0.3931	0.0612	0.0600	0.0643	0.0730	0.0683	0.0671
		300	0.3907	0.3749	0.3741	0.4112	0.4052	0.4086	0.0565	0.0630	0.0620	0.0718	0.0703	0.0705
	4	25	0.2854	0.2726	0.2718	0.2899	0.2595	0.2607	0.0349	0.0428	0.0421	0.0409	0.0466	0.0508
		100	0.3397	0.3217	0.3187	0.3349	0.3281	0.3296	0.0354	0.0287	0.0292	0.0292	0.0359	0.0274
		300	0.3674	0.3543	0.3533	0.3451	0.3327	0.3326	0.0178	0.0240	0.0221	0.0194	0.0239	0.0263
	6	25	0.2450	0.2332	0.2319	0.2864	0.2732	0.2722	0.0294	0.0261	0.0241	0.0253	0.0336	0.0361
		100	0.3061	0.2943	0.2924	0.3154	0.3036	0.3038	0.0214	0.0149	0.0120	0.0139	0.0144	0.0132
		300	0.3454	0.3313	0.3311	0.3411	0.3349	0.3349	0.0056	0.0073	0.0077	0.0119	0.0076	0.0101



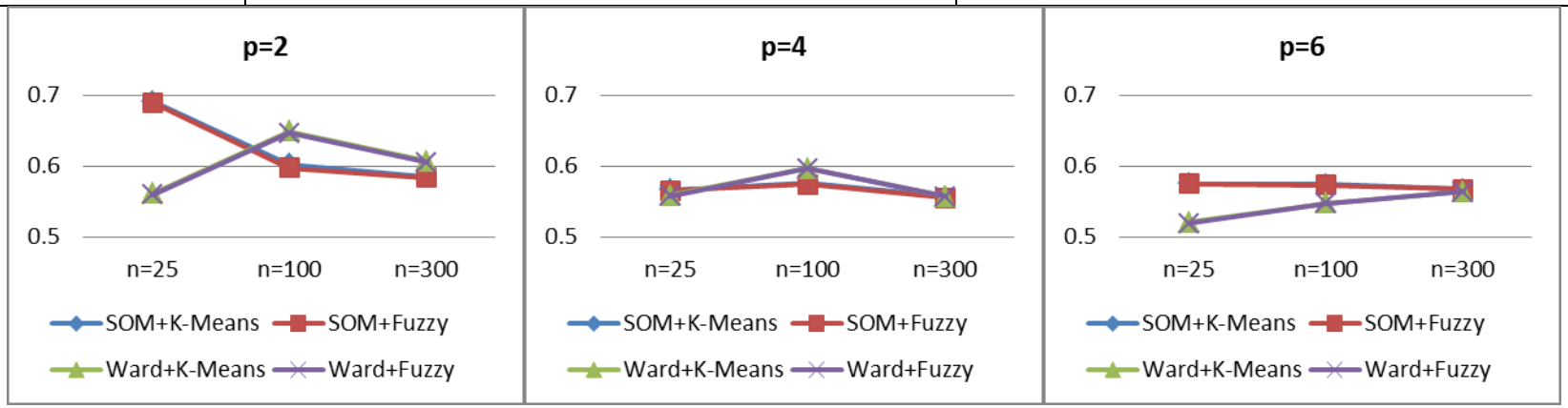
ตารางที่ 4.1 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.3956	0.3842	0.3742	0.3686	0.3655	0.3514	0.1043	0.1043	0.0969	0.1159	0.1166	0.1147
		100	0.4776	0.4651	0.4626	0.4526	0.4467	0.4454	0.0542	0.0571	0.0551	0.0653	0.0604	0.0589
		300	0.4593	0.4414	0.4402	0.4845	0.4756	0.4768	0.0418	0.0455	0.0456	0.0568	0.0503	0.0525
	4	25	0.3076	0.2914	0.2909	0.4379	0.4256	0.4230	0.0777	0.0888	0.0936	0.0898	0.0847	0.0834
		100	0.3668	0.3635	0.3533	0.4268	0.4265	0.4211	0.0467	0.0466	0.0408	0.0494	0.0436	0.0430
		300	0.3714	0.3638	0.3582	0.3885	0.3764	0.3756	0.0313	0.0397	0.0359	0.0461	0.0470	0.0466
	6	25	0.3263	0.3249	0.3195	0.3667	0.3586	0.3576	0.0615	0.0756	0.0727	0.0767	0.0730	0.0764
		100	0.3754	0.3717	0.3619	0.3582	0.3415	0.3385	0.0404	0.0404	0.0338	0.0547	0.0475	0.0472
		300	0.3624	0.3565	0.3519	0.3954	0.3868	0.3839	0.0188	0.0269	0.0247	0.0411	0.0358	0.0346



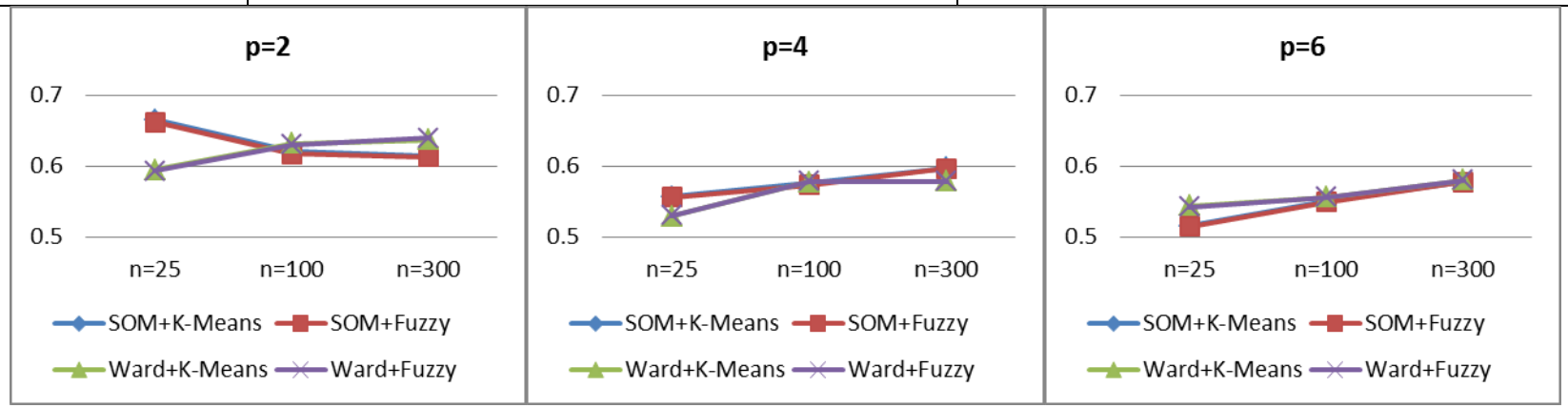
ตารางที่ 4.2 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.7014	0.6919	0.6897	0.5848	0.5613	0.5591	0.0664	0.0727	0.0684	0.0659	0.0672	0.0619
		100	0.6040	0.6024	0.5977	0.6684	0.6492	0.6467	0.0406	0.0420	0.0374	0.0393	0.0346	0.0327
		300	0.6005	0.5844	0.5834	0.6148	0.6066	0.6049	0.0185	0.0235	0.0244	0.0309	0.0246	0.0256
	4	25	0.5867	0.5661	0.5656	0.5798	0.5583	0.5576	0.0145	0.0301	0.0247	0.0280	0.0323	0.0365
		100	0.5811	0.5762	0.5739	0.6083	0.5968	0.5963	0.0172	0.0165	0.0126	0.0123	0.0151	0.0146
		300	0.5907	0.5570	0.5554	0.5720	0.5565	0.5567	0.0052	0.0107	0.0090	0.0112	0.0099	0.0096
	6	25	0.5804	0.5747	0.5753	0.5264	0.5202	0.5190	0.0118	0.0159	0.0190	0.0052	0.0148	0.0136
		100	0.5912	0.5741	0.5734	0.5611	0.5476	0.5475	0.0088	0.0074	0.0050	0.0084	0.0049	0.0064
		300	0.5792	0.5671	0.5672	0.5684	0.5637	0.5638	0.0012	0.0038	0.0032	0.0043	0.0030	0.0025



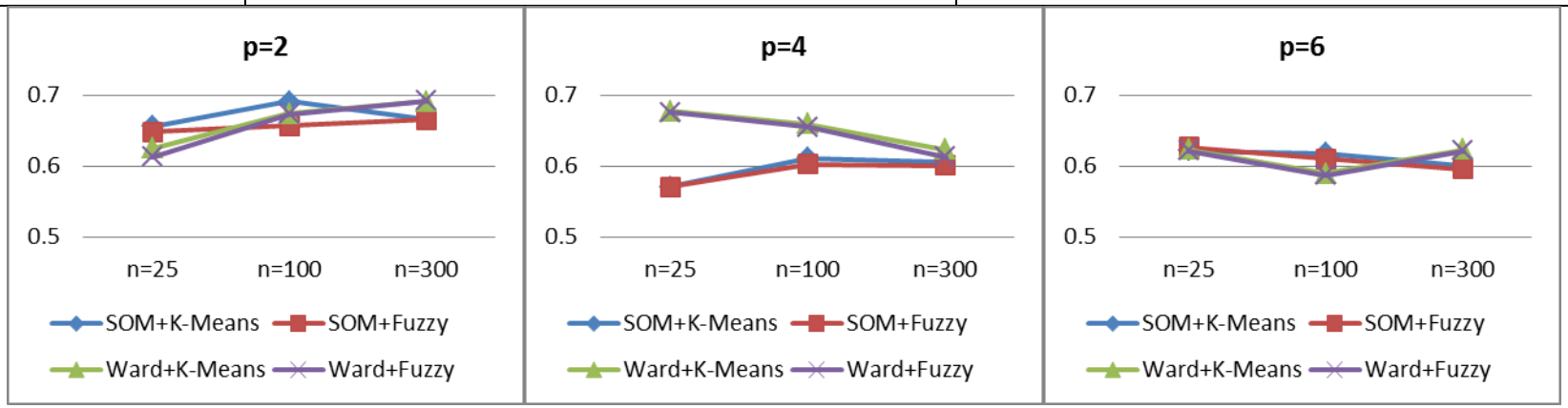
ตารางที่ 4.2 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.6841	0.6653	0.6621	0.6019	0.5951	0.5930	0.0663	0.0713	0.0728	0.0741	0.0735	0.0760
		100	0.6335	0.6209	0.6172	0.6455	0.6317	0.6303	0.0475	0.0466	0.0498	0.0595	0.0572	0.0557
		300	0.6233	0.6134	0.6127	0.6457	0.6369	0.6394	0.0472	0.0523	0.0514	0.0552	0.0541	0.0543
	4	25	0.5697	0.5568	0.5562	0.5412	0.5288	0.5298	0.0345	0.0403	0.0394	0.0337	0.0371	0.0402
		100	0.5863	0.5754	0.5728	0.5853	0.5772	0.5785	0.0306	0.0249	0.0252	0.0244	0.0296	0.0233
		300	0.6031	0.5974	0.5966	0.5877	0.5782	0.5781	0.0150	0.0200	0.0185	0.0157	0.0191	0.0210
	6	25	0.5196	0.5156	0.5142	0.5561	0.5435	0.5425	0.0275	0.0252	0.0232	0.0227	0.0280	0.0309
		100	0.5629	0.5507	0.5490	0.5697	0.5560	0.5562	0.0188	0.0134	0.0110	0.0121	0.0120	0.0109
		300	0.5849	0.5775	0.5773	0.6013	0.5804	0.5804	0.0052	0.0068	0.0071	0.0099	0.0064	0.0086



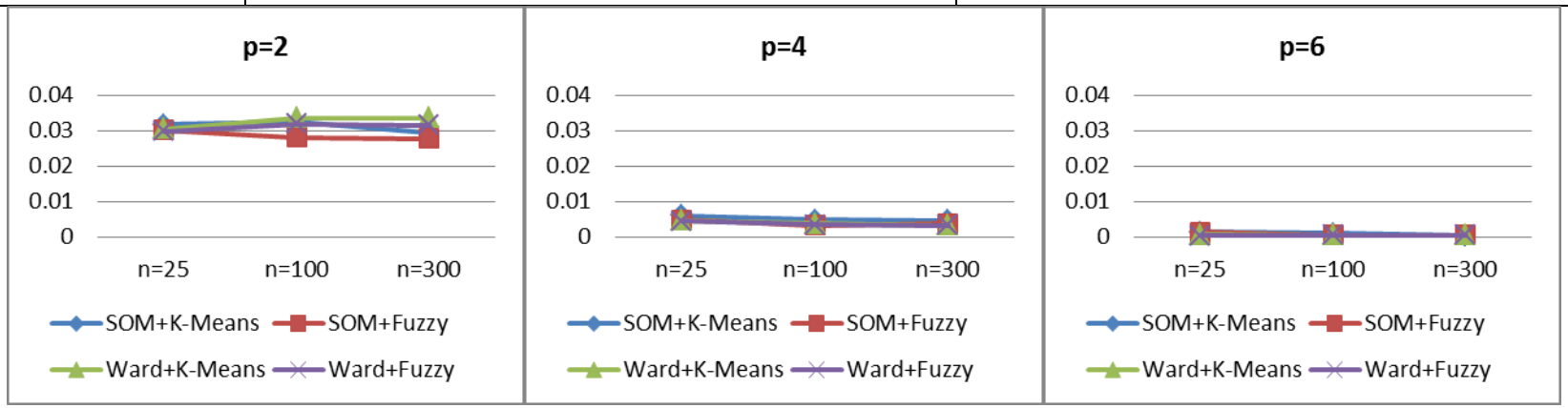
ตารางที่ 4.2 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.6601	0.6566	0.6484	0.6247	0.6254	0.6129	0.0877	0.0875	0.0832	0.0996	0.1005	0.0991
		100	0.7007	0.6913	0.6569	0.6777	0.6745	0.6734	0.0400	0.0428	0.0410	0.0491	0.0464	0.0446
		300	0.6700	0.6668	0.6659	0.6997	0.6914	0.6922	0.0321	0.0342	0.0342	0.0408	0.0364	0.0378
	4	25	0.5879	0.5705	0.5704	0.6870	0.6778	0.6757	0.0747	0.0839	0.0878	0.0683	0.0643	0.0635
		100	0.6138	0.6110	0.6026	0.6588	0.6596	0.6554	0.0390	0.0390	0.0347	0.0376	0.0340	0.0335
		300	0.6166	0.6053	0.6008	0.6278	0.6231	0.6131	0.0267	0.0331	0.0302	0.0373	0.0380	0.0377
	6	25	0.6363	0.6216	0.6264	0.6285	0.6231	0.6219	0.0539	0.0629	0.0636	0.0632	0.0603	0.0638
		100	0.6213	0.6183	0.6102	0.5956	0.5892	0.5866	0.0331	0.0330	0.0283	0.0458	0.0410	0.0405
		300	0.6079	0.5996	0.5958	0.6339	0.6237	0.6214	0.0160	0.0225	0.0208	0.0324	0.0285	0.0274



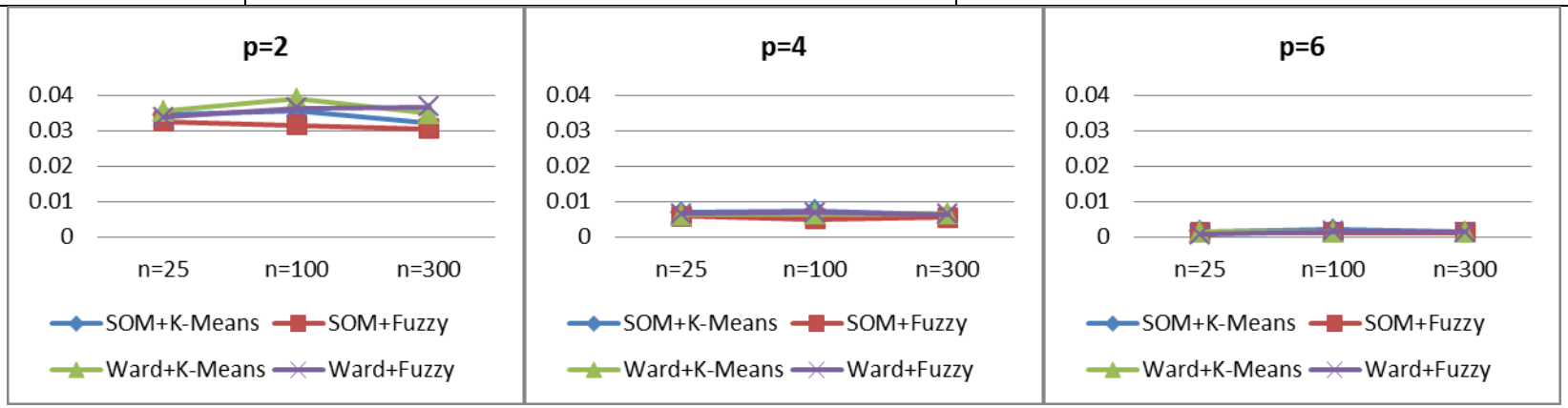
ตารางที่ 4.3 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.0443	0.0318	0.0302	0.0561	0.0305	0.0298	0.0275	0.0348	0.0322	0.0348	0.0352	0.0349
		100	0.0341	0.0324	0.0282	0.0378	0.0336	0.0319	0.0173	0.0243	0.0194	0.0293	0.0283	0.0245
		300	0.0312	0.0295	0.0278	0.0459	0.0336	0.0315	0.0082	0.0163	0.0133	0.0098	0.0193	0.0183
	4	25	0.0071	0.0060	0.0047	0.0051	0.0046	0.0044	0.0040	0.0115	0.0095	0.0151	0.0171	0.0189
		100	0.0050	0.0047	0.0032	0.0035	0.0037	0.0034	0.0080	0.0088	0.0057	0.0054	0.0073	0.0058
		300	0.0058	0.0046	0.0038	0.0035	0.0033	0.0032	0.0026	0.0053	0.0042	0.0049	0.0041	0.0037
	6	25	0.0017	0.0014	0.0013	0.0006	0.0006	0.0003	0.0044	0.0069	0.0069	0.0003	0.0004	0.0004
		100	0.0013	0.0010	0.0004	0.0035	0.0005	0.0004	0.0031	0.0036	0.0017	0.0035	0.0020	0.0029
		300	0.0005	0.0003	0.0004	0.0008	0.0004	0.0004	0.0004	0.0017	0.0013	0.0016	0.0012	0.0010



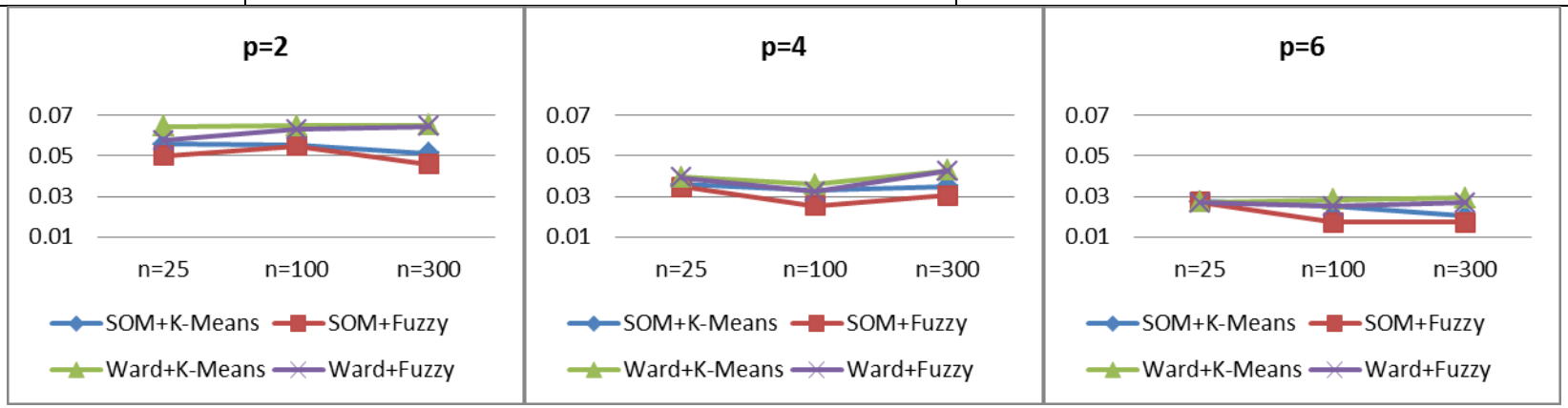
ตารางที่ 4.3 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.0457	0.0345	0.0326	0.0489	0.0355	0.0338	0.0326	0.0423	0.0373	0.0412	0.0452	0.0446
		100	0.0451	0.0356	0.0315	0.0474	0.0390	0.0364	0.0278	0.0322	0.0339	0.0329	0.0368	0.0338
		300	0.0411	0.0321	0.0306	0.0447	0.0349	0.0368	0.0234	0.0319	0.0292	0.0342	0.0400	0.0405
	4	25	0.0076	0.0069	0.0058	0.0067	0.0062	0.0065	0.0128	0.0167	0.0142	0.0228	0.0235	0.0234
		100	0.0082	0.0073	0.0050	0.0075	0.0063	0.0069	0.0128	0.0130	0.0118	0.0116	0.0146	0.0108
		300	0.0073	0.0061	0.0055	0.0072	0.0066	0.0064	0.0185	0.0077	0.0089	0.0096	0.0117	0.0145
	6	25	0.0020	0.0015	0.0012	0.0018	0.0014	0.0007	0.0104	0.0130	0.0104	0.0077	0.0123	0.0123
		100	0.0025	0.0021	0.0011	0.0019	0.0014	0.0015	0.0074	0.0042	0.0033	0.0043	0.0068	0.0045
		300	0.0022	0.0013	0.0011	0.0016	0.0014	0.0014	0.0020	0.0026	0.0027	0.0047	0.0027	0.0036



ตารางที่ 4.3 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.0628	0.0560	0.0500	0.0641	0.0648	0.0576	0.0403	0.0469	0.0445	0.0496	0.0497	0.0495
		100	0.0558	0.0553	0.0550	0.0615	0.0651	0.0630	0.0257	0.0308	0.0308	0.0358	0.0406	0.0384
		300	0.0686	0.0512	0.0459	0.0795	0.0653	0.0647	0.0177	0.0269	0.0264	0.0325	0.0354	0.0349
	4	25	0.0420	0.0360	0.0348	0.0455	0.0396	0.0392	0.0323	0.0412	0.0428	0.0431	0.0400	0.0373
		100	0.0306	0.0330	0.0253	0.0346	0.0359	0.0324	0.0201	0.0256	0.0176	0.0261	0.0251	0.0236
		300	0.0463	0.0350	0.0306	0.0570	0.0429	0.0427	0.0118	0.0187	0.0136	0.0229	0.0267	0.0276
	6	25	0.0364	0.0276	0.0272	0.0284	0.0272	0.0268	0.0242	0.0363	0.0350	0.0359	0.0355	0.0360
		100	0.0241	0.0249	0.0171	0.0347	0.0284	0.0250	0.0168	0.0209	0.0119	0.0232	0.0248	0.0216
		300	0.0256	0.0201	0.0173	0.0324	0.0292	0.0268	0.0284	0.0128	0.0099	0.0209	0.0210	0.0198



จากตารางที่ 4.1-4.3 ซึ่งแสดงผลการเปรียบเทียบค่า Wilk' s Lambda ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ตามลำดับ ของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 2 กลุ่ม พบว่า การจัดกลุ่มแบบ 2 ขั้นตอน ด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี Ward ร่วมกับ Fuzzy C-Means มีประสิทธิภาพดีกว่าวิธี SOM ร่วมกับ K-Means และวิธี Ward ร่วมกับ K-Means ในทุกกรณีสำหรับทุกขนาดตัวอย่าง และทุกจำนวนตัวแปร โดยให้ค่า Wilk's Lambda, ค่า RMSSTD และค่า Misclassification Rate ต่ำที่สุด

กรณีที่ 1 ตัวแปรอิสระต่อกัน

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

กรณีที่ 2 ตัวแปรมีความสัมพันธ์กัน โดย $\rho = Corr(X_i, X_j) \sim Unif(-1,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ ถ้าขนาดตัวอย่างเพิ่มขึ้น (ยกเว้น จำนวนตัวแปรเท่ากับ 2 พบว่าค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าขนาดตัวอย่างเพิ่มขึ้น)

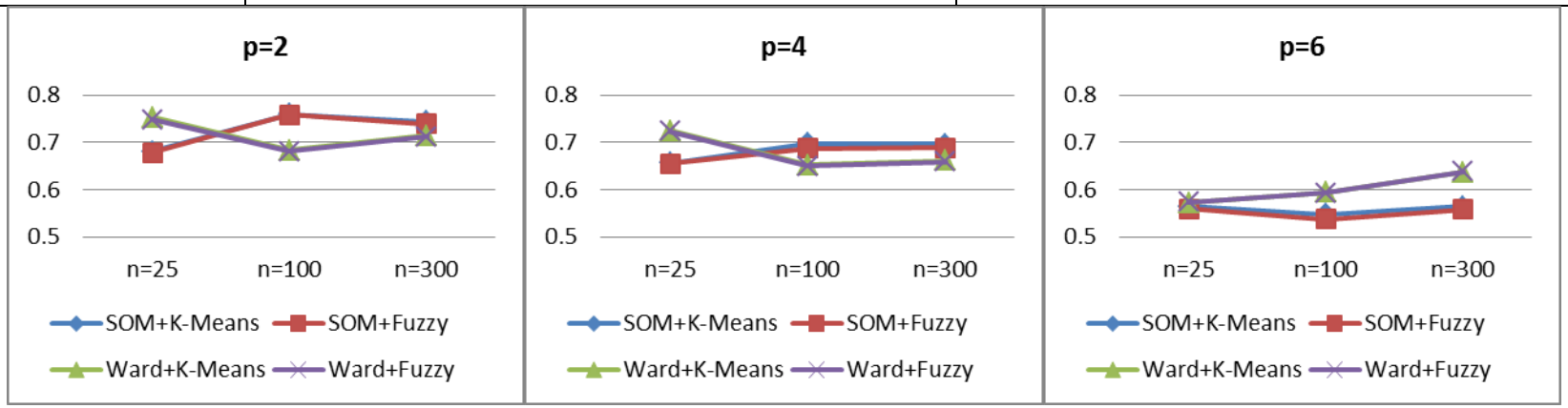
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

กรณีที่ 3 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(0,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

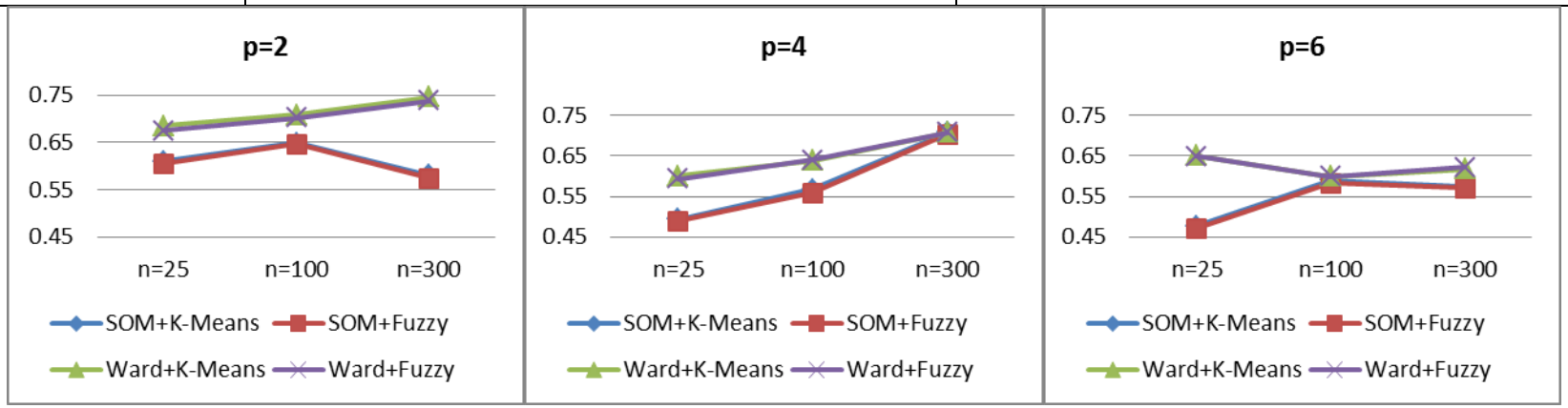
ตารางที่ 4.4 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.6957	0.6810	0.6790	0.7760	0.7533	0.7486	0.0893	0.0871	0.0800	0.0928	0.1043	0.1011
		100	0.7723	0.7604	0.7594	0.7042	0.6831	0.6808	0.0608	0.0615	0.0583	0.0593	0.0595	0.0552
		300	0.7540	0.7452	0.7401	0.7314	0.7145	0.7119	0.0388	0.0352	0.0338	0.0418	0.0361	0.0347
	4	25	0.6678	0.6562	0.6550	0.7409	0.7249	0.7233	0.0698	0.0716	0.0647	0.0750	0.0800	0.0719
		100	0.7169	0.6980	0.6880	0.6657	0.6521	0.6503	0.0450	0.0435	0.0433	0.0547	0.0453	0.0469
		300	0.6845	0.6961	0.6886	0.6745	0.6622	0.6597	0.0240	0.0265	0.0242	0.0333	0.0277	0.0294
	6	25	0.5864	0.5638	0.5594	0.5881	0.5716	0.5735	0.0510	0.0607	0.0567	0.0617	0.0595	0.0577
		100	0.5496	0.5478	0.5366	0.6143	0.5939	0.5935	0.0311	0.0317	0.0268	0.0371	0.0310	0.0347
		300	0.5813	0.5649	0.5577	0.6451	0.6370	0.6386	0.0148	0.0204	0.0158	0.0213	0.0219	0.0220



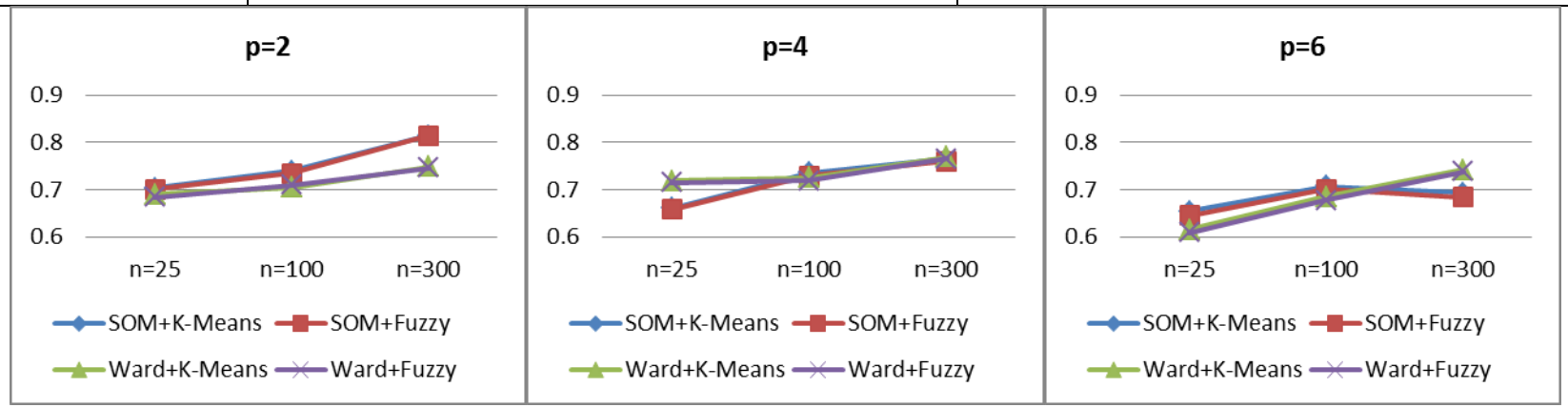
ตารางที่ 4.4 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.6198	0.6110	0.6054	0.6924	0.6855	0.6750	0.1238	0.1290	0.1245	0.1307	0.0511	0.1396
		100	0.6547	0.6496	0.6463	0.7155	0.7087	0.7030	0.0843	0.0867	0.0839	0.0823	0.0878	0.0865
		300	0.5883	0.5802	0.5745	0.7563	0.7467	0.7383	0.0856	0.0853	0.0865	0.0912	0.0825	0.0825
	4	25	0.5124	0.4937	0.4884	0.6223	0.6006	0.5930	0.0993	0.0874	0.0796	0.0956	0.0998	0.0926
		100	0.5865	0.5670	0.5590	0.6499	0.6382	0.6400	0.0621	0.0554	0.0545	0.0804	0.0706	0.0687
		300	0.7214	0.7087	0.7039	0.7146	0.7076	0.7077	0.0598	0.0561	0.0552	0.0620	0.0559	0.0499
	6	25	0.4798	0.4760	0.4703	0.6658	0.6506	0.6485	0.0593	0.0665	0.0620	0.0614	0.0695	0.0660
		100	0.5965	0.5899	0.5831	0.6012	0.5993	0.5993	0.0511	0.0523	0.0383	0.0450	0.0465	0.0412
		300	0.5844	0.5730	0.5702	0.6357	0.6172	0.6215	0.0411	0.0357	0.0344	0.0447	0.0453	0.0468



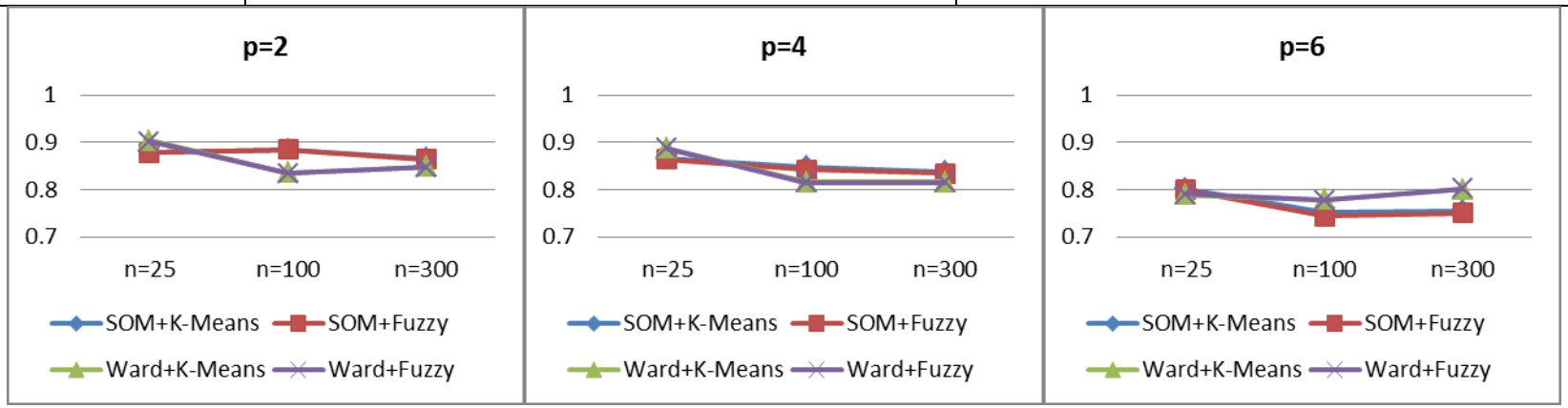
ตารางที่ 4.4 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.7081	0.7039	0.7001	0.7159	0.6907	0.6845	0.0830	0.0889	0.0855	0.0985	0.1079	0.1015
		100	0.7529	0.7402	0.7349	0.7186	0.7052	0.7104	0.0680	0.0618	0.0694	0.0825	0.0785	0.0767
		300	0.8159	0.8154	0.8152	0.7658	0.7488	0.7462	0.0472	0.0430	0.0400	0.0546	0.0482	0.0494
	4	25	0.6675	0.6608	0.6584	0.7321	0.7189	0.7156	0.0914	0.0918	0.0777	0.0658	0.0773	0.0731
		100	0.7451	0.7361	0.7294	0.7338	0.7249	0.7194	0.0500	0.0477	0.0530	0.0698	0.0605	0.0565
		300	0.7676	0.7658	0.7607	0.7764	0.7693	0.7654	0.0320	0.0269	0.0292	0.0387	0.0345	0.0311
	6	25	0.6639	0.6537	0.6451	0.6268	0.6157	0.6082	0.1104	0.0950	0.0970	0.0991	0.0986	0.0916
		100	0.7171	0.7071	0.7013	0.6926	0.6851	0.6779	0.0469	0.0402	0.0399	0.0579	0.0506	0.0515
		300	0.7043	0.6923	0.6844	0.7535	0.7423	0.7388	0.0416	0.0377	0.0373	0.0365	0.0342	0.0331



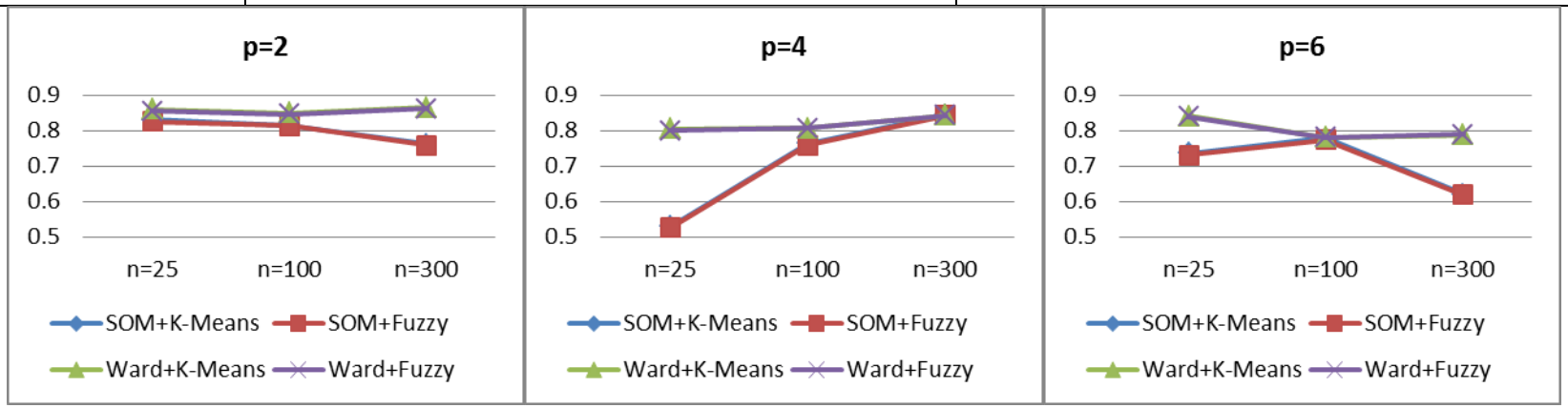
ตารางที่ 4.5 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.8834	0.8799	0.8787	0.9210	0.9045	0.9014	0.0573	0.0568	0.0513	0.0572	0.0599	0.0581
		100	0.8895	0.8851	0.8844	0.8529	0.8362	0.8346	0.0386	0.0383	0.0364	0.0357	0.0357	0.0334
		300	0.8814	0.8673	0.8644	0.8611	0.8486	0.8470	0.0224	0.0203	0.0195	0.0241	0.0210	0.0201
	4	25	0.8763	0.8650	0.8643	0.8910	0.8878	0.8870	0.0475	0.0483	0.0439	0.0479	0.0467	0.0438
		100	0.8507	0.8481	0.8421	0.8256	0.8160	0.8148	0.0299	0.0290	0.0289	0.0352	0.0294	0.0298
		300	0.8491	0.8384	0.8339	0.8224	0.8168	0.8153	0.0147	0.0162	0.0148	0.0204	0.0175	0.0181
	6	25	0.8179	0.8018	0.7988	0.7998	0.7892	0.7904	0.0367	0.0435	0.0406	0.0410	0.0360	0.0356
		100	0.7618	0.7512	0.7436	0.7852	0.7782	0.7780	0.0206	0.0210	0.0179	0.0239	0.0203	0.0225
		300	0.7682	0.7552	0.7506	0.8066	0.8008	0.8018	0.0092	0.0126	0.0098	0.0136	0.0142	0.0142



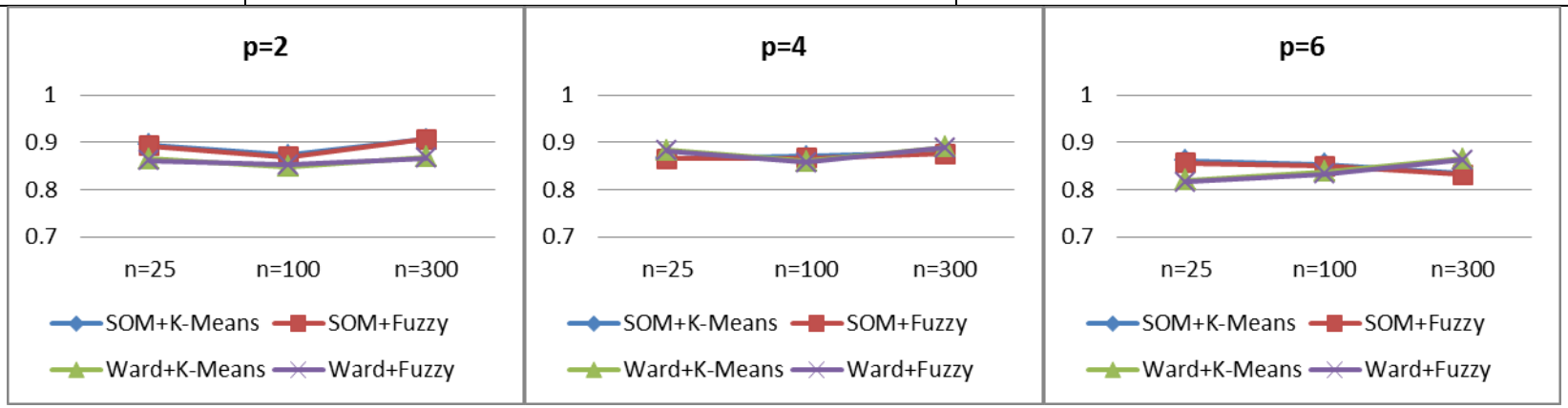
ตารางที่ 4.5 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.8439	0.8309	0.8269	0.8673	0.8611	0.8552	0.0883	0.0906	0.0878	0.0868	0.0957	0.0903
		100	0.8198	0.8164	0.8142	0.8587	0.8504	0.8470	0.0511	0.0521	0.0506	0.0489	0.0524	0.0520
		300	0.7694	0.7621	0.7583	0.8797	0.8666	0.8618	0.0511	0.0511	0.0518	0.0565	0.0516	0.0516
	4	25	0.5369	0.5283	0.5256	0.8145	0.8053	0.80080	0.0676	0.0601	0.0552	0.0621	0.0595	0.0571
		100	0.7811	0.7634	0.7581	0.8172	0.8066	0.8079	0.0393	0.0355	0.0351	0.0529	0.0474	0.0463
		300	0.8598	0.8456	0.8429	0.8553	0.8435	0.8437	0.0370	0.0350	0.0345	0.0373	0.0341	0.0305
	6	25	0.7485	0.7359	0.7314	0.8547	0.8417	0.8405	0.0453	0.0507	0.0472	0.0403	0.0453	0.0436
		100	0.7845	0.7795	0.7751	0.8001	0.7810	0.7812	0.0349	0.0356	0.0268	0.0289	0.0302	0.0265
		300	0.6298	0.6208	0.6193	0.8042	0.7879	0.7905	0.0277	0.0242	0.0237	0.0285	0.0290	0.0298



ตารางที่ 4.5 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.9014	0.8948	0.8926	0.8811	0.8653	0.8618	0.0528	0.0575	0.0543	0.0617	0.0657	0.0616
		100	0.8884	0.8729	0.8696	0.8547	0.8491	0.8523	0.0405	0.0365	0.0412	0.0495	0.0477	0.0460
		300	0.9119	0.9073	0.9072	0.8773	0.8696	0.8673	0.0265	0.0239	0.0222	0.0315	0.0280	0.0287
	4	25	0.8714	0.8669	0.8659	0.8922	0.8838	0.8820	0.0607	0.0613	0.0512	0.0409	0.0448	0.0424
		100	0.8761	0.8708	0.8667	0.8642	0.8604	0.8572	0.0298	0.0283	0.0316	0.0415	0.0363	0.0338
		300	0.8846	0.8794	0.8764	0.8918	0.8899	0.8892	0.0185	0.0154	0.0168	0.0199	0.0168	0.0149
	6	25	0.8747	0.8621	0.8562	0.8235	0.8204	0.8157	0.0733	0.0632	0.0651	0.0662	0.0646	0.0601
		100	0.8595	0.8536	0.8501	0.8398	0.8377	0.8332	0.0283	0.0243	0.0242	0.0351	0.0308	0.0311
		300	0.8410	0.8359	0.8312	0.8706	0.8652	0.8631	0.0253	0.0227	0.0227	0.0209	0.0197	0.0191

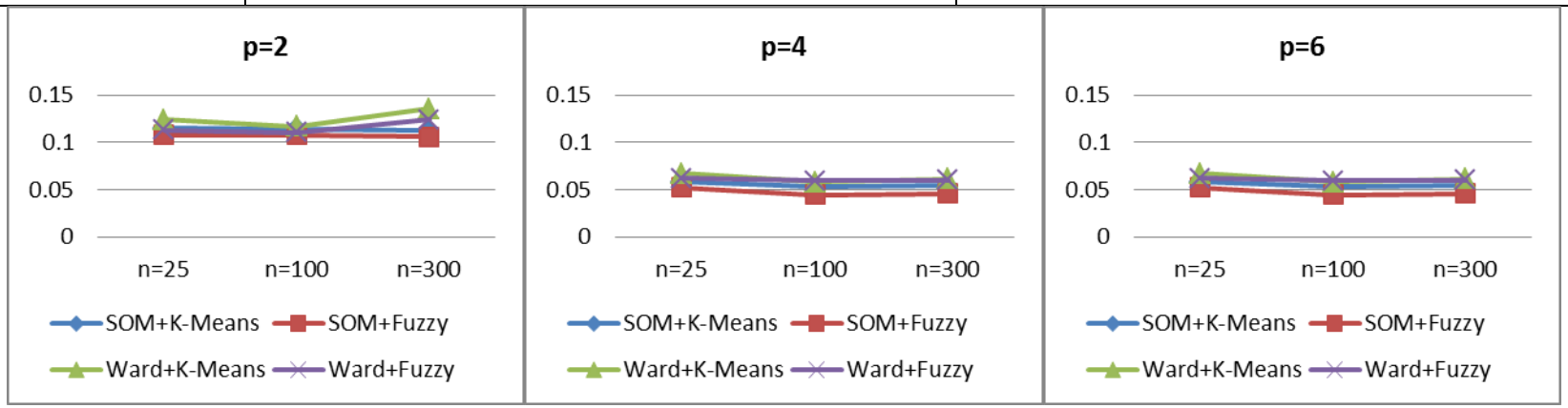


ตารางที่ 4.6 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.1346	0.1220	0.1220	0.1355	0.1301	0.1266	0.0667	0.0592	0.0604	0.0695	0.0731	0.0677
		100	0.1442	0.1327	0.1262	0.1407	0.1359	0.1367	0.0381	0.0433	0.0394	0.0498	0.0504	0.0485
		300	0.1457	0.1354	0.1255	0.1477	0.1393	0.1359	0.0318	0.0385	0.0339	0.0458	0.0389	0.0368
	4	25	0.0696	0.0634	0.0586	0.0660	0.0579	0.0567	0.0421	0.0546	0.0522	0.0585	0.0700	0.0513
		100	0.0788	0.0652	0.0510	0.0689	0.0619	0.0647	0.0261	0.0331	0.0270	0.0401	0.0362	0.0379
		300	0.0668	0.0609	0.0490	0.0681	0.0604	0.0650	0.0183	0.0277	0.0209	0.0283	0.0279	0.0335
	6	25	0.0363	0.0322	0.0262	0.0368	0.0337	0.0340	0.0329	0.0394	0.0392	0.0457	0.0403	0.0374
		100	0.0364	0.0304	0.0208	0.0299	0.0255	0.0259	0.0187	0.0235	0.0157	0.0272	0.0200	0.0256
		300	0.0363	0.0273	0.0209	0.0343	0.0247	0.0275	0.0102	0.0193	0.0110	0.0144	0.0193	0.0201

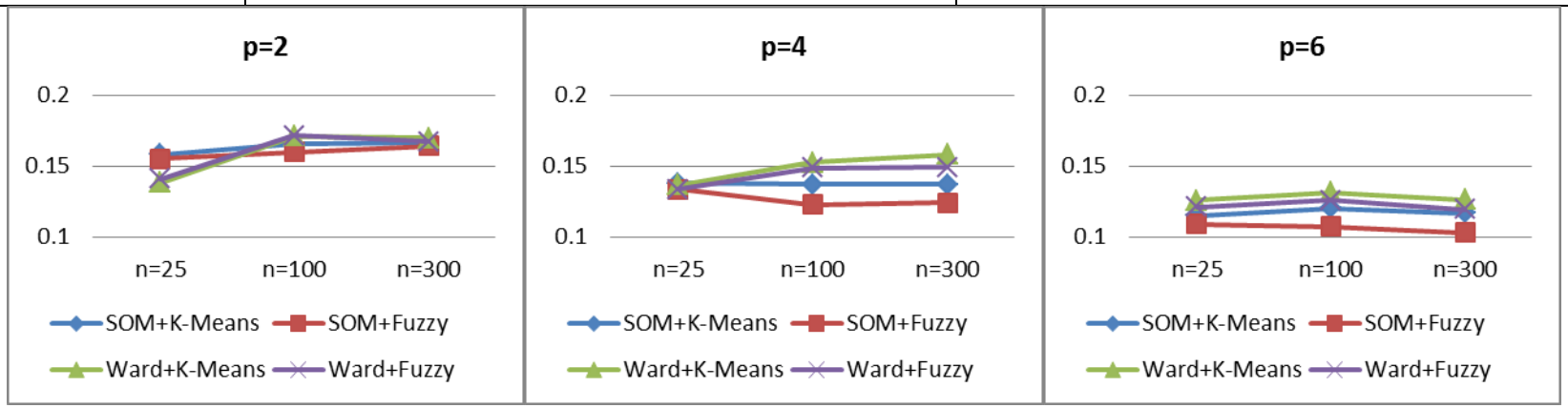
ตารางที่ 4.6 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.1187	0.1154	0.1082	0.1279	0.1240	0.1132	0.0714	0.0814	0.0762	0.0853	0.0924	0.0872
		100	0.1199	0.1143	0.1083	0.1235	0.1174	0.1107	0.0677	0.0717	0.0709	0.0738	0.0822	0.0819
		300	0.1185	0.1125	0.1064	0.1454	0.1357	0.1244	0.0643	0.0736	0.0729	0.0703	0.0684	0.0676
	4	25	0.0685	0.0580	0.0524	0.0688	0.0673	0.0618	0.0686	0.0594	0.0588	0.0728	0.0703	0.0634
		100	0.0655	0.0530	0.0444	0.0647	0.0584	0.0597	0.0458	0.0487	0.0387	0.0549	0.0525	0.0479
		300	0.0641	0.0543	0.0458	0.0764	0.0611	0.0599	0.0438	0.0460	0.0403	0.0557	0.0552	0.0513
	6	25	0.0456	0.0322	0.0266	0.0458	0.0342	0.0314	0.0352	0.0445	0.0398	0.0441	0.0438	0.0481
		100	0.0298	0.0275	0.0203	0.0459	0.0314	0.0307	0.0299	0.0354	0.0212	0.0342	0.0354	0.0293
		300	0.0345	0.0250	0.0212	0.0356	0.0272	0.0314	0.0247	0.0249	0.0204	0.0324	0.0349	0.0386



ตารางที่ 4.6 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 2 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.1620	0.1580	0.1552	0.1504	0.1384	0.1408	0.0704	0.0678	0.0659	0.0635	0.0695	0.0659
		100	0.1706	0.1654	0.1598	0.1715	0.1712	0.1716	0.0466	0.0466	0.0499	0.0551	0.0550	0.0519
		300	0.1683	0.1667	0.1641	0.1801	0.1700	0.1671	0.0427	0.0404	0.0394	0.0525	0.0448	0.0446
	4	25	0.1256	0.1380	0.1336	0.1445	0.1364	0.1336	0.0710	0.0766	0.0711	0.0678	0.0631	0.0618
		100	0.1282	0.1371	0.1229	0.1451	0.1526	0.1489	0.0368	0.0462	0.0410	0.0541	0.0503	0.0467
		300	0.1379	0.1372	0.1240	0.1610	0.1581	0.1493	0.0276	0.0322	0.0301	0.0310	0.0358	0.0374
	6	25	0.1220	0.1148	0.1088	0.1324	0.1256	0.1212	0.0561	0.0556	0.0537	0.0760	0.0696	0.0654
		100	0.1127	0.1202	0.1073	0.1266	0.1316	0.1257	0.0366	0.0416	0.0405	0.0474	0.0436	0.0452
		300	0.1223	0.1167	0.1028	0.1280	0.1263	0.1193	0.0228	0.0333	0.0264	0.0396	0.0375	0.0353



จากตารางที่ 4.4-4.6 ซึ่งแสดงผลการเปรียบเทียบค่า Wilk' s Lambda ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ตามลำดับ ของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองเป็น 2 กลุ่ม พบว่า การจัดกลุ่มแบบ 2 ขั้นตอน ด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี Ward ร่วมกับ Fuzzy C-Means มีประสิทธิภาพดีกว่าวิธี SOM ร่วมกับ K-Means และวิธี Ward ร่วมกับ K-Means ในทุกกรณีสำหรับทุกขนาดตัวอย่าง และทุกจำนวนตัวแปร โดยให้ค่า Wilk's Lambda, ค่า RMSSTD และค่า Misclassification Rate ต่ำที่สุด

กรณีที่ 1 ตัวแปรอิสระต่อกัน

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

กรณีที่ 2 ตัวแปรมีความสัมพันธ์กัน โดย $\rho = Corr(X_i, X_j) \sim Unif(-1,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น

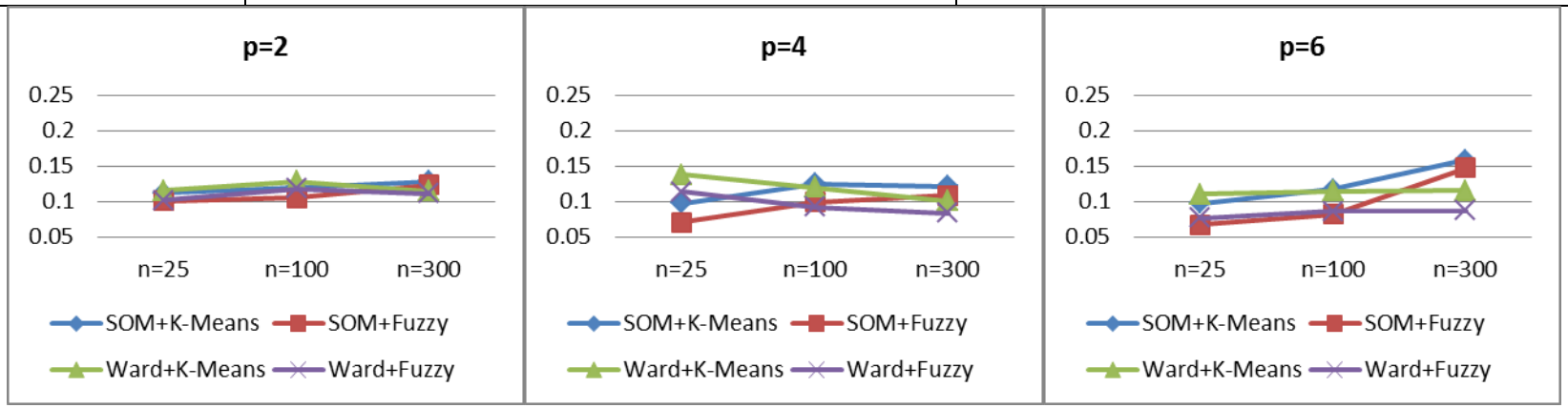
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

กรณีที่ 3 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(0,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ ถ้าขนาดตัวอย่างเพิ่มขึ้น

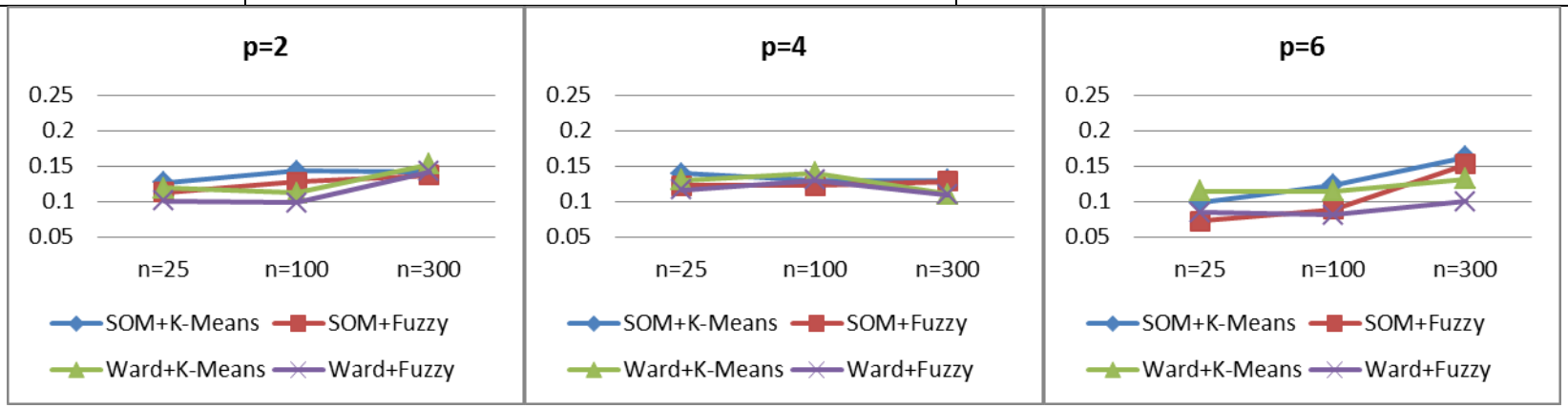
ตารางที่ 4.7 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.1992	0.1123	0.1003	0.1971	0.1153	0.1013	0.0443	0.0568	0.0445	0.0433	0.0453	0.0310
		100	0.2294	0.1199	0.1050	0.2316	0.1286	0.1183	0.0434	0.0400	0.0171	0.0321	0.0399	0.0182
		300	0.2761	0.1287	0.1232	0.2203	0.1146	0.1108	0.0102	0.0243	0.0114	0.0252	0.0258	0.0129
	4	25	0.2122	0.0976	0.0704	0.2260	0.1380	0.1133	0.0266	0.0490	0.0249	0.0235	0.0409	0.0252
		100	0.2112	0.1245	0.0986	0.2113	0.1195	0.0920	0.0250	0.0450	0.0060	0.0262	0.0425	0.0068
		300	0.2657	0.1213	0.1085	0.1918	0.1011	0.0838	0.0062	0.0488	0.0476	0.0358	0.0462	0.0119
	6	25	0.1893	0.0974	0.0674	0.1936	0.1102	0.0769	0.0521	0.0484	0.0188	0.0241	0.0462	0.0176
		100	0.1665	0.1172	0.0817	0.1970	0.1142	0.0859	0.0236	0.0493	0.0021	0.0207	0.0503	0.0036
		300	0.2844	0.1586	0.1478	0.2007	0.1151	0.0869	0.0062	0.0500	0.0570	0.0285	0.0481	0.0011



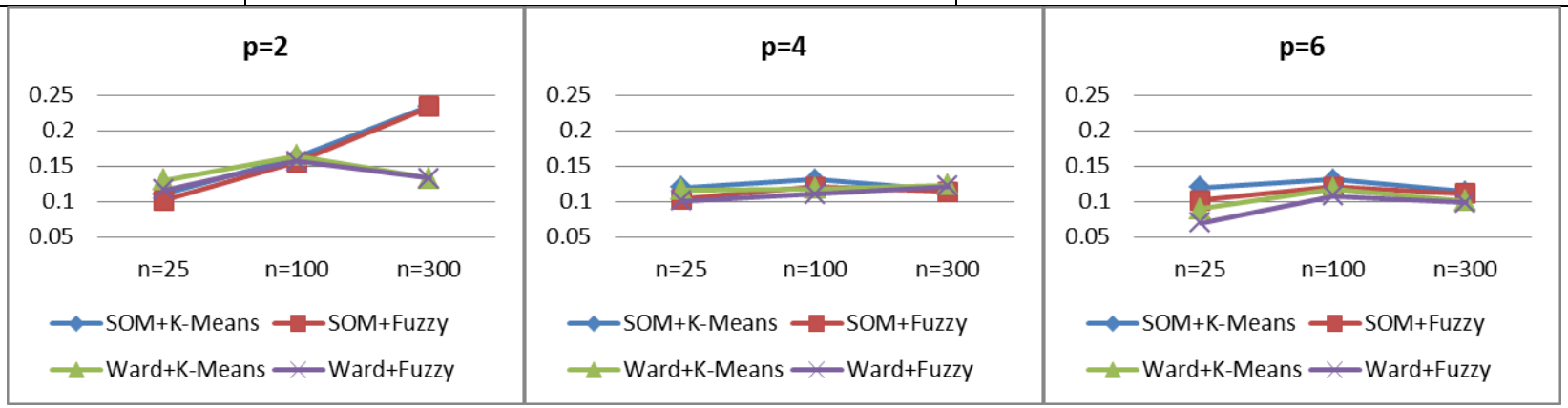
ตารางที่ 4.7 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.2015	0.1269	0.1132	0.2002	0.1187	0.1004	0.0390	0.0547	0.049	0.0483	0.0538	0.0493
		100	0.2268	0.1435	0.1279	0.2188	0.1121	0.0979	0.0433	0.0445	0.0302	0.0336	0.0431	0.0303
		300	0.2483	0.1417	0.1369	0.2605	0.1526	0.1423	0.0177	0.0438	0.0401	0.0422	0.0354	0.0270
	4	25	0.2045	0.1393	0.1219	0.1897	0.1301	0.1166	0.0238	0.0533	0.0289	0.0393	0.0551	0.0257
		100	0.2369	0.1297	0.1226	0.2113	0.1404	0.1299	0.0256	0.0430	0.0107	0.0238	0.0484	0.0105
		300	0.2458	0.1300	0.1284	0.2122	0.1105	0.1092	0.0088	0.0469	0.0514	0.0267	0.0481	0.013
	6	25	0.2231	0.0985	0.0721	0.1965	0.1147	0.0841	0.0240	0.0459	0.0191	0.0178	0.0499	0.0180
		100	0.2181	0.1231	0.0882	0.1975	0.1142	0.0808	0.0290	0.0532	0.0144	0.0326	0.0477	0.0038
		300	0.2553	0.1630	0.1528	0.1961	0.1319	0.1000	0.0064	0.0491	0.0562	0.0344	0.0557	0.0024



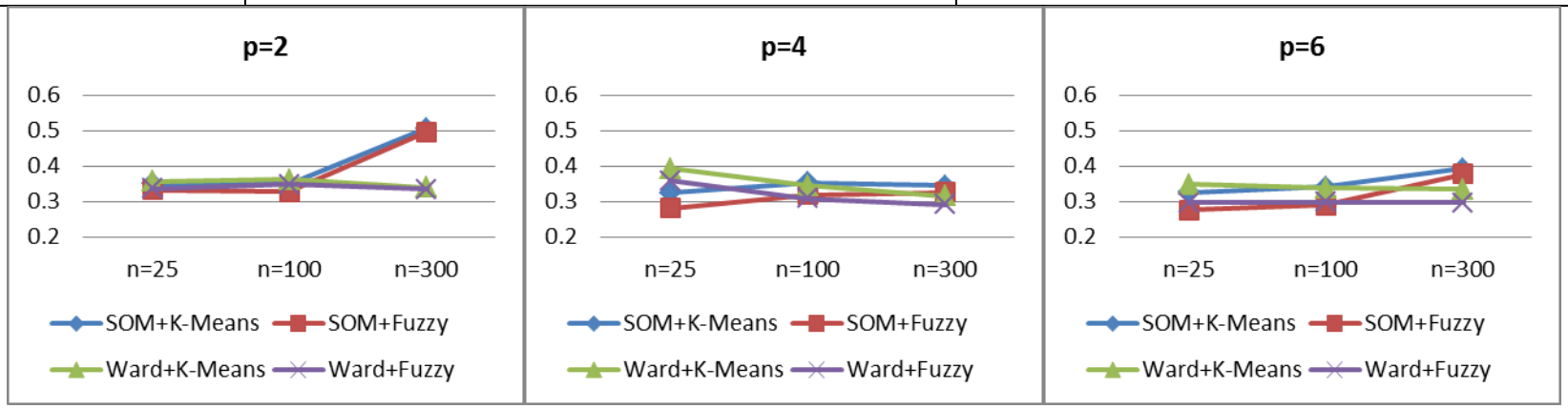
ตารางที่ 4.7 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.1984	0.1106	0.1016	0.2086	0.1301	0.1162	0.0520	0.0502	0.0492	0.0455	0.0513	0.0511
		100	0.2439	0.1622	0.1552	0.2586	0.1649	0.1567	0.0308	0.0372	0.0293	0.0399	0.0403	0.0292
		300	0.3642	0.2350	0.2338	0.2364	0.1331	0.1324	0.0163	0.0284	0.0255	0.0272	0.0234	0.0203
	4	25	0.2576	0.1200	0.1031	0.2143	0.1163	0.1006	0.0422	0.0481	0.0370	0.0341	0.0470	0.0372
		100	0.1996	0.1308	0.1207	0.2198	0.1176	0.1102	0.0246	0.0313	0.0175	0.0266	0.0341	0.0187
		300	0.2512	0.1154	0.1140	0.2333	0.1228	0.1212	0.0109	0.0198	0.0137	0.0273	0.0171	0.0152
	6	25	0.2242	0.1201	0.1019	0.1785	0.0896	0.0695	0.0363	0.0461	0.0336	0.0310	0.0399	0.0304
		100	0.2000	0.1307	0.1207	0.2258	0.1181	0.1076	0.0228	0.0393	0.0138	0.0167	0.0339	0.0137
		300	0.2747	0.1140	0.1110	0.2158	0.1004	0.0985	0.0088	0.0167	0.0097	0.0253	0.0202	0.0101



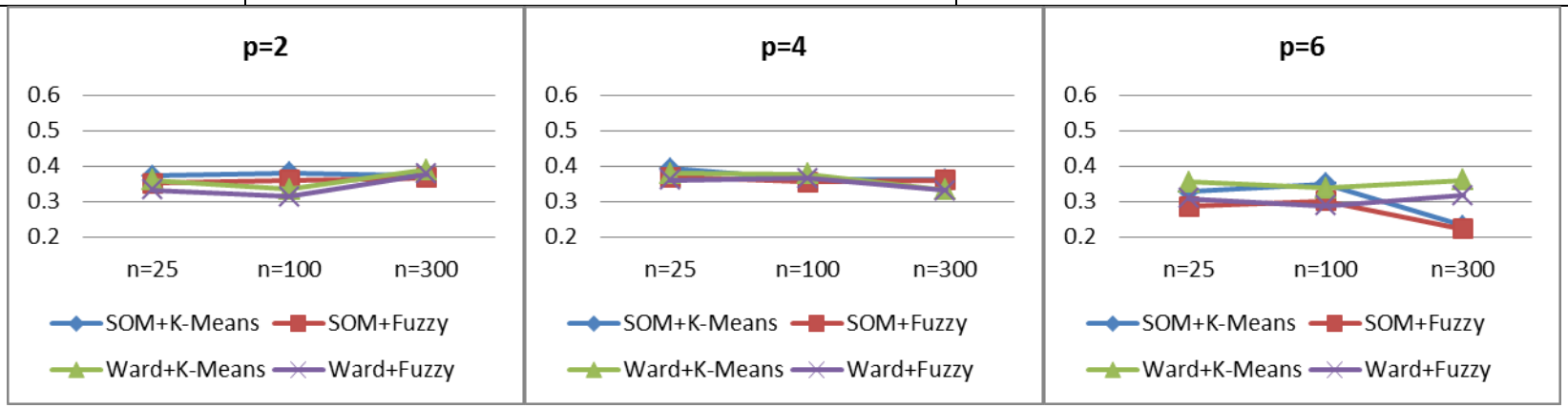
ตารางที่ 4.8 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.4746	0.3504	0.3328	0.4456	0.3571	0.3354	0.0435	0.0782	0.0632	0.0486	0.0729	0.0540
		100	0.4849	0.3476	0.3279	0.4732	0.3611	0.3481	0.0419	0.0509	0.0250	0.0294	0.0503	0.0252
		300	0.7467	0.5069	0.4976	0.4579	0.3387	0.3341	0.0102	0.0321	0.0169	0.0223	0.0304	0.0172
	4	25	0.4909	0.3234	0.2801	0.4799	0.3925	0.3589	0.0338	0.0742	0.0379	0.0206	0.0609	0.0379
		100	0.4658	0.3530	0.3188	0.4466	0.3445	0.3076	0.0301	0.0633	0.0103	0.0218	0.0558	0.0109
		300	0.5179	0.3438	0.3262	0.4225	0.3149	0.2907	0.0063	0.0678	0.0649	0.0317	0.0639	0.0168
	6	25	0.4641	0.3252	0.2756	0.4369	0.3474	0.2958	0.0226	0.0203	0.0073	0.0216	0.0719	0.0275
		100	0.4134	0.3406	0.2902	0.4306	0.3369	0.2976	0.0294	0.0685	0.0034	0.0162	0.0696	0.0060
		300	0.5355	0.3931	0.3778	0.4324	0.3351	0.2962	0.0058	0.0673	0.0756	0.0243	0.0682	0.0020



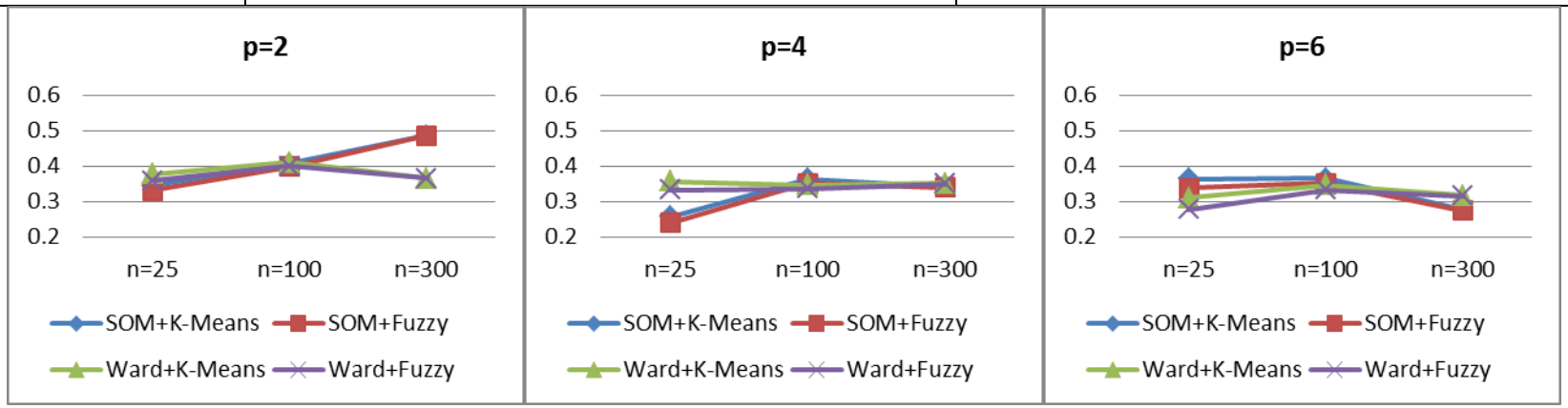
ตารางที่ 4.8 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.4870	0.3723	0.3518	0.4721	0.3609	0.3326	0.0461	0.0901	0.0837	0.0496	0.0748	0.0691
		100	0.4984	0.3805	0.3606	0.4501	0.3342	0.3136	0.0443	0.0598	0.0437	0.0329	0.0542	0.0399
		300	0.5004	0.3744	0.3677	0.5264	0.3900	0.3773	0.0181	0.0579	0.0543	0.0421	0.0475	0.0379
	4	25	0.5723	0.3928	0.3685	0.4301	0.3795	0.3605	0.0241	0.0754	0.0421	0.0377	0.0851	0.0437
		100	0.4890	0.3635	0.3547	0.4337	0.3779	0.3650	0.0264	0.0576	0.0165	0.0211	0.0675	0.0175
		300	0.4980	0.3613	0.3595	0.4461	0.3331	0.3315	0.0089	0.0653	0.0700	0.0225	0.0631	0.0189
	6	25	0.5042	0.3280	0.2851	0.4420	0.3552	0.3087	0.0280	0.0691	0.0295	0.0149	0.0754	0.0280
		100	0.4732	0.3499	0.3011	0.4323	0.3369	0.2883	0.0317	0.0718	0.0197	0.0284	0.0666	0.0063
		300	0.5074	0.2306	0.2226	0.4265	0.3598	0.3177	0.0062	0.0663	0.0749	0.0299	0.0756	0.0041



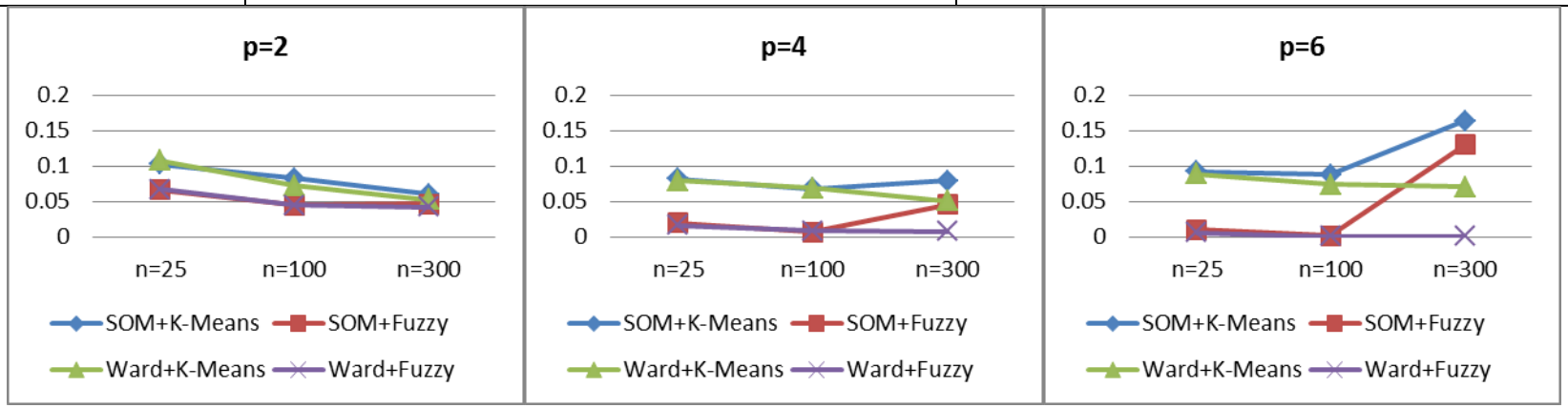
ตารางที่ 4.8 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.4714	0.3442	0.3301	0.4601	0.3783	0.3578	0.0556	0.0610	0.0609	0.0485	0.0678	0.068
		100	0.4999	0.4067	0.3987	0.5040	0.4099	0.4003	0.0310	0.0447	0.0364	0.0376	0.0513	0.0400
		300	0.6064	0.4866	0.4856	0.4778	0.3655	0.3647	0.0159	0.0381	0.0346	0.0255	0.0298	0.0265
	4	25	0.3820	0.2564	0.2389	0.4633	0.3566	0.3330	0.0604	0.0919	0.0743	0.0378	0.0683	0.0569
		100	0.4526	0.3645	0.3517	0.4602	0.3457	0.3359	0.0273	0.0408	0.0251	0.0253	0.0431	0.0258
		300	0.5036	0.3404	0.3388	0.4728	0.3512	0.3495	0.0107	0.0270	0.0199	0.0251	0.0229	0.0208
	6	25	0.5043	0.3641	0.3374	0.4172	0.3100	0.2764	0.0399	0.0668	0.0503	0.0302	0.0528	0.0409
		100	0.4533	0.3648	0.3522	0.4658	0.3458	0.3322	0.0261	0.0509	0.0207	0.0158	0.0424	0.0194
		300	0.4301	0.2761	0.2731	0.4505	0.3174	0.3150	0.0084	0.0222	0.0141	0.0217	0.0271	0.0152



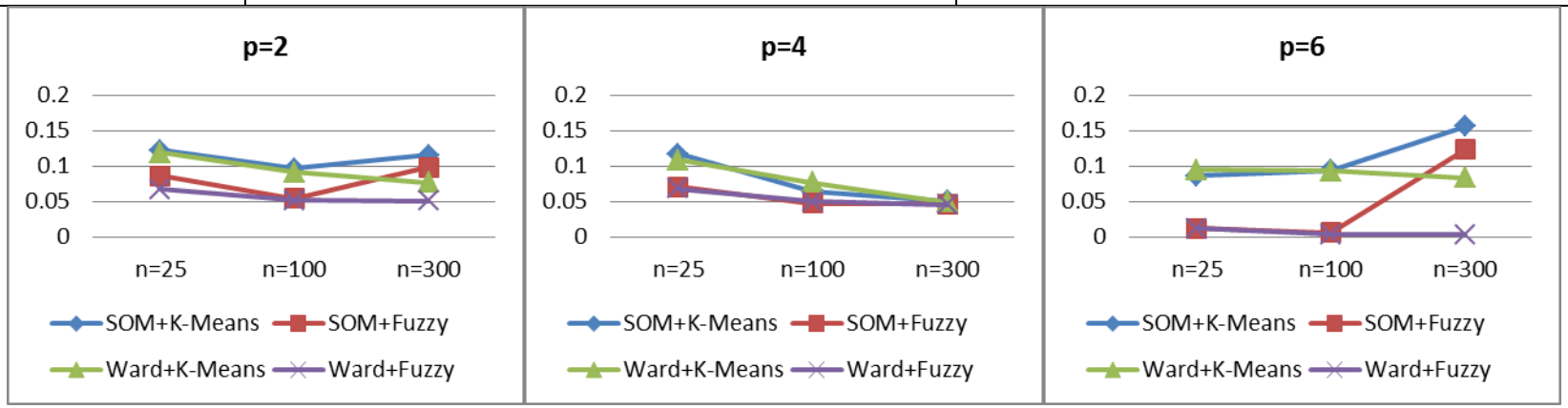
ตารางที่ 4.9 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.6371	0.1027	0.0664	0.7331	0.1078	0.0675	0.0731	0.1030	0.0612	0.0277	0.1017	0.0600
		100	0.7107	0.0824	0.0447	0.7249	0.0724	0.0451	0.0613	0.0841	0.0233	0.0131	0.0902	0.0301
		300	0.5746	0.0601	0.0462	0.7273	0.0523	0.0422	0.0267	0.0505	0.0144	0.0114	0.0448	0.0168
	4	25	0.6630	0.0820	0.0197	0.7553	0.0794	0.0165	0.0825	0.1157	0.0488	0.0098	0.1037	0.0544
		100	0.7910	0.0678	0.0067	0.7470	0.0693	0.0080	0.0521	0.1116	0.0084	0.0050	0.0905	0.0080
		300	0.5598	0.0792	0.0457	0.7468	0.0507	0.0071	0.0290	0.1151	0.1008	0.0028	0.1060	0.0250
	6	25	0.6685	0.0924	0.0103	0.7590	0.0888	0.0058	0.0720	0.1174	0.0420	0.0069	0.1179	0.0416
		100	0.8330	0.0883	0.0014	0.7495	0.0738	0.0008	0.0500	0.1245	0.0033	0.0025	0.1164	0.0042
		300	0.5473	0.1638	0.1303	0.7496	0.0708	0.0010	0.0256	0.1218	0.1259	0.0009	0.1184	0.0015



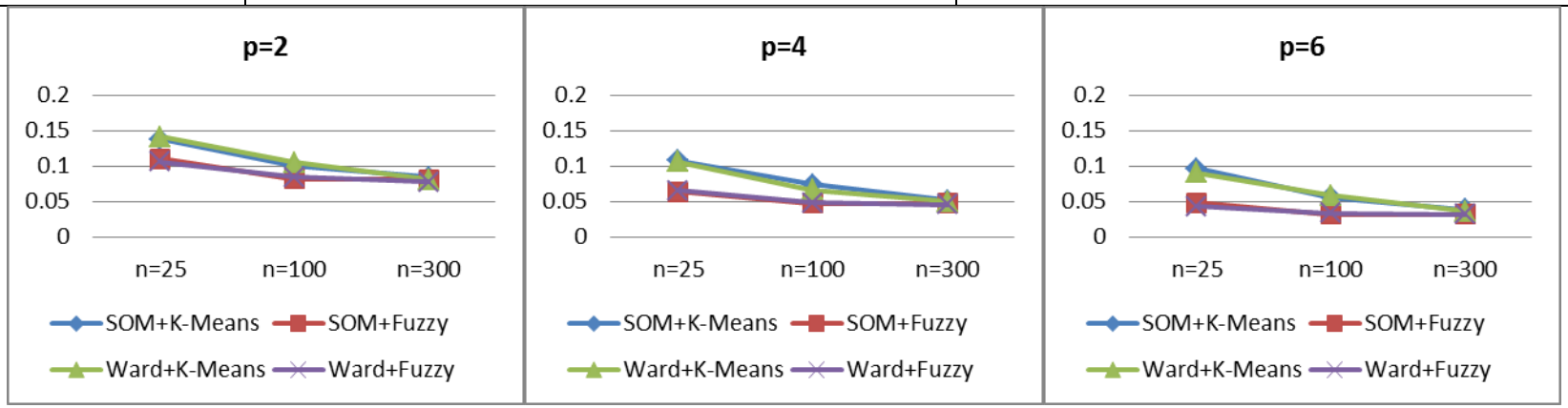
ตารางที่ 4.9 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.3628	0.1225	0.0862	0.2707	0.1191	0.0670	0.0798	0.1115	0.0952	0.0363	0.1050	0.0870
		100	0.2850	0.0966	0.0539	0.2734	0.0911	0.0520	0.0725	0.0989	0.0537	0.0236	0.0909	0.0489
		300	0.4291	0.1158	0.0979	0.2757	0.0770	0.0505	0.0311	0.0958	0.0875	0.0278	0.0801	0.0371
	4	25	0.3524	0.1173	0.0707	0.2467	0.1089	0.0681	0.0655	0.1172	0.0616	0.0141	0.1126	0.0473
		100	0.2434	0.0647	0.0474	0.2542	0.0763	0.0497	0.0522	0.1130	0.0148	0.0081	0.1088	0.0135
		300	0.4434	0.0510	0.0461	0.2433	0.0493	0.0454	0.0334	0.1229	0.1176	0.0079	0.1091	0.0290
	6	25	0.3544	0.0862	0.0121	0.2419	0.0947	0.0116	0.0703	0.1119	0.0447	0.0057	0.1200	0.0419
		100	0.2181	0.0937	0.0057	0.2515	0.0931	0.0035	0.0542	0.1166	0.0256	0.0040	0.1246	0.0056
		300	0.4568	0.1562	0.1235	0.2506	0.0836	0.0030	0.0303	0.1311	0.1339	0.0023	0.1251	0.0035



ตารางที่ 4.9 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.3601	0.1384	0.1107	0.2848	0.1417	0.1057	0.0788	0.1032	0.0921	0.0361	0.0914	0.0850
		100	0.2432	0.1003	0.0816	0.2964	0.1056	0.0846	0.0535	0.0766	0.0420	0.0241	0.0694	0.0456
		300	0.4245	0.0845	0.0806	0.2947	0.0812	0.0777	0.0246	0.0548	0.0380	0.0188	0.0411	0.0305
	4	25	0.3462	0.1079	0.0633	0.2697	0.1061	0.0658	0.0836	0.1029	0.0689	0.0237	0.0967	0.0652
		100	0.2194	0.0740	0.0471	0.2776	0.0661	0.0479	0.0508	0.0724	0.0257	0.0152	0.0686	0.0258
		300	0.4279	0.0515	0.0475	0.2765	0.0498	0.0455	0.0204	0.0424	0.0176	0.0159	0.0290	0.0197
	6	25	0.3402	0.0968	0.0478	0.2588	0.0907	0.0427	0.0761	0.1033	0.0660	0.0190	0.0988	0.0541
		100	0.1966	0.0559	0.0314	0.2700	0.0584	0.0332	0.0483	0.0868	0.0178	0.0126	0.0748	0.0194
		300	0.4294	0.0389	0.0315	0.2683	0.0363	0.0313	0.0205	0.0343	0.0123	0.0083	0.0417	0.0137



จากตารางที่ 4.7-4.9 ซึ่งแสดงผลการเปรียบเทียบค่า Wilk' s Lambda ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ตามลำดับ ของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 4 กลุ่ม พบว่า การจัดกลุ่มแบบ 2 ขั้นตอน ด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี Ward ร่วมกับ Fuzzy C-Means มีประสิทธิภาพดีกว่าวิธี SOM ร่วมกับ K-Means และวิธี Ward ร่วมกับ K-Means ในทุกกรณีสำหรับทุกขนาดตัวอย่าง และทุกจำนวนตัวแปร โดยให้ค่า Wilk's Lambda, ค่า RMSSTD และค่า Misclassification Rate ต่ำที่สุด

กรณีที่ 1 ตัวแปรอิสระต่อกัน

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่จะลดลงเรื่อยๆ เมื่อจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ เมื่อขนาดตัวอย่างเพิ่มขึ้น (ยกเว้น จำนวนตัวแปรเท่ากับ 4)
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

กรณีที่ 2 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(-1,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น

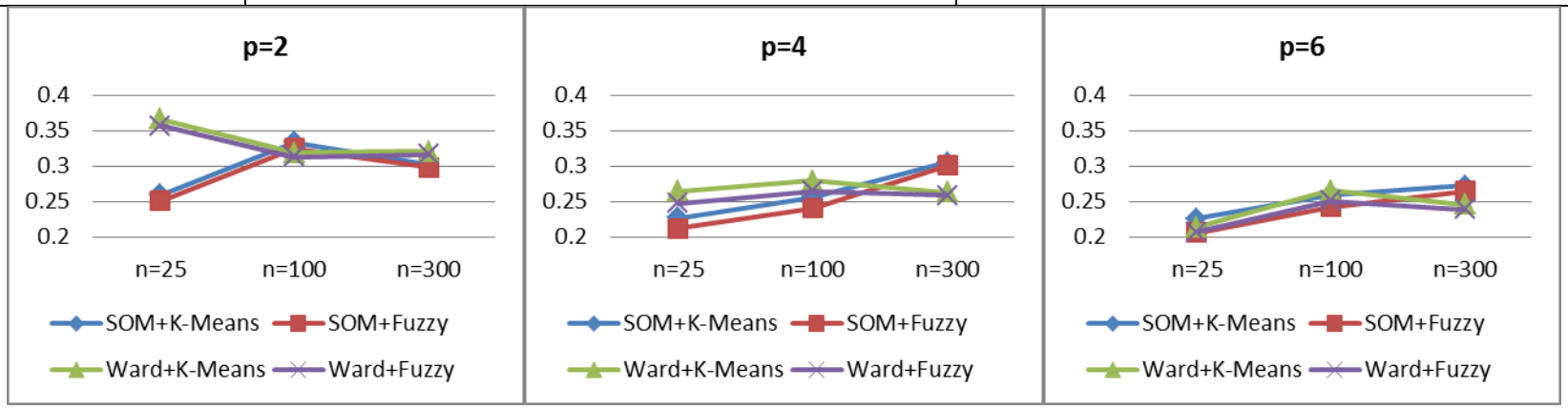
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

กรณีที่ 3 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(0,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

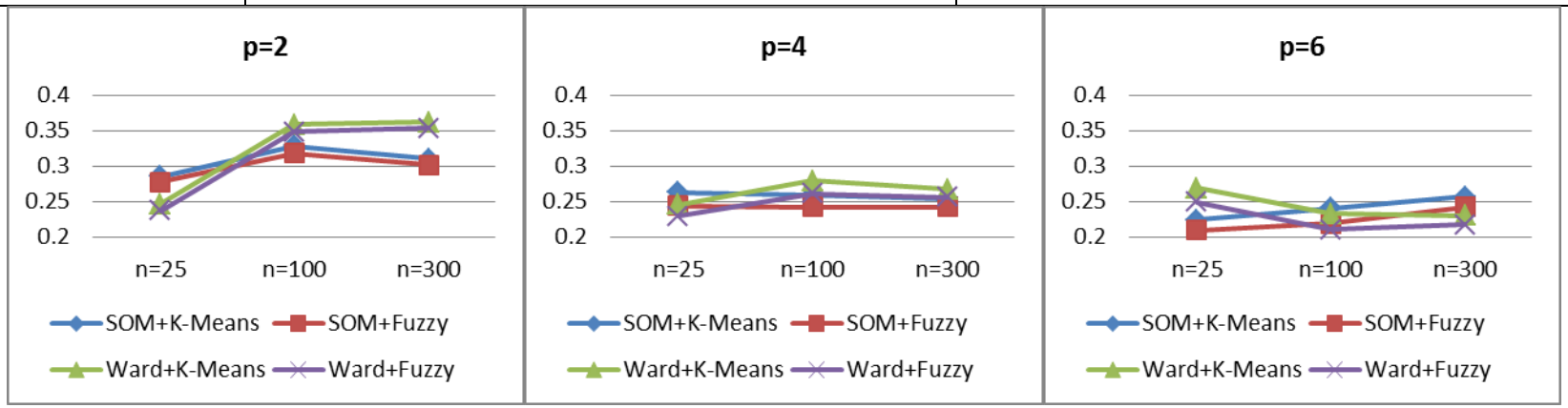
ตารางที่ 4.10 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.3957	0.2596	0.2507	0.4082	0.3664	0.3573	0.0690	0.0666	0.0639	0.0544	0.0691	0.0663
		100	0.4393	0.3337	0.3249	0.3657	0.3194	0.3127	0.0380	0.0355	0.0360	0.0355	0.0476	0.0407
		300	0.3756	0.3029	0.2981	0.4017	0.3204	0.3166	0.0200	0.0212	0.0198	0.0289	0.0267	0.0254
	4	25	0.3324	0.2268	0.2120	0.3064	0.2649	0.2472	0.0372	0.0467	0.0388	0.0370	0.0506	0.0422
		100	0.3847	0.2555	0.2400	0.3365	0.2791	0.2647	0.0275	0.0378	0.0238	0.0431	0.0389	0.0231
		300	0.4340	0.3055	0.3012	0.3504	0.2631	0.2585	0.0109	0.0282	0.0207	0.0173	0.0231	0.0158
	6	25	0.3023	0.2253	0.2056	0.2928	0.2137	0.1869	0.0299	0.0455	0.0314	0.0285	0.0509	0.0369
		100	0.3571	0.2593	0.2419	0.3383	0.2658	0.2504	0.0316	0.0346	0.0160	0.0165	0.0391	0.0149
		300	0.3610	0.2719	0.2646	0.3486	0.2453	0.2383	0.0085	0.0319	0.0184	0.0230	0.0211	0.0084



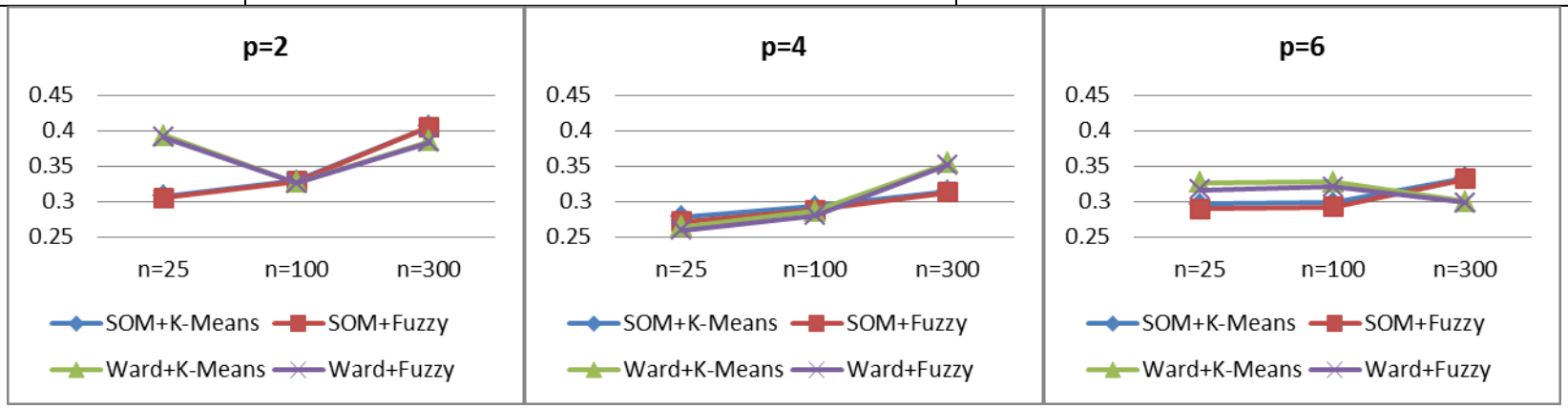
ตารางที่ 4.10 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.3804	0.2857	0.2778	0.3047	0.2468	0.2373	0.0597	0.0577	0.0540	0.0832	0.0772	0.0744
		100	0.3854	0.3287	0.3184	0.4032	0.3588	0.3486	0.0545	0.0573	0.0610	0.0504	0.0432	0.0492
		300	0.3563	0.3106	0.3017	0.4071	0.3624	0.3540	0.0398	0.0397	0.0447	0.0438	0.0366	0.0450
	4	25	0.3459	0.2630	0.2440	0.3156	0.2452	0.2289	0.0362	0.0427	0.0470	0.0446	0.0542	0.0465
		100	0.3530	0.2597	0.2423	0.3639	0.2792	0.2607	0.0365	0.0446	0.0347	0.0400	0.0400	0.0374
		300	0.3810	0.2545	0.2418	0.3125	0.2673	0.2556	0.0190	0.0354	0.0317	0.0301	0.0369	0.0287
	6	25	0.3647	0.2235	0.1988	0.3156	0.2691	0.2487	0.0332	0.0553	0.0455	0.0366	0.0474	0.0361
		100	0.3815	0.2407	0.2185	0.3228	0.2328	0.2097	0.0303	0.0417	0.0225	0.0384	0.0376	0.0223
		300	0.3984	0.2566	0.2422	0.3326	0.2302	0.2167	0.0116	0.0332	0.0247	0.0265	0.0357	0.0201



ตารางที่ 4.10 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.3605	0.3077	0.3051	0.4326	0.3930	0.3904	0.0801	0.0668	0.0708	0.0552	0.0643	0.0602
		100	0.4054	0.3296	0.3281	0.3890	0.3268	0.3261	0.0506	0.0441	0.0469	0.0448	0.0393	0.0438
		300	0.4653	0.4066	0.4054	0.4522	0.3852	0.3836	0.0327	0.0324	0.0323	0.0267	0.0278	0.0285
	4	25	0.3450	0.2783	0.2714	0.3283	0.2647	0.2588	0.0704	0.0506	0.0465	0.0369	0.0525	0.0451
		100	0.3553	0.2927	0.2880	0.3517	0.2861	0.2797	0.0412	0.0346	0.0358	0.0319	0.0379	0.0399
		300	0.3936	0.3146	0.3132	0.4227	0.3539	0.3520	0.0249	0.0256	0.0243	0.0274	0.0245	0.0228
	6	25	0.3767	0.2963	0.2896	0.3984	0.3263	0.3159	0.0617	0.0517	0.0486	0.0369	0.0438	0.0454
		100	0.3867	0.2981	0.2917	0.4077	0.3276	0.3207	0.0249	0.0319	0.0305	0.0416	0.0334	0.0331
		300	0.4298	0.3335	0.3318	0.3776	0.2997	0.2979	0.0205	0.0209	0.0205	0.0246	0.0202	0.0183

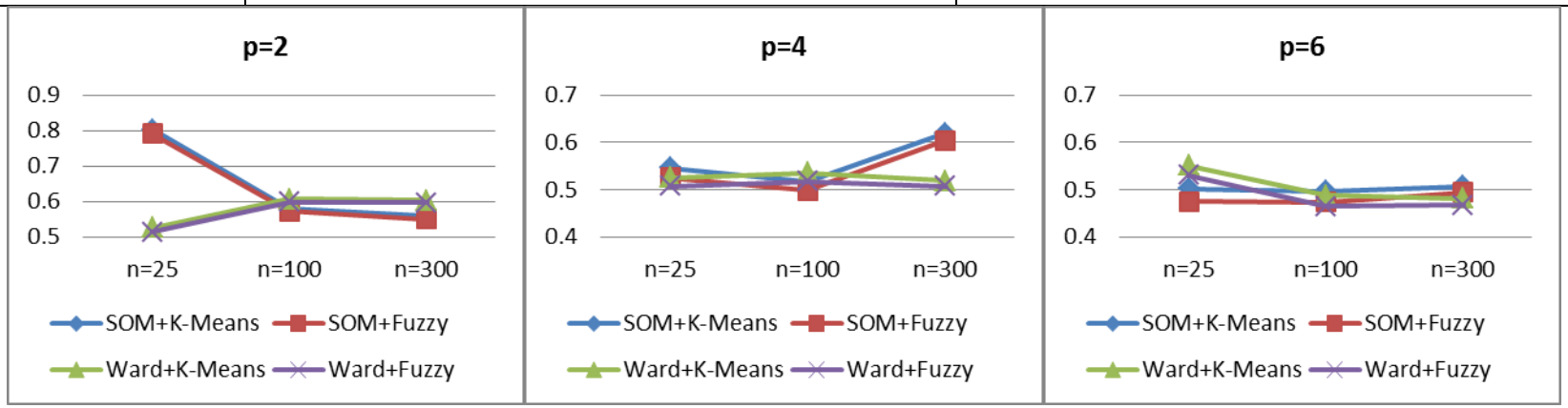


ตารางที่ 4.11 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.6721	0.5415	0.5320	0.6923	0.6404	0.6326	0.0542	0.0576	0.0570	0.0480	0.0645	0.0626
		100	0.6525	0.5856	0.5779	0.6371	0.5730	0.5670	0.0288	0.0298	0.0306	0.0280	0.0389	0.0347
		300	0.6158	0.5527	0.5483	0.6385	0.5685	0.5653	0.0224	0.0263	0.0246	0.0210	0.0217	0.0208
	4	25	0.6054	0.5062	0.4895	0.6245	0.5453	0.5272	0.0345	0.0522	0.0446	0.0339	0.0544	0.0459
		100	0.6257	0.5119	0.4969	0.5807	0.5354	0.5220	0.0222	0.0360	0.0238	0.0349	0.0376	0.0233
		300	0.6620	0.5494	0.5456	0.5893	0.5151	0.5108	0.0083	0.0246	0.0183	0.0139	0.0224	0.0159
	6	25	0.5871	0.5050	0.4835	0.5556	0.4903	0.4597	0.0284	0.0502	0.0358	0.0282	0.0583	0.0432
		100	0.6063	0.5156	0.4991	0.5819	0.5226	0.5080	0.0266	0.0342	0.0167	0.0137	0.0380	0.0155
		300	0.6138	0.5233	0.5167	0.5779	0.4972	0.4905	0.0071	0.0302	0.0175	0.0171	0.0198	0.0084

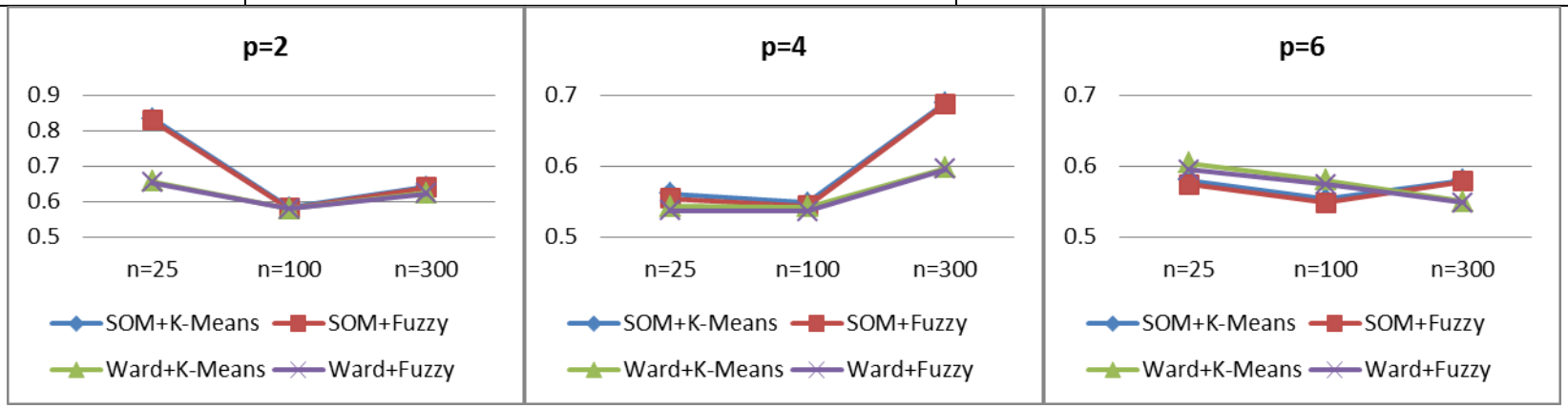
ตารางที่ 4.11 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.9438	0.8036	0.7919	0.5671	0.5248	0.5137	0.0707	0.0771	0.0738	0.0732	0.0728	0.0708
		100	0.6089	0.5809	0.5713	0.6458	0.6072	0.5982	0.0444	0.0528	0.0567	0.0394	0.0364	0.0415
		300	0.5874	0.5589	0.5504	0.6368	0.6040	0.5966	0.0316	0.0338	0.0387	0.0351	0.0313	0.0391
	4	25	0.6279	0.5461	0.5260	0.6172	0.5252	0.5074	0.0326	0.0458	0.0512	0.0369	0.0525	0.0451
		100	0.5479	0.5158	0.4984	0.5691	0.5351	0.5173	0.0294	0.0418	0.0334	0.0349	0.0403	0.0379
		300	0.7202	0.6194	0.6040	0.5612	0.5184	0.5072	0.0154	0.0347	0.0316	0.0248	0.0332	0.0263
	6	25	0.6450	0.5026	0.4747	0.6674	0.5512	0.5299	0.0299	0.0609	0.0509	0.0346	0.0539	0.0420
		100	0.6167	0.4964	0.4739	0.5883	0.4881	0.4644	0.0247	0.0418	0.0237	0.0314	0.0379	0.0235
		300	0.6343	0.5077	0.4937	0.5141	0.4808	0.4673	0.0092	0.0311	0.0235	0.0214	0.034	0.0199



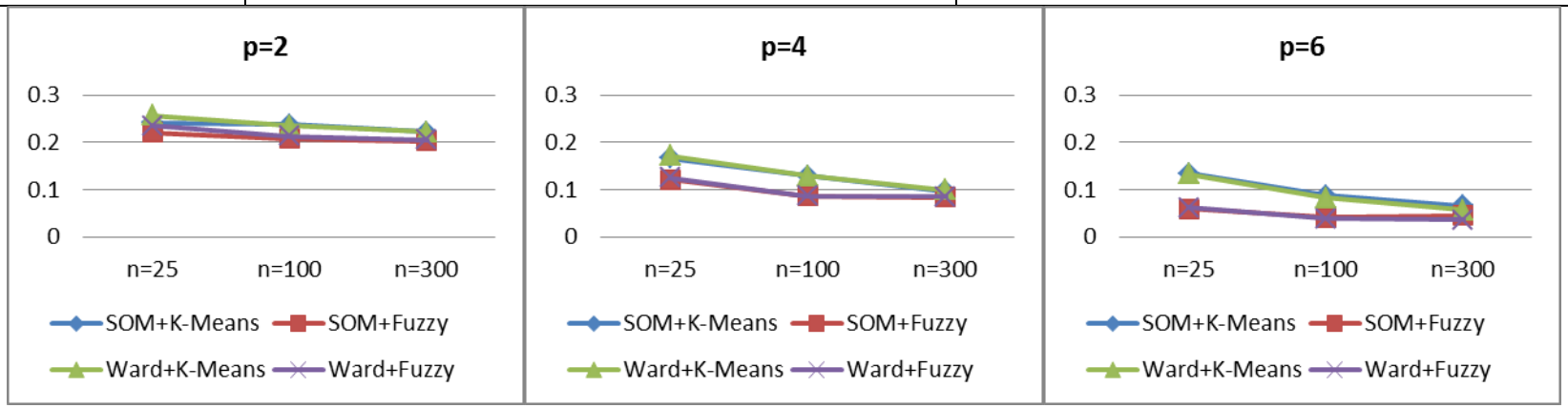
ตารางที่ 4.11 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.9021	0.8341	0.8304	0.6808	0.6559	0.6536	0.0676	0.0663	0.0690	0.0459	0.0584	0.0550
		100	0.6453	0.5820	0.5806	0.6265	0.5790	0.5784	0.0403	0.0382	0.0405	0.0369	0.0359	0.0394
		300	0.6852	0.6405	0.6395	0.6721	0.6231	0.6217	0.0267	0.0281	0.0280	0.0203	0.0231	0.0237
	4	25	0.6247	0.5607	0.5539	0.5900	0.5432	0.5373	0.0601	0.0485	0.0470	0.0348	0.1018	0.0880
		100	0.6043	0.5484	0.5440	0.5949	0.5419	0.5358	0.0334	0.0311	0.0324	0.0268	0.0352	0.0368
		300	0.7720	0.6900	0.6884	0.6500	0.5975	0.5959	0.0203	0.0233	0.0221	0.0210	0.0213	0.0198
	6	25	0.6548	0.5804	0.5738	0.6528	0.6043	0.5948	0.0539	0.0522	0.0498	0.0313	0.0400	0.0418
		100	0.6307	0.5536	0.5477	0.6416	0.5803	0.5742	0.0199	0.0292	0.0281	0.0331	0.0300	0.0301
		300	0.6588	0.5801	0.5786	0.6133	0.5498	0.5482	0.0164	0.0188	0.0186	0.0190	0.0182	0.0166



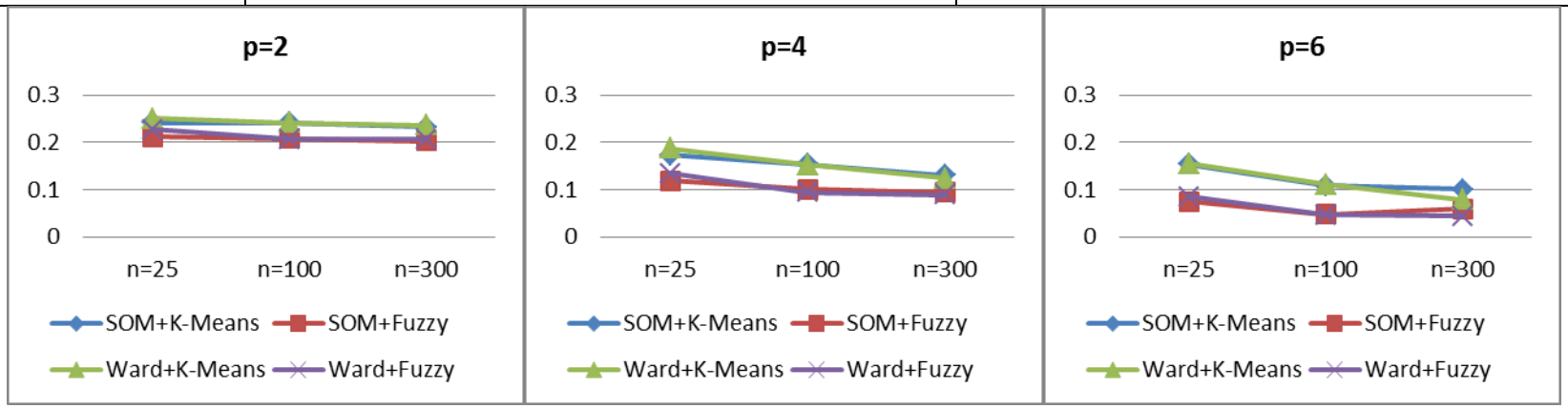
ตารางที่ 4.12 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.3948	0.2414	0.2216	0.3405	0.2581	0.2358	0.0584	0.0754	0.0761	0.0426	0.0950	0.0935
		100	0.4215	0.2377	0.2077	0.3600	0.2360	0.2132	0.0517	0.0711	0.0611	0.0315	0.0718	0.0654
		300	0.4276	0.2220	0.2026	0.3437	0.2228	0.2056	0.0161	0.0533	0.0481	0.0245	0.0548	0.0466
	4	25	0.3760	0.1670	0.1214	0.2940	0.1720	0.1253	0.0643	0.0992	0.0799	0.0362	0.1056	0.0865
		100	0.3319	0.1298	0.0870	0.3021	0.1289	0.0860	0.0588	0.0946	0.0345	0.0293	0.1022	0.0369
		300	0.4289	0.0964	0.0834	0.2965	0.0997	0.0849	0.0213	0.0601	0.0416	0.0165	0.0513	0.0233
	6	25	0.3548	0.1338	0.0597	0.2724	0.1329	0.0622	0.0707	0.1128	0.0649	0.0289	0.1040	0.0641
		100	0.3329	0.0876	0.0405	0.2727	0.0821	0.0381	0.0627	0.0931	0.0231	0.0138	0.0964	0.0230
		300	0.4297	0.0661	0.0446	0.2707	0.0564	0.0363	0.0239	0.0791	0.0415	0.0082	0.0581	0.0115



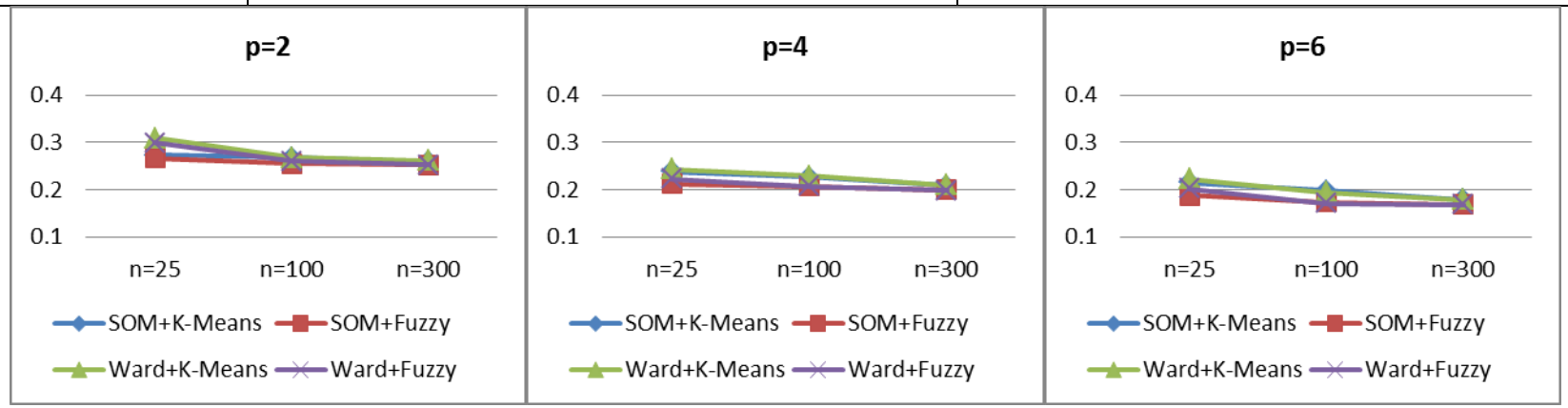
ตารางที่ 4.12 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.3648	0.2423	0.2126	0.3351	0.2517	0.2278	0.0670	0.0838	0.0761	0.0451	0.0951	0.0962
		100	0.3747	0.2425	0.2083	0.3496	0.2423	0.2072	0.0637	0.0767	0.0789	0.0437	0.0825	0.0829
		300	0.4260	0.2332	0.2035	0.3494	0.2374	0.2067	0.0235	0.0793	0.0853	0.0442	0.0641	0.0773
	4	25	0.3592	0.1734	0.1192	0.2927	0.1879	0.1342	0.0696	0.0996	0.0936	0.0476	0.1157	0.0942
		100	0.3980	0.1541	0.1005	0.2972	0.1528	0.0937	0.0567	0.1025	0.0631	0.0403	0.0924	0.0651
		300	0.4233	0.1307	0.0946	0.2741	0.1244	0.0891	0.0278	0.0830	0.0506	0.0323	0.0938	0.0519
	6	25	0.3820	0.1534	0.0746	0.3951	0.1550	0.0858	0.0624	0.1035	0.0714	0.0390	0.1137	0.0744
		100	0.3658	0.1079	0.0476	0.2714	0.1115	0.0461	0.0617	0.1023	0.0382	0.0245	0.1031	0.0346
		300	0.4284	0.1006	0.0593	0.2677	0.0791	0.0436	0.0259	0.0900	0.0614	0.0231	0.0911	0.0333



ตารางที่ 4.12 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 4 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.3916	0.2738	0.2672	0.3578	0.3102	0.2993	0.0668	0.0832	0.0877	0.0387	0.0868	0.0829
		100	0.3670	0.2679	0.2546	0.3776	0.2685	0.2600	0.0667	0.0651	0.0589	0.0313	0.0554	0.0605
		300	0.4352	0.2597	0.2525	0.3786	0.2619	0.2532	0.0157	0.0459	0.0451	0.0248	0.0506	0.0512
	4	25	0.3762	0.2388	0.2118	0.3365	0.2434	0.2232	0.0645	0.0838	0.0823	0.0348	0.1018	0.0880
		100	0.3687	0.2267	0.2069	0.3555	0.2288	0.2067	0.0506	0.0725	0.0582	0.0277	0.0598	0.0541
		300	0.4304	0.2079	0.1992	0.3566	0.2092	0.1981	0.0154	0.0471	0.0384	0.0249	0.0483	0.0416
	6	25	0.3757	0.2153	0.1879	0.3274	0.2222	0.2014	0.0705	0.0853	0.0764	0.0415	0.0914	0.0850
		100	0.3497	0.1986	0.1733	0.3415	0.1942	0.1715	0.0521	0.0665	0.0478	0.0280	0.0665	0.0473
		300	0.4289	0.1788	0.1687	0.3427	0.1785	0.1685	0.0144	0.0412	0.0298	0.0245	0.0354	0.0285



จากตารางที่ 4.10-4.12 ซึ่งแสดงผลการเปรียบเทียบค่า Wilk' s Lambda ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ตามลำดับ ของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองเป็น 4 กลุ่ม พบว่า การจัดกลุ่มแบบ 2 ขั้นตอน ด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี Ward ร่วมกับ Fuzzy C-Means มีประสิทธิภาพดีกว่าวิธี SOM ร่วมกับ K-Means และวิธี Ward ร่วมกับ K-Means ในทุกกรณีสำหรับทุกขนาดตัวอย่าง และทุกจำนวนตัวแปร โดยให้ค่า Wilk's Lambda, ค่า RMSSTD และค่า Misclassification Rate ต่ำที่สุด

กรณีที่ 1 ตัวแปรอิสระต่อกัน

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่คงที่ ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่คงที่ ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ เมื่อจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ เมื่อขนาดตัวอย่างเพิ่มขึ้น

กรณีที่ 2 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(-1,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น

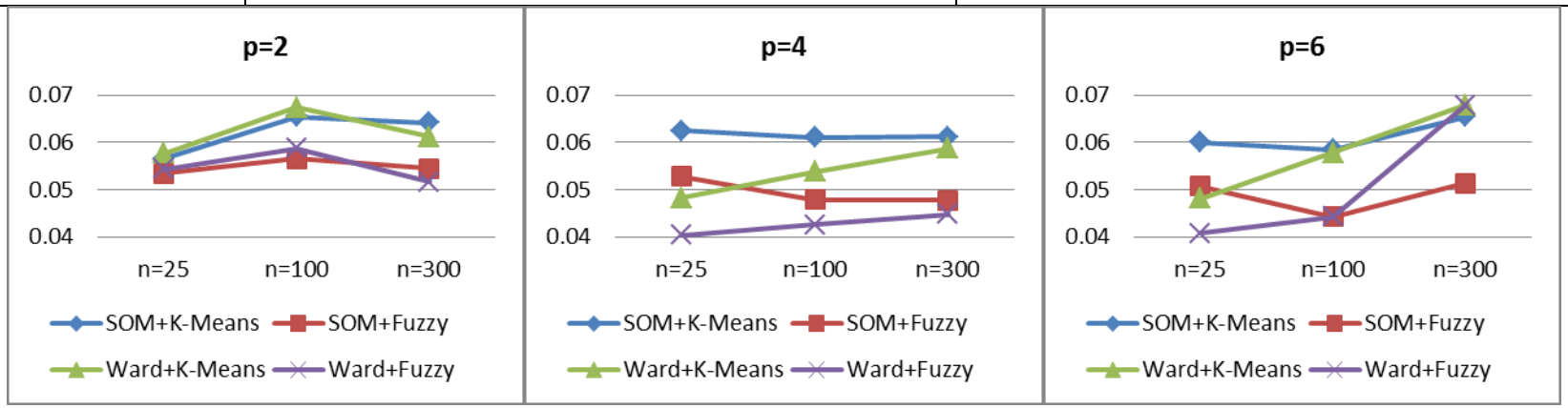
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ เมื่อขนาดตัวอย่างเพิ่มขึ้น

กรณีที่ 3 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(0,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ เมื่อจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

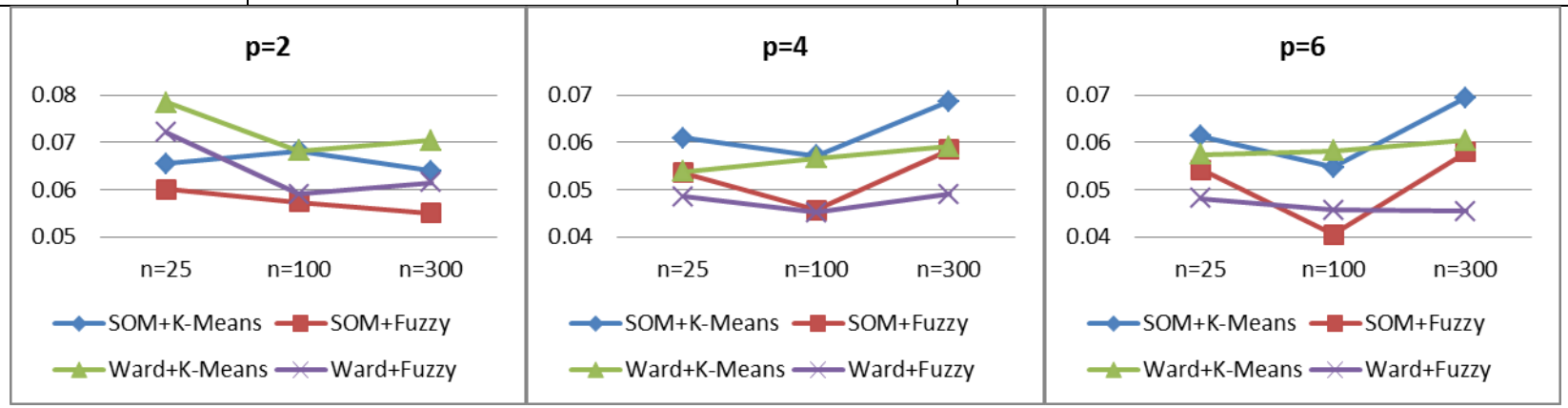
ตารางที่ 4.13 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.1658	0.0565	0.0534	0.1744	0.0577	0.0543	0.0395	0.0219	0.0172	0.0308	0.0439	0.0335
		100	0.1072	0.0655	0.0565	0.1241	0.0674	0.0587	0.0137	0.0165	0.0146	0.0273	0.0238	0.0223
		300	0.1389	0.0642	0.0545	0.1224	0.0612	0.0516	0.0094	0.0168	0.0124	0.0176	0.0174	0.0149
	4	25	0.1559	0.0625	0.0528	0.1191	0.0482	0.0404	0.0408	0.0166	0.0185	0.0080	0.0158	0.0165
		100	0.0824	0.0611	0.0479	0.1550	0.0538	0.0426	0.0081	0.0163	0.0150	0.0051	0.0185	0.0117
		300	0.1782	0.0612	0.0477	0.1425	0.0588	0.0448	0.0117	0.0186	0.0143	0.0103	0.0183	0.0102
	6	25	0.1455	0.0601	0.0507	0.1353	0.0481	0.0407	0.0314	0.0176	0.0173	0.0068	0.0164	0.0165
		100	0.0760	0.0585	0.0443	0.1377	0.0579	0.0442	0.0088	0.0164	0.0158	0.0067	0.0184	0.0139
		300	0.1795	0.0655	0.0514	0.1371	0.0679	0.0679	0.0110	0.0178	0.0162	0.0050	0.0170	0.0129



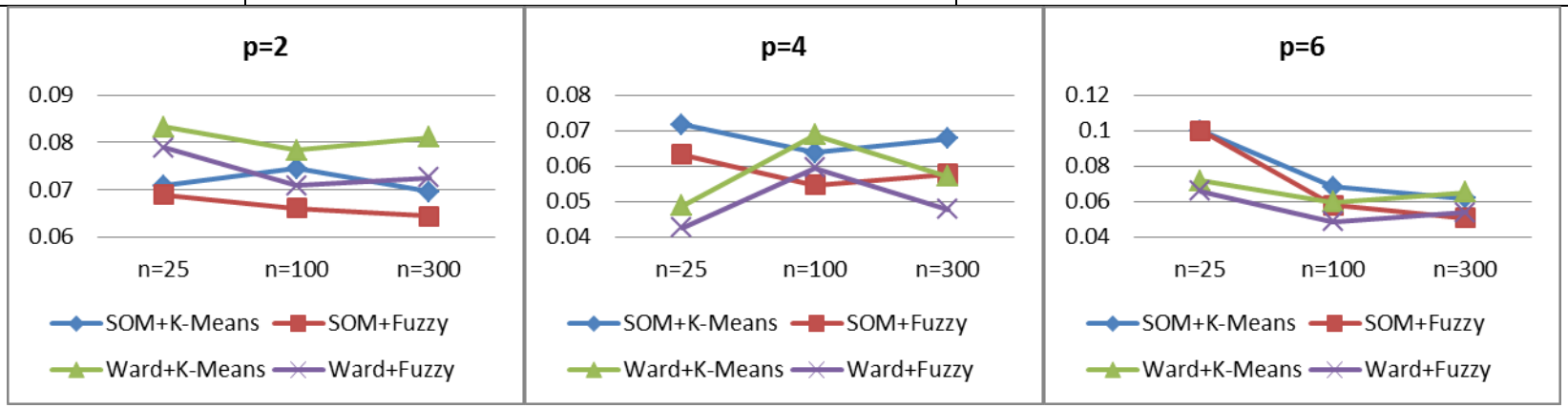
ตารางที่ 4.13 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.1433	0.0655	0.0601	0.1320	0.0785	0.0721	0.0371	0.0257	0.0245	0.0207	0.0221	0.0223
		100	0.0996	0.0683	0.0573	0.1409	0.0683	0.0590	0.0144	0.0183	0.0181	0.0219	0.0252	0.0213
		300	0.1746	0.0639	0.0550	0.1508	0.0705	0.0615	0.0166	0.0209	0.0151	0.0139	0.0156	0.0132
	4	25	0.1253	0.0610	0.0536	0.1293	0.0538	0.0485	0.0393	0.0246	0.0219	0.0073	0.0181	0.0157
		100	0.0978	0.0572	0.0457	0.1285	0.0567	0.0451	0.0087	0.0197	0.0145	0.0163	0.0166	0.0157
		300	0.1513	0.0688	0.0585	0.1303	0.0591	0.0490	0.0108	0.0186	0.0170	0.0064	0.0196	0.0140
	6	25	0.1430	0.0613	0.0542	0.1351	0.0574	0.0481	0.0337	0.0189	0.0173	0.0213	0.0224	0.0169
		100	0.0803	0.0548	0.0405	0.1281	0.0582	0.0457	0.0083	0.0167	0.0158	0.0049	0.0174	0.0128
		300	0.1428	0.0695	0.0580	0.1354	0.0604	0.0454	0.0064	0.0287	0.0209	0.0054	0.0171	0.0133



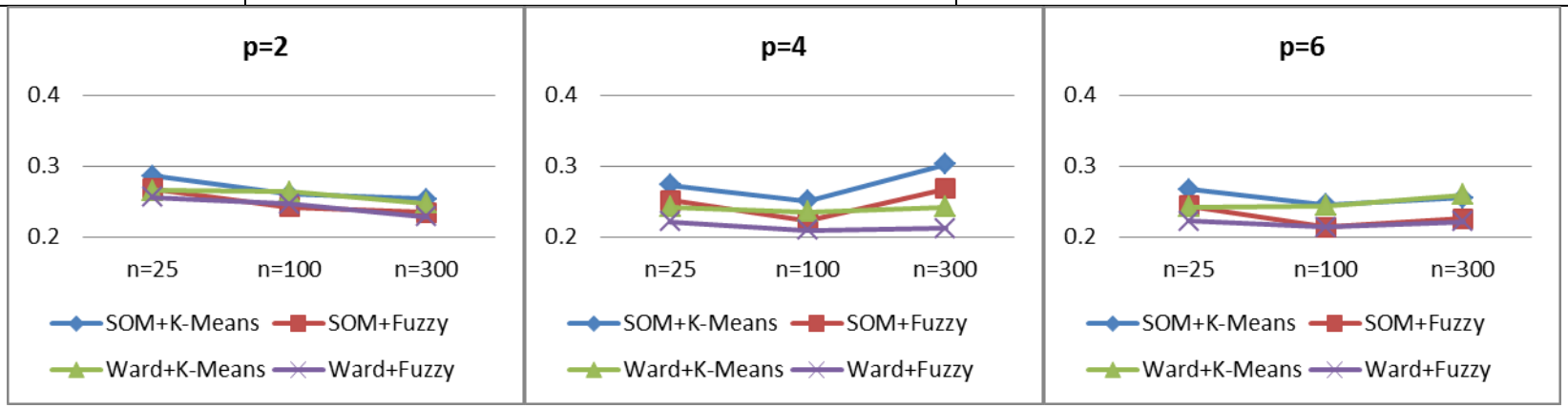
ตารางที่ 4.13 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.1518	0.0708	0.0689	0.1597	0.0833	0.0789	0.0468	0.0220	0.0201	0.0217	0.0261	0.0227
		100	0.0911	0.0745	0.0661	0.1547	0.0784	0.0709	0.0101	0.0165	0.0164	0.0152	0.0182	0.0151
		300	0.1429	0.0696	0.0644	0.1688	0.0811	0.0725	0.0145	0.0174	0.0117	0.0152	0.0150	0.0101
	4	25	0.1509	0.0718	0.0632	0.1154	0.0487	0.0425	0.0319	0.0216	0.0210	0.0108	0.0196	0.0191
		100	0.0882	0.0638	0.0546	0.1413	0.0688	0.0595	0.0101	0.0151	0.0128	0.0153	0.0172	0.0130
		300	0.1555	0.0678	0.0577	0.1348	0.0571	0.0478	0.0107	0.0174	0.0084	0.0112	0.0176	0.0098
	6	25	0.1925	0.1001	0.1001	0.1415	0.0716	0.0658	0.0291	0.0208	0.0219	0.0119	0.0205	0.0184
		100	0.0889	0.0683	0.0579	0.1299	0.0596	0.0487	0.0085	0.0191	0.0117	0.0106	0.0173	0.0126
		300	0.1554	0.0616	0.0507	0.1428	0.0650	0.0536	0.0070	0.0168	0.0102	0.0111	0.0179	0.0053



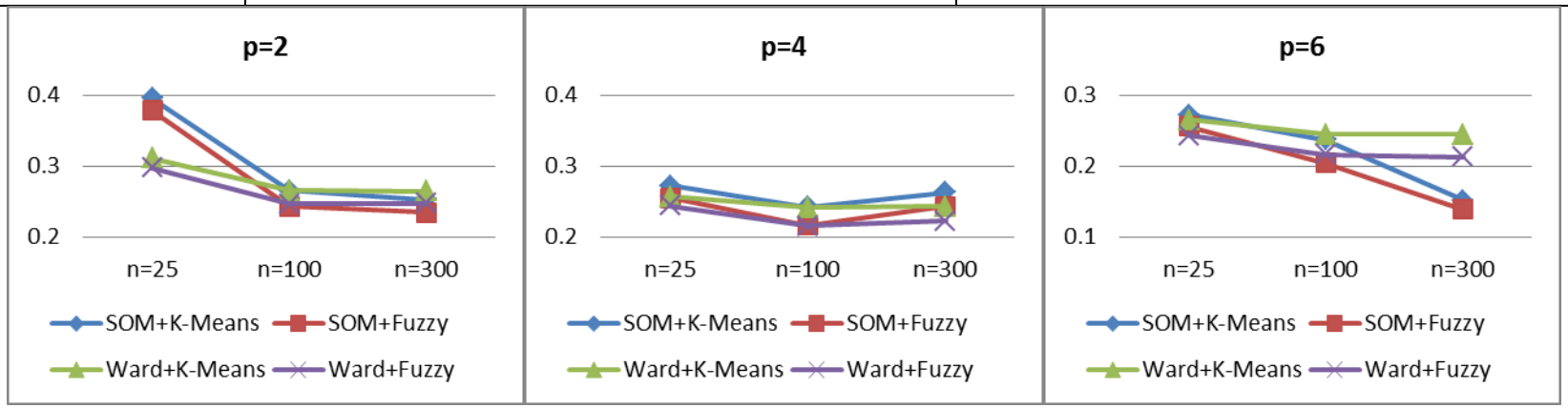
ตารางที่ 4.14 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.4371	0.2862	0.2676	0.4158	0.2652	0.2558	0.0561	0.0445	0.0379	0.0362	0.0585	0.0464
		100	0.3392	0.2606	0.2423	0.4015	0.2641	0.2467	0.0223	0.0350	0.0320	0.0274	0.0298	0.0281
		300	0.3657	0.2533	0.2344	0.3565	0.2470	0.2280	0.0129	0.0332	0.0249	0.0172	0.0198	0.0172
	4	25	0.4410	0.2735	0.2512	0.3467	0.2420	0.2209	0.0592	0.0409	0.0454	0.0129	0.0430	0.0451
		100	0.2942	0.2507	0.2221	0.3346	0.2350	0.2096	0.0145	0.0345	0.0329	0.0076	0.0368	0.0236
		300	0.4724	0.3025	0.2681	0.3419	0.2418	0.2121	0.0147	0.0390	0.0306	0.0119	0.0386	0.0220
	6	25	0.4217	0.2674	0.2442	0.3785	0.2421	0.2224	0.0454	0.0359	0.0369	0.0126	0.0435	0.0440
		100	0.4023	0.2453	0.2135	0.3691	0.2436	0.2136	0.0173	0.0363	0.0359	0.0068	0.0400	0.0316
		300	0.4349	0.2550	0.2256	0.3457	0.2597	0.2211	0.0138	0.0384	0.0357	0.0069	0.0359	0.0273



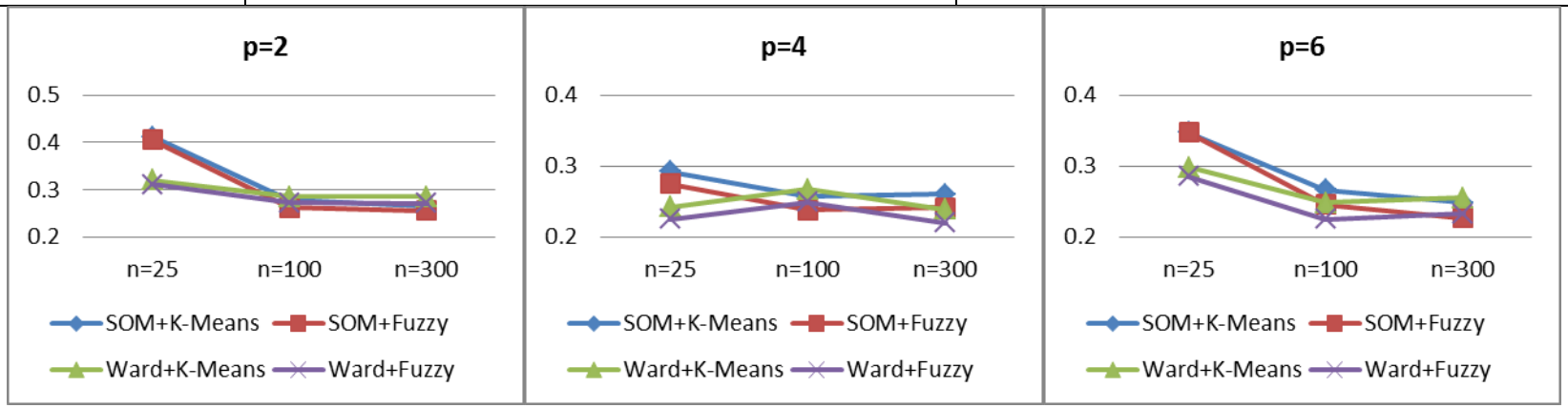
ตารางที่ 4.14 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.4036	0.3963	0.3787	0.3823	0.3108	0.2976	0.0553	0.0569	0.0565	0.0292	0.0427	0.0444
		100	0.3290	0.2655	0.2429	0.3576	0.2654	0.2462	0.0219	0.0347	0.0353	0.0291	0.0432	0.0392
		300	0.4228	0.2524	0.2341	0.3580	0.2646	0.2473	0.0197	0.0377	0.0306	0.0198	0.0311	0.0283
	4	25	0.4262	0.2726	0.2548	0.4025	0.2565	0.2439	0.0616	0.0539	0.0513	0.0112	0.0402	0.0369
		100	0.3206	0.2422	0.2163	0.3398	0.2415	0.2154	0.0141	0.0410	0.0321	0.0223	0.0358	0.0350
		300	0.3818	0.2624	0.2430	0.3414	0.2429	0.2225	0.0143	0.0373	0.0345	0.0091	0.0372	0.0275
	6	25	0.4229	0.2730	0.2562	0.3597	0.2657	0.2433	0.0514	0.0427	0.0432	0.0303	0.0501	0.0409
		100	0.2804	0.2373	0.2040	0.3389	0.2446	0.2166	0.0149	0.0359	0.0365	0.0067	0.0381	0.0292
		300	0.3807	0.1514	0.1382	0.3364	0.2450	0.2131	0.0085	0.0509	0.0426	0.0071	0.0353	0.0283



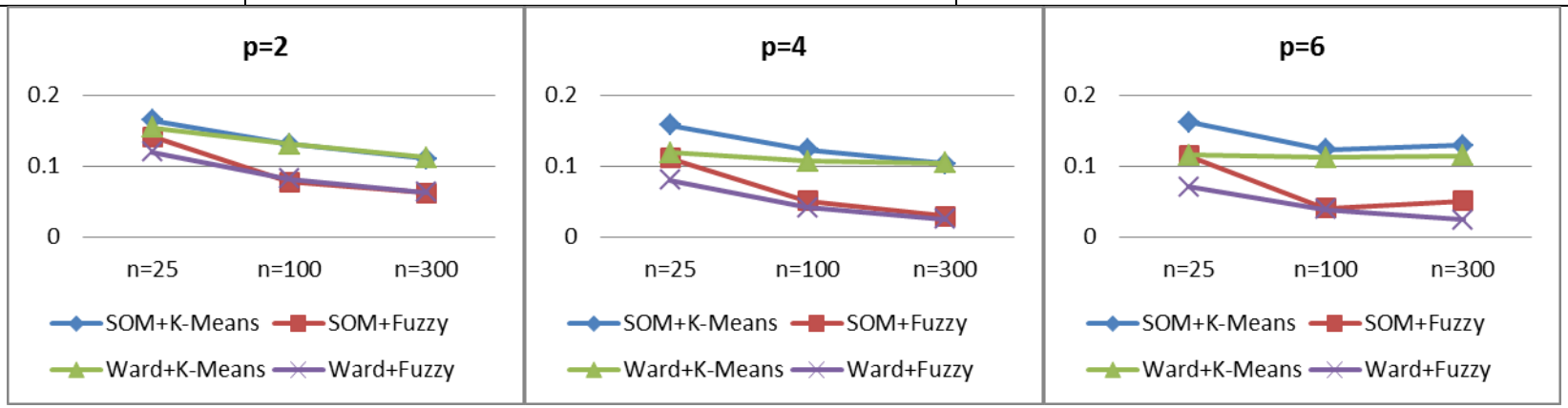
ตารางที่ 4.14 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.6047	0.4123	0.4061	0.4049	0.3198	0.3114	0.0687	0.0459	0.0432	0.0351	0.0576	0.0525
		100	0.3091	0.2780	0.2620	0.3825	0.2857	0.2718	0.0157	0.0300	0.0303	0.0202	0.0355	0.0300
		300	0.3809	0.2647	0.2550	0.3961	0.2856	0.2706	0.0170	0.0307	0.0213	0.0177	0.0273	0.0193
	4	25	0.4283	0.2924	0.2746	0.3340	0.2419	0.2248	0.0455	0.0410	0.0411	0.0190	0.0478	0.0500
		100	0.3043	0.2570	0.2380	0.3637	0.2672	0.2489	0.0169	0.0296	0.0261	0.0193	0.0335	0.0266
		300	0.3975	0.2606	0.2416	0.3499	0.2389	0.2199	0.0131	0.0324	0.0167	0.0136	0.0349	0.0200
	6	25	0.4798	0.3475	0.3474	0.3804	0.2979	0.2855	0.0420	0.0376	0.0406	0.0205	0.0467	0.0449
		100	0.3056	0.2660	0.2454	0.3463	0.2482	0.2248	0.0140	0.0372	0.0242	0.0141	0.0355	0.0266
		300	0.3975	0.2479	0.2264	0.3614	0.2552	0.2329	0.0084	0.0310	0.0193	0.0131	0.0359	0.0118



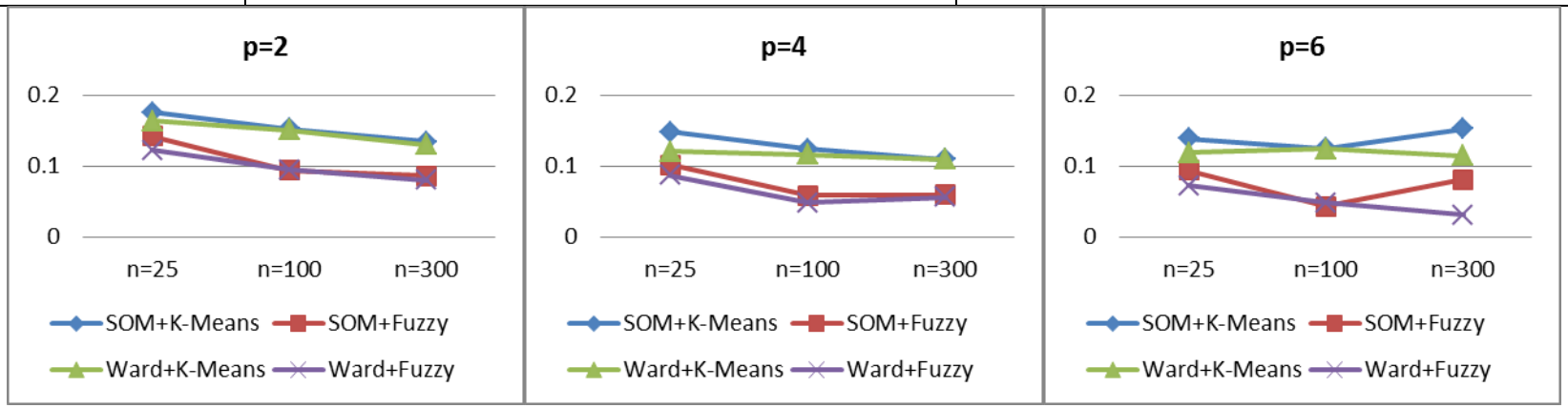
ตารางที่ 4.15 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.3776	0.1649	0.1415	0.3655	0.1544	0.1198	0.0454	0.1002	0.0838	0.0269	0.0705	0.0600
		100	0.2871	0.1309	0.0772	0.3395	0.1307	0.0818	0.0363	0.0824	0.0673	0.0214	0.0503	0.0497
		300	0.4478	0.1105	0.0622	0.4514	0.1119	0.0625	0.0291	0.0765	0.0489	0.0188	0.0487	0.0471
	4	25	0.3880	0.1576	0.1112	0.3240	0.1194	0.0799	0.0520	0.0945	0.0866	0.0077	0.0852	0.0800
		100	0.2481	0.1229	0.0507	0.3235	0.1071	0.0414	0.0375	0.0375	0.0679	0.0042	0.0854	0.0519
		300	0.4490	0.1036	0.0287	0.3562	0.1044	0.0250	0.0255	0.0967	0.0666	0.0026	0.0890	0.0520
	6	25	0.3756	0.1621	0.1150	0.3741	0.1154	0.0702	0.0511	0.0930	0.0874	0.0028	0.0863	0.0798
		100	0.2067	0.1235	0.0401	0.3505	0.1125	0.0386	0.0340	0.0884	0.0684	0.0014	0.0957	0.0719
		300	0.4628	0.1290	0.0510	0.3341	0.1149	0.0238	0.0242	0.0940	0.0766	0.0014	0.0922	0.0584



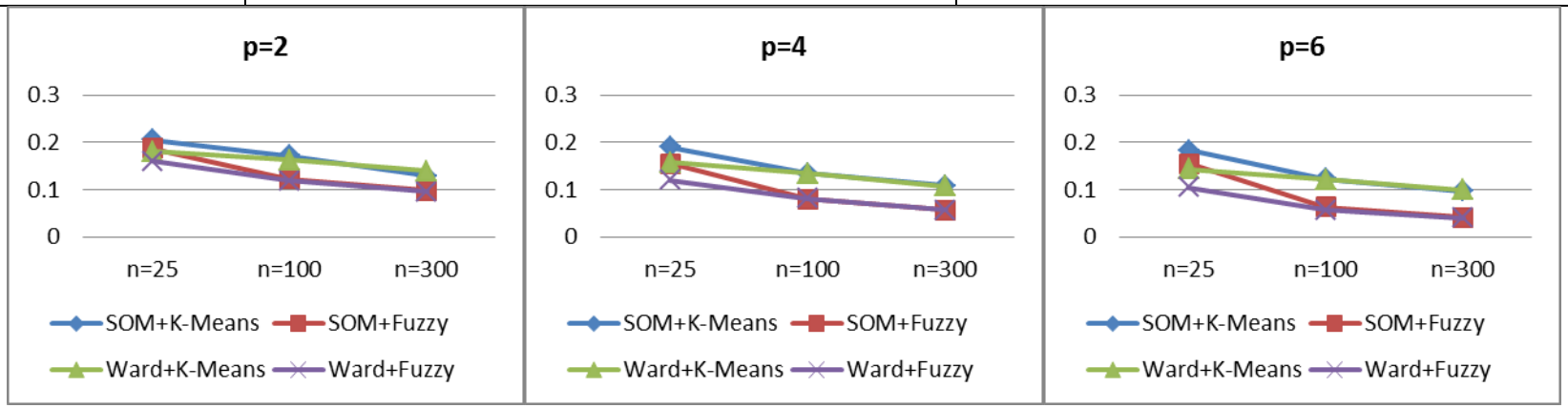
ตารางที่ 4.15 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.3748	0.1760	0.1424	0.3314	0.1639	0.1223	0.0501	0.0960	0.0893	0.0243	0.0771	0.0881
		100	0.2791	0.1527	0.0938	0.3470	0.1511	0.0950	0.0334	0.0860	0.0714	0.0252	0.0829	0.0740
		300	0.4621	0.1347	0.0860	0.3673	0.1302	0.0800	0.0293	0.0858	0.0642	0.0188	0.0804	0.0557
	4	25	0.3796	0.1486	0.1014	0.3143	0.1205	0.0868	0.0517	0.1004	0.0888	0.0095	0.0937	0.0804
		100	0.2402	0.1243	0.0587	0.3055	0.1162	0.0482	0.0340	0.0958	0.0693	0.0094	0.0848	0.0765
		300	0.4336	0.1094	0.0595	0.3017	0.1097	0.0561	0.0289	0.0952	0.0824	0.0069	0.0954	0.0663
	6	25	0.3616	0.1390	0.0941	0.3111	0.1188	0.0719	0.0525	0.0970	0.0911	0.0043	0.0864	0.0859
		100	0.1993	0.1253	0.0426	0.3213	0.1246	0.0481	0.0328	0.0899	0.0746	0.0033	0.0948	0.0626
		300	0.4717	0.1529	0.0813	0.3139	0.1150	0.0308	0.0227	0.1253	0.1107	0.0020	0.0921	0.0632



ตารางที่ 4.15 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.3885	0.2059	0.1875	0.3572	0.1814	0.1604	0.0539	0.0888	0.0856	0.0252	0.0909	0.0878
		100	0.2586	0.1719	0.1228	0.3607	0.1635	0.1189	0.0375	0.0779	0.0643	0.0181	0.0810	0.0699
		300	0.4625	0.1294	0.0987	0.3740	0.1402	0.0959	0.0282	0.0817	0.0490	0.0138	0.0731	0.0400
	4	25	0.3857	0.1896	0.1550	0.3452	0.1582	0.1198	0.0482	0.0938	0.0820	0.0200	0.0861	0.0838
		100	0.2399	0.1344	0.0801	0.3456	0.1345	0.0811	0.0359	0.0747	0.0622	0.0133	0.0778	0.0622
		300	0.4588	0.1083	0.0564	0.3583	0.1073	0.0561	0.0233	0.0822	0.0314	0.0112	0.0838	0.0407
	6	25	0.3861	0.1842	0.1547	0.3380	0.1441	0.1043	0.0477	0.0846	0.0799	0.0209	0.0837	0.0778
		100	0.2277	0.1230	0.0625	0.3373	0.1213	0.0572	0.0345	0.0936	0.0564	0.0112	0.0843	0.0577
		300	0.4598	0.0963	0.0411	0.3501	0.0995	0.0400	0.0181	0.0817	0.0457	0.0085	0.0838	0.0211



จากตารางที่ 4.13-4.15 ซึ่งแสดงผลการเปรียบเทียบค่า Wilk' s Lambda ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ตามลำดับ ของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มไม่ซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 6 กลุ่ม พบว่า การจัดกลุ่มแบบ 2 ขั้นตอน ด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี Ward ร่วมกับ Fuzzy C-Means มีประสิทธิภาพดีกว่าวิธี SOM ร่วมกับ K-Means และวิธี Ward ร่วมกับ K-Means ในทุกกรณีสำหรับทุกขนาดตัวอย่าง และทุกจำนวนตัวแปร โดยให้ค่า Wilk's Lambda, ค่า RMSSTD และค่า Misclassification Rate ต่ำที่สุด

กรณีที่ 1 ตัวแปรอิสระต่อกัน

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ เมื่อขนาดตัวอย่างเพิ่มขึ้น(ยกเว้น จำนวนตัวแปรเท่ากับ 6)

กรณีที่ 2 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(-1,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น

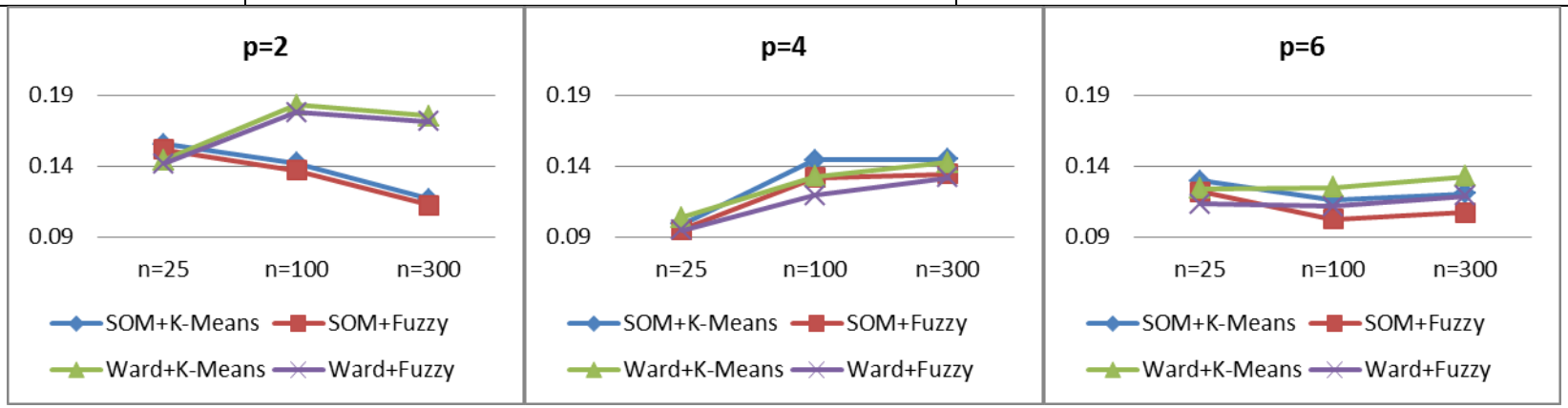
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

กรณีที่ 3 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(0,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ เมื่อจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ เมื่อขนาดตัวอย่างเพิ่มขึ้น

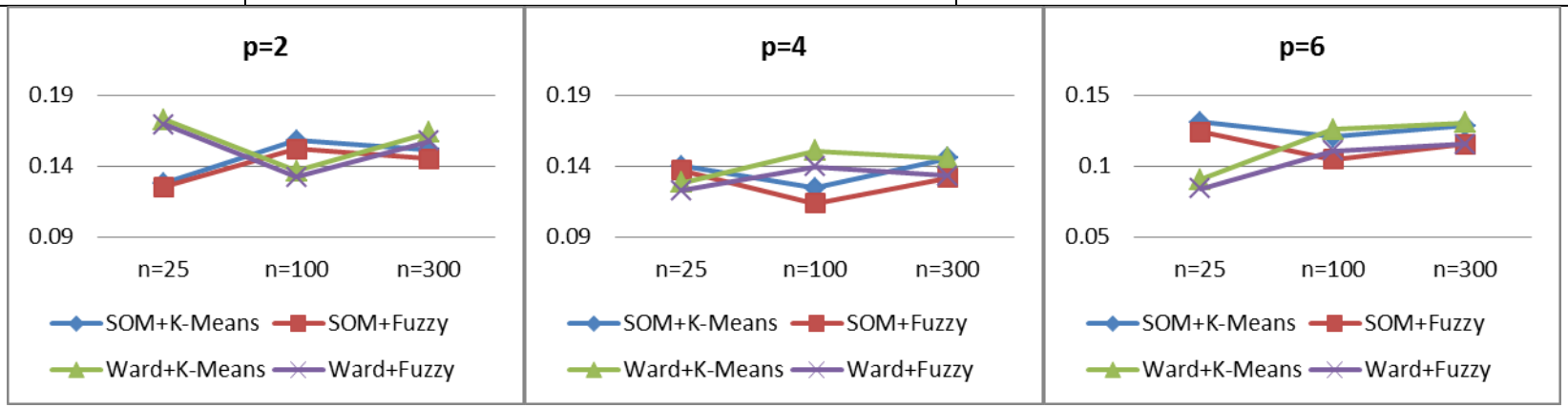
ตารางที่ 4.16 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวน	ขนาด	Mean						S.D.					
	ตัวแปร	ตัวอย่าง	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.2459	0.1556	0.1516	0.2019	0.1442	0.1418	0.0464	0.0321	0.0292	0.0433	0.0401	0.0305
		100	0.1977	0.1421	0.1369	0.2459	0.1832	0.1780	0.0284	0.0207	0.0188	0.0213	0.0175	0.0179
		300	0.2186	0.1167	0.1122	0.2351	0.1754	0.1715	0.0173	0.0128	0.0118	0.0190	0.0136	0.0126
	4	25	0.1916	0.0991	0.0949	0.2047	0.1035	0.0947	0.0392	0.0196	0.0219	0.0306	0.0211	0.0218
		100	0.1674	0.1442	0.1320	0.1846	0.1328	0.1194	0.0145	0.0160	0.0136	0.0153	0.0172	0.0124
		300	0.2216	0.1448	0.1345	0.2135	0.1428	0.1316	0.0105	0.0148	0.0065	0.0109	0.0167	0.0089
	6	25	0.2266	0.1296	0.1220	0.2076	0.1238	0.1132	0.0389	0.0209	0.0177	0.0355	0.0174	0.0175
		100	0.1490	0.1163	0.1025	0.1684	0.1250	0.1116	0.0114	0.0163	0.0114	0.0163	0.0165	0.0094
		300	0.2422	0.1207	0.1070	0.2659	0.1327	0.1187	0.0101	0.0179	0.0078	0.0146	0.0166	0.0055



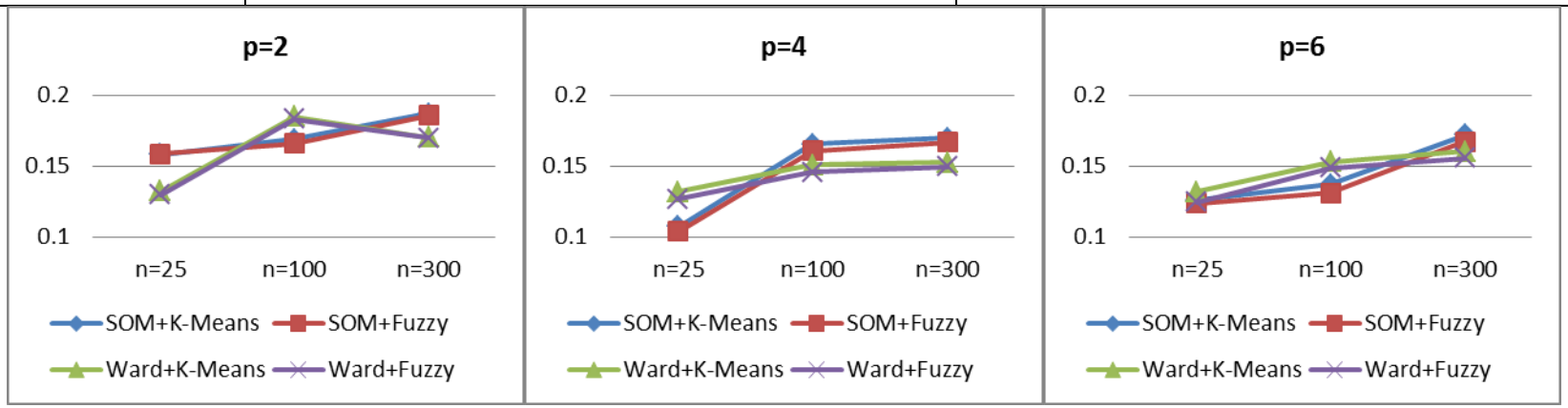
ตารางที่ 4.16 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.2012	0.1277	0.1252	0.2894	0.1730	0.1693	0.0671	0.0357	0.0385	0.0654	0.0440	0.0416
		100	0.1847	0.1581	0.1522	0.1787	0.1365	0.1323	0.0286	0.0263	0.0252	0.0200	0.0273	0.0299
		300	0.2001	0.1518	0.1452	0.2201	0.1636	0.1579	0.0221	0.0178	0.0199	0.0264	0.0206	0.0239
	4	25	0.2293	0.1400	0.1371	0.2458	0.1283	0.1229	0.0366	0.0219	0.0185	0.0337	0.0288	0.0256
		100	0.1705	0.1244	0.1136	0.2055	0.1510	0.1396	0.0175	0.0168	0.0183	0.0145	0.0186	0.0195
		300	0.2138	0.1454	0.1317	0.2078	0.1454	0.1331	0.0139	0.0169	0.0171	0.0192	0.0168	0.0172
	6	25	0.2228	0.1312	0.1240	0.1895	0.0907	0.0841	0.0305	0.0192	0.0192	0.0378	0.0175	0.0185
		100	0.1603	0.1209	0.1047	0.1543	0.1259	0.1108	0.0160	0.0153	0.0140	0.0175	0.0163	0.0128
		300	0.2058	0.1287	0.1156	0.2147	0.1307	0.1156	0.0104	0.0149	0.0135	0.0134	0.0178	0.0121



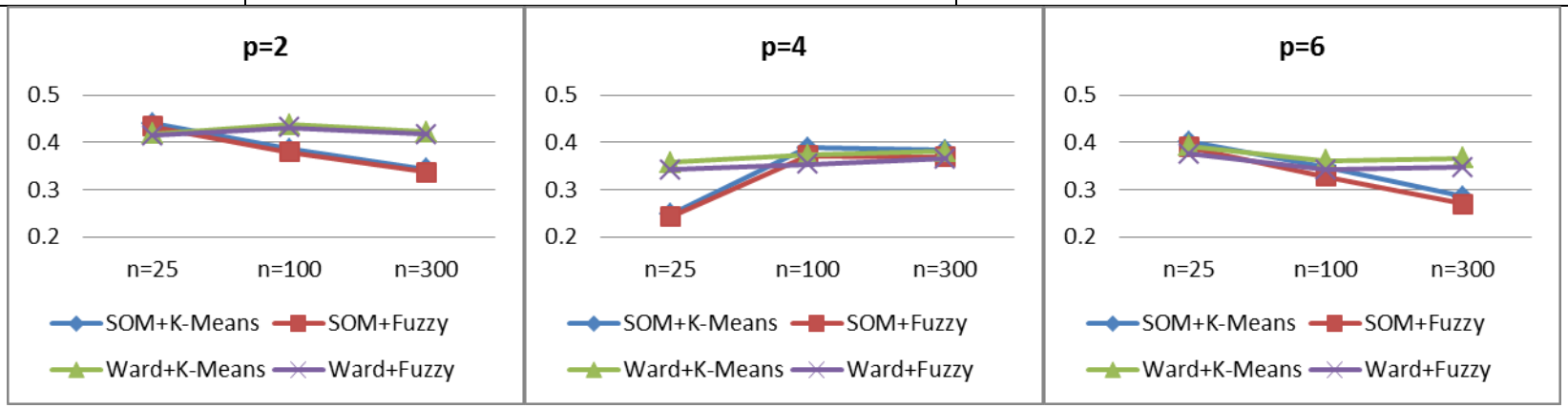
ตารางที่ 4.16 แสดงค่า Wilk' s Lambda ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.2558	0.1584	0.1588	0.1870	0.1323	0.1297	0.0676	0.0391	0.0397	0.0681	0.0476	0.0442
		100	0.2066	0.1691	0.1662	0.2457	0.1852	0.1831	0.0232	0.0190	0.0195	0.0269	0.0238	0.0257
		300	0.2434	0.1871	0.1857	0.2359	0.1705	0.1697	0.0193	0.0158	0.0165	0.0194	0.0198	0.0199
	4	25	0.1784	0.1074	0.1039	0.1969	0.1318	0.1268	0.0454	0.0303	0.0272	0.0490	0.0405	0.0321
		100	0.1851	0.1654	0.1608	0.2148	0.1507	0.1455	0.0242	0.0161	0.0177	0.0284	0.0171	0.0142
		300	0.2172	0.1698	0.1668	0.2219	0.1524	0.1495	0.0143	0.0128	0.0131	0.0115	0.0126	0.0121
	6	25	0.2101	0.1258	0.1237	0.2063	0.1319	0.1245	0.0334	0.0224	0.0218	0.0311	0.0333	0.0295
		100	0.1576	0.1370	0.1311	0.2223	0.1529	0.1485	0.0170	0.0169	0.0157	0.0141	0.0152	0.0149
		300	0.2473	0.1719	0.1671	0.2288	0.1603	0.1556	0.0111	0.0127	0.0125	0.0166	0.0118	0.0110



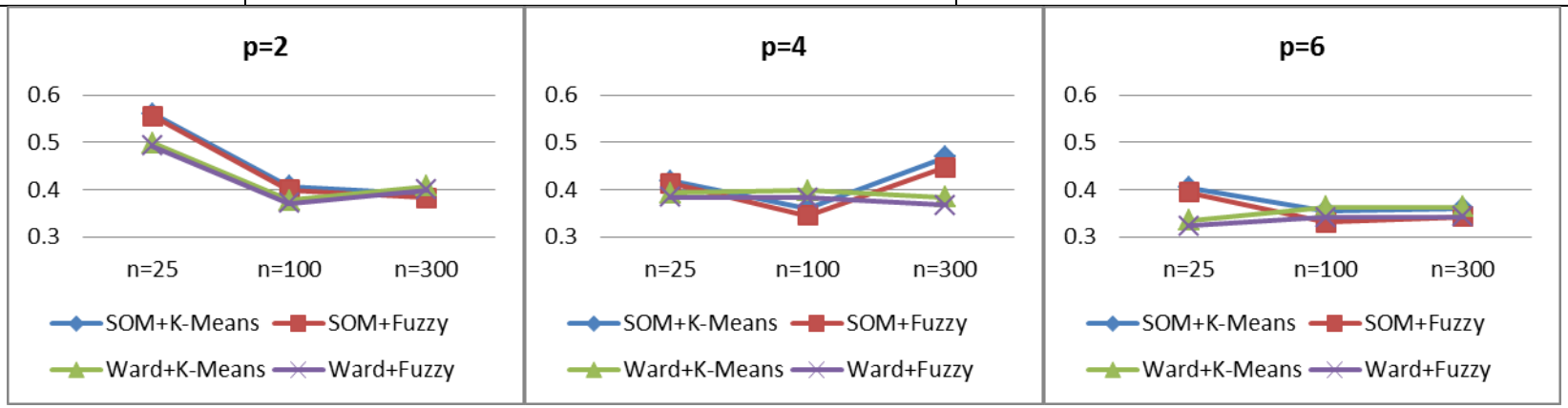
ตารางที่ 4.17 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.5674	0.4405	0.4346	0.4885	0.4183	0.4148	0.0585	0.0504	0.0454	0.0516	0.0503	0.0362
		100	0.4552	0.3862	0.3792	0.5215	0.4384	0.4321	0.0325	0.0270	0.0250	0.0212	0.0217	0.0225
		300	0.4474	0.3440	0.3374	0.5124	0.4220	0.4174	0.0186	0.0161	0.0150	0.0169	0.0165	0.0154
	4	25	0.3841	0.2477	0.2425	0.4766	0.3573	0.3418	0.0480	0.0316	0.0357	0.0425	0.0354	0.0378
		100	0.4557	0.3892	0.3724	0.4665	0.3733	0.3540	0.0176	0.0212	0.0184	0.0124	0.0245	0.0186
		300	0.4489	0.3833	0.3696	0.4395	0.3805	0.3657	0.0113	0.0193	0.0088	0.0269	0.0220	0.0123
	6	25	0.4711	0.4024	0.3904	0.4958	0.3928	0.3757	0.0468	0.0357	0.0313	0.0361	0.0318	0.0325
		100	0.3966	0.3492	0.3281	0.4115	0.3621	0.3425	0.0150	0.0240	0.0172	0.0127	0.0238	0.0145
		300	0.3877	0.2854	0.2693	0.4521	0.3666	0.3473	0.0105	0.0250	0.0110	0.0207	0.0228	0.0080



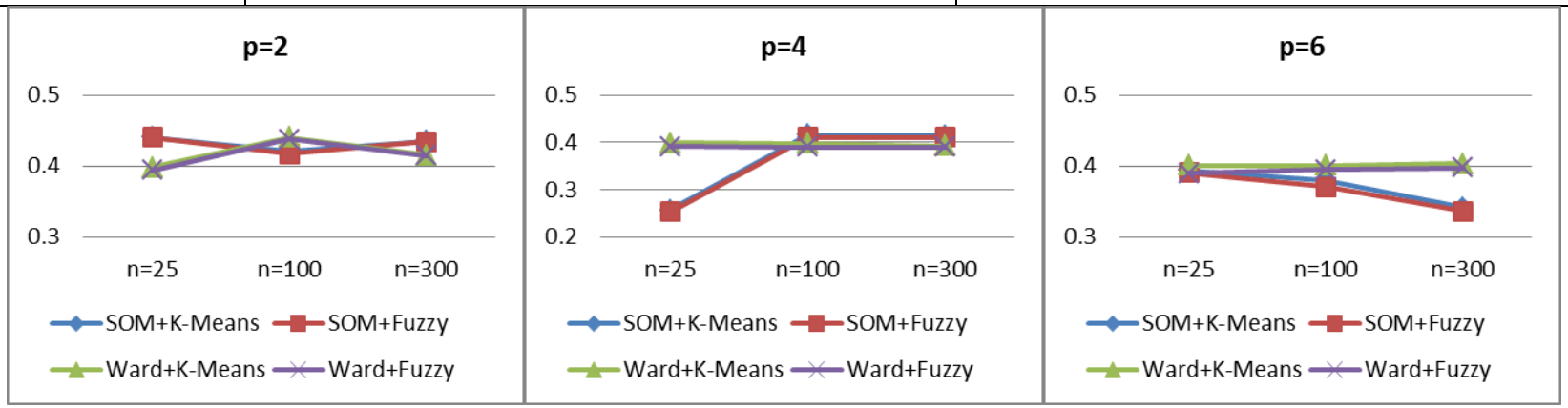
ตารางที่ 4.17 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.6685	0.5621	0.5569	0.5413	0.4991	0.4930	0.0769	0.0465	0.0499	0.0637	0.0531	0.0521
		100	0.4871	0.4069	0.3989	0.4687	0.3775	0.3713	0.0334	0.0335	0.0327	0.0396	0.0318	0.0354
		300	0.4652	0.3919	0.3829	0.4712	0.4071	0.3996	0.0252	0.0224	0.0260	0.0247	0.0236	0.0277
	4	25	0.4952	0.4184	0.4139	0.4852	0.3926	0.3838	0.0458	0.0351	0.0314	0.0477	0.0427	0.0394
		100	0.4125	0.3609	0.3446	0.4795	0.3979	0.3825	0.0199	0.0220	0.0249	0.0229	0.0246	0.0265
		300	0.5284	0.4700	0.4473	0.4788	0.3838	0.3672	0.0154	0.0230	0.0237	0.0260	0.0214	0.0227
	6	25	0.4774	0.4052	0.3938	0.4256	0.3352	0.3225	0.0380	0.0311	0.0318	0.0387	0.0311	0.0329
		100	0.4012	0.3558	0.3313	0.4526	0.3632	0.3408	0.0200	0.0235	0.0221	0.0238	0.0238	0.0194
		300	0.4159	0.3610	0.3425	0.4569	0.3637	0.3424	0.0117	0.0217	0.0202	0.0291	0.0241	0.0171



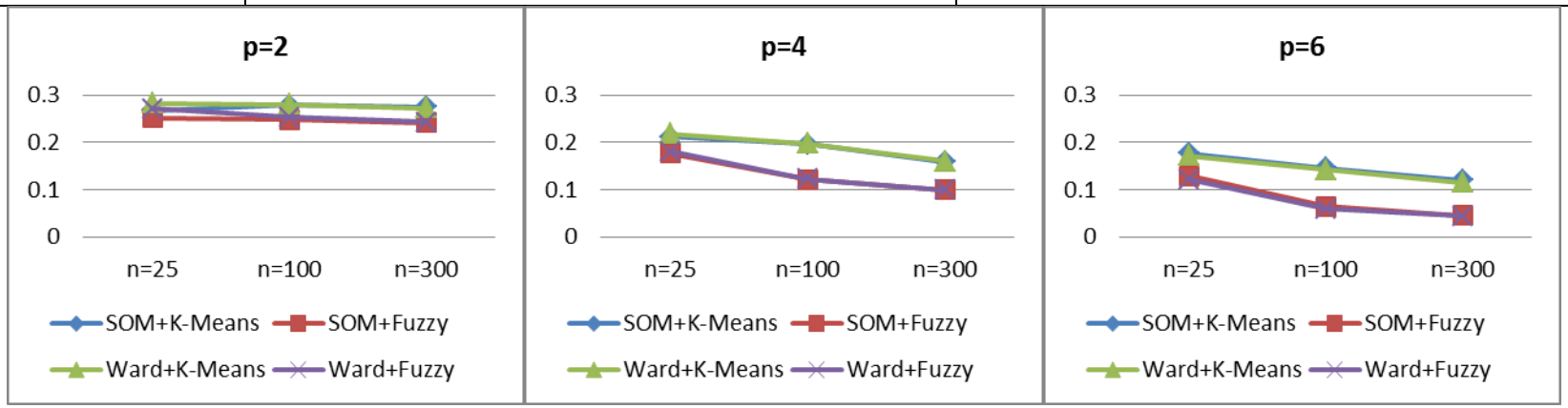
ตารางที่ 4.17 แสดงค่า RMSSTD ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.5603	0.4402	0.4405	0.4489	0.3979	0.3939	0.0806	0.0589	0.0598	0.0743	0.0622	0.0591
		100	0.4654	0.4211	0.4174	0.4970	0.4407	0.4382	0.0270	0.0232	0.0239	0.0268	0.0279	0.0296
		300	0.4972	0.4358	0.4342	0.4815	0.4159	0.4149	0.0194	0.0188	0.0197	0.0274	0.0223	0.0223
	4	25	0.3314	0.2564	0.2521	0.4600	0.3992	0.3915	0.0542	0.0441	0.0401	0.0560	0.0517	0.0432
		100	0.4406	0.4165	0.4107	0.4622	0.3975	0.3904	0.0283	0.0205	0.0223	0.0280	0.0220	0.0185
		300	0.4698	0.4153	0.4115	0.4664	0.3933	0.3896	0.0147	0.0156	0.0161	0.0199	0.0150	0.0146
	6	25	0.5102	0.3941	0.3909	0.4698	0.4010	0.3898	0.0422	0.0365	0.0361	0.0490	0.0424	0.0381
		100	0.4067	0.3792	0.3709	0.4698	0.4005	0.3947	0.0200	0.0209	0.0199	0.0228	0.0203	0.0202
		300	0.4094	0.3412	0.3364	0.4739	0.4034	0.3976	0.0114	0.0161	0.0161	0.0145	0.0148	0.0140



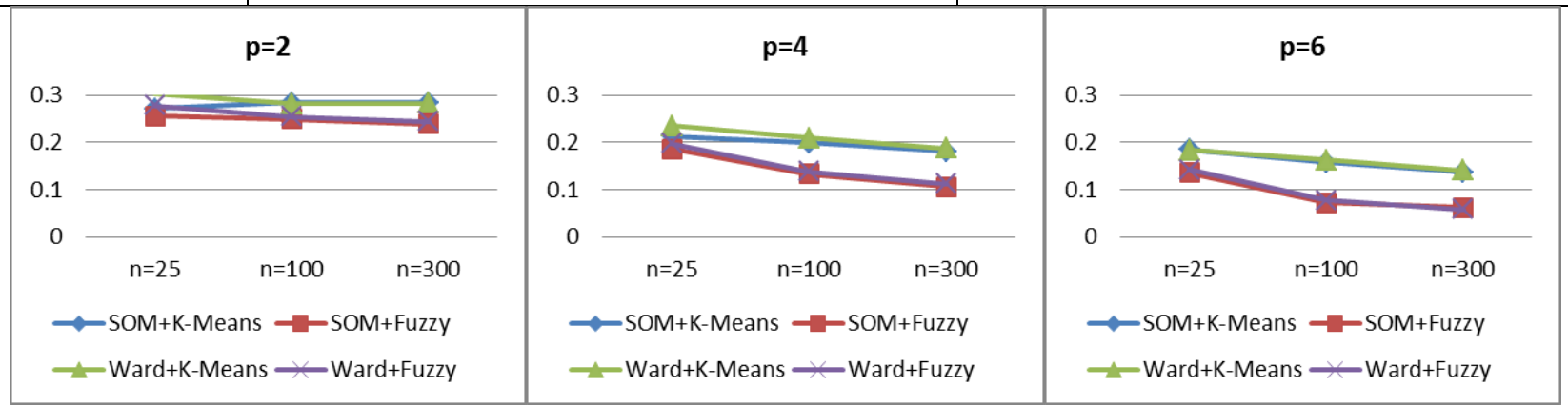
ตารางที่ 4.18 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case1	2	25	0.3940	0.2682	0.2522	0.4154	0.2838	0.2715	0.0463	0.0669	0.0720	0.0413	0.0708	0.0689
		100	0.3765	0.2798	0.2480	0.3885	0.2815	0.2535	0.0443	0.0614	0.0569	0.0481	0.0571	0.0555
		300	0.4884	0.2764	0.2415	0.4874	0.2729	0.2431	0.0270	0.0529	0.0502	0.0236	0.0562	0.0471
	4	25	0.3924	0.2122	0.1770	0.4004	0.2191	0.1806	0.0481	0.0770	0.0853	0.0487	0.0783	0.0795
		100	0.3226	0.1972	0.1218	0.3321	0.1978	0.1223	0.0455	0.0732	0.0554	0.0672	0.0908	0.0505
		300	0.4615	0.1596	0.0997	0.4595	0.1605	0.0991	0.0214	0.0762	0.0239	0.0415	0.0787	0.0254
	6	25	0.3832	0.1776	0.1292	0.3954	0.1718	0.1217	0.0503	0.0953	0.0761	0.0703	0.0841	0.0826
		100	0.2890	0.1467	0.0644	0.2958	0.1417	0.0598	0.0373	0.0858	0.0426	0.0579	0.0736	0.0433
		300	0.4564	0.1204	0.0444	0.4395	0.1153	0.0440	0.0225	0.0794	0.0267	0.0214	0.0794	0.0148



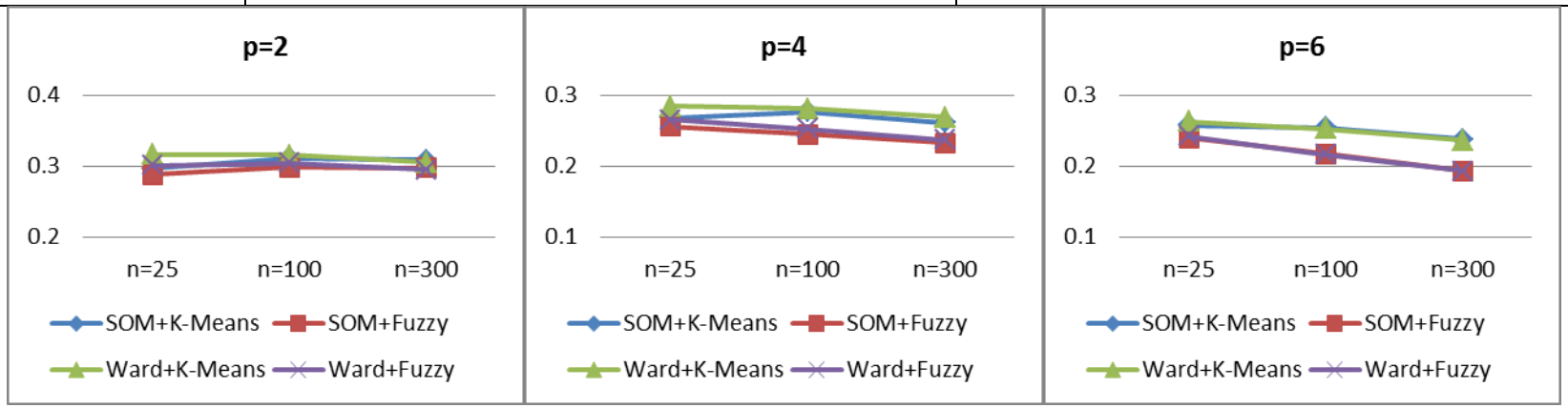
ตารางที่ 4.18 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case2	2	25	0.3936	0.2722	0.2569	0.4251	0.3025	0.2782	0.0514	0.0881	0.0865	0.0644	0.0900	0.0911
		100	0.3827	0.2842	0.2497	0.3915	0.2824	0.2550	0.0486	0.0633	0.0765	0.0454	0.0717	0.0923
		300	0.4742	0.2846	0.2396	0.4998	0.2841	0.2446	0.0393	0.0637	0.0754	0.0336	0.0683	0.0827
	4	25	0.3924	0.2134	0.1862	0.4015	0.2356	0.1981	0.0582	0.0796	0.0813	0.0606	0.0881	0.0993
		100	0.3561	0.1999	0.1325	0.3952	0.2092	0.1385	0.0427	0.0739	0.0827	0.0467	0.0757	0.0815
		300	0.4651	0.1815	0.1056	0.3854	0.1871	0.1126	0.0300	0.0793	0.0611	0.0379	0.0753	0.0609
	6	25	0.3933	0.1854	0.1359	0.3758	0.1842	0.1425	0.0516	0.0839	0.0842	0.0514	0.0856	0.0777
		100	0.3302	0.1576	0.0726	0.3554	0.1636	0.0779	0.0422	0.0797	0.0625	0.0476	0.0802	0.0593
		300	0.4628	0.1365	0.0620	0.4758	0.1406	0.0576	0.0247	0.0785	0.0520	0.0305	0.0824	0.0436



ตารางที่ 4.18 แสดงค่า Misclassification ของการจัดกลุ่มแบบ 2 ขั้นตอนสำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มจริงที่จำลองเป็น 6 กลุ่ม

	จำนวนตัวแปร	ขนาดตัวอย่าง	Mean						S.D.					
			SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F	SOM	SOM+K	SOM+F	Ward	Ward+K	Ward+F
Case3	2	25	0.3982	0.2964	0.2881	0.4172	0.3162	0.3006	0.0581	0.0766	0.0814	0.0587	0.0773	0.0795
		100	0.3665	0.3111	0.2986	0.4336	0.3157	0.3039	0.0444	0.0519	0.0582	0.0442	0.0595	0.0560
		300	0.4981	0.3087	0.2975	0.4510	0.3055	0.2947	0.0248	0.0499	0.0501	0.0310	0.0532	0.0468
	4	25	0.3958	0.2682	0.2558	0.4043	0.2851	0.2656	0.0502	0.0825	0.0723	0.0430	0.0789	0.0718
		100	0.3511	0.2767	0.2450	0.4150	0.2815	0.2526	0.0438	0.0531	0.0586	0.0456	0.0550	0.0541
		300	0.4884	0.2616	0.2327	0.4328	0.2690	0.2375	0.0217	0.0537	0.0512	0.0337	0.0605	0.0425
	6	25	0.3943	0.2582	0.2397	0.3952	0.2630	0.2427	0.0425	0.0646	0.0653	0.0585	0.0799	0.0765
		100	0.3350	0.2548	0.2173	0.4035	0.2525	0.2166	0.0375	0.0624	0.0668	0.0306	0.0624	0.0534
		300	0.4827	0.2382	0.1934	0.4186	0.2360	0.1934	0.0182	0.0655	0.0440	0.0181	0.0641	0.0396



จากตารางที่ 4.16-4.18 ซึ่งแสดงผลการเปรียบเทียบค่า Wilk' s Lambda ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ตามลำดับ ของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองเป็น 6 กลุ่ม พบว่า การจัดกลุ่มแบบ 2 ขั้นตอน ด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี Ward ร่วมกับ Fuzzy C-Means มีประสิทธิภาพดีกว่าวิธี SOM ร่วมกับ K-Means และวิธี Ward ร่วมกับ K-Means ในทุกกรณีสำหรับทุกขนาดตัวอย่าง และทุกจำนวนตัวแปร โดยให้ค่า Wilk's Lambda, ค่า RMSSTD และค่า Misclassification Rate ต่ำที่สุด

กรณีที่ 1 ตัวแปรอิสระต่อกัน

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ เมื่อจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

กรณีที่ 2 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(-1,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะเพิ่มขึ้นเรื่อยๆ ถ้าจำนวนตัวแปรเพิ่มขึ้น

- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

กรณีที่ 3 ตัวแปรมีความสัมพันธ์กัน $\rho = Corr(X_i, X_j) \sim Unif(0,1)$

- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับจำนวนตัวแปร
- เมื่อพิจารณาจำนวนตัวแปรคงที่ พบว่า ค่า Wilk's Lambda และค่า RMSSTD มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง
- เมื่อพิจารณาขนาดตัวอย่างคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่จะลดลงเรื่อยๆ เมื่อจำนวนตัวแปรเพิ่มขึ้น
- เมื่อพิจารณาที่จำนวนตัวแปรคงที่ พบว่า ค่า Misclassification Rate มีแนวโน้มที่ไม่ชัดเจน ขึ้นอยู่กับขนาดตัวอย่าง

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

จากการศึกษาเปรียบเทียบประสิทธิภาพการจัดกลุ่มข้อมูลโดยใช้อัลกอริทึมการจัดกลุ่มแบบ 2 ขั้นตอนทั้ง 4 วิธี ได้แก่ วิธีที่ 1 SOM ร่วมกับ K-Means วิธีที่ 2 SOM ร่วมกับ Fuzzy C-Means วิธีที่ 3 Ward ร่วมกับ K-Means และวิธีที่ 4 ward ร่วมกับ Fuzzy C-Means สำหรับข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) โดยทำการศึกษาสำหรับข้อมูลแต่ละกลุ่มไม่มีการซ้อนทับกันและข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองได้แก่ 2 กลุ่ม, 4 กลุ่ม และ 6 กลุ่ม ขนาดตัวอย่างที่ศึกษาเป็น 25, 100 และ 300 จำนวนตัวแปรที่ใช้ในการศึกษาเป็น 2, 4 และ 6 ตัวแปร โดยศึกษาทั้งในกรณีที่ตัวแปรอิสระต่อกัน และกรณีที่ตัวแปรมีความสัมพันธ์กัน โดยค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ สุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ค่าที่สุ่มอยู่ในช่วง (-1,1) และ (0,1) จากการศึกษาครั้งนี้เกณฑ์การพิจารณาเปรียบเทียบคือ ค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate สามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการวิจัย

จากการพิจารณาค่า Wilk' s Lambda ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate พบว่า การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี ward ร่วมกับ Fuzzy C-Means มีประสิทธิภาพดีกว่าวิธี SOM ร่วมกับ K-Means และวิธี Ward ร่วมกับ K-Means ในทุกขนาดตัวอย่าง และทุกมิติ (จำนวนตัวแปร) ที่ศึกษา โดย การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี ward ร่วมกับ Fuzzy C-Means ให้ประสิทธิภาพที่แตกต่างกันออกไปในแต่ละกรณี ขึ้นอยู่กับจำนวนตัวอย่าง และจำนวนตัวแปร โดยภาพรวมแล้วสามารถสรุปผลได้ดังนี้

กรณีที่ 1 การจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 2 กลุ่ม

เนื่องจากขนาดตัวอย่างที่ศึกษาเป็น 25, 100 และ 300 จำนวนตัวแปรที่ใช้ในการศึกษาเป็น 2, 4 และ 6 ตัวแปร โดยศึกษาทั้งในกรณีที่ตัวแปรอิสระต่อกัน และกรณีที่ตัวแปรมีความสัมพันธ์กัน โดยค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ สุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ค่าที่สุ่มอยู่ในช่วง (-1,1) และ (0,1) ซึ่งปัจจัยที่ใช้ศึกษาเหล่านี้ส่งผลต่อค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ที่แตกต่างกันในแต่ละกรณีที่ศึกษา ทางผู้วิจัยจึงได้พิจารณาค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของทุกกรณีที่กล่าวมาประกอบการสรุปผลดังนี้

ตารางที่ 5.1 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 2 กลุ่ม

Method	Mean			S.D.		
	Wilk	RMSSTD	Mis	Wilk	RMSSTD	Mis
SOM	0.3534	0.6110	0.0254	0.0531	0.0428	0.0215
SOM+K	0.3431	0.6003	0.0214	0.0522	0.0438	0.0181
SOM+F	0.3400	0.5969	0.0191	0.0512	0.0409	0.0169
Ward	0.3603	0.6094	0.0282	0.0574	0.0452	0.0247
Ward+K	0.3500	0.5991	0.0238	0.0586	0.0472	0.0214
Ward+F	0.3484	0.5975	0.0227	0.0582	0.0465	0.0205

จากตารางที่ 5.1 พบว่า สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 2 กลุ่ม การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means มีประสิทธิภาพในการจัดกลุ่มดีที่สุด โดยให้ค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate โดยเฉลี่ยในทุกกรณีต่ำที่สุด

กรณีที่ 2 การจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองเป็น 2 กลุ่ม

เนื่องจากขนาดตัวอย่างที่ศึกษาเป็น 25, 100 และ 300 จำนวนตัวแปรที่ใช้ในการศึกษาเป็น 2, 4 และ 6 ตัวแปร โดยศึกษาทั้งในกรณีที่ตัวแปรอิสระต่อกัน และกรณีที่ตัวแปรมีความสัมพันธ์กัน โดยค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ สุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ค่าที่สุ่มอยู่ในช่วง (-1,1) และ (0,1) ซึ่งปัจจัยที่ใช้ศึกษาเหล่านี้ส่งผลต่อค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ที่แตกต่างกันในแต่ละกรณีที่ศึกษา ทางผู้วิจัยจึงได้พิจารณาค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของทุกกรณีที่กล่าวมาประกอบการสรุปผลดังนี้

ตารางที่ 5.2 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองเป็น 2 กลุ่ม

Method	Mean			S.D.		
	Wilk	RMSSTD	Mis	Wilk	RMSSTD	Mis
SOM	0.6628	0.8245	0.0987	0.0854	0.0847	0.0456
SOM+K	0.6533	0.8155	0.0932	0.0870	0.0855	0.0479
SOM+F	0.6479	0.8122	0.0856	0.0879	0.0859	0.0482
Ward	0.6929	0.8164	0.0993	0.0556	0.1624	0.0488
Ward+K	0.6796	0.8072	0.0932	0.0559	0.1610	0.0502
Ward+F	0.6768	0.8056	0.0915	0.0548	0.1606	0.0482

จากตารางที่ 5.2 พบว่า สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 2 กลุ่ม การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means มีประสิทธิภาพในการจัดกลุ่มดีที่สุด โดยให้ค่า Wilk' s Lambda และค่า Misclassification Rate โดยเฉลี่ยในทุกกรณีต่ำที่สุด

กรณีที่ 3 การจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 4 กลุ่ม

เนื่องจากขนาดตัวอย่างที่ศึกษาเป็น 25, 100 และ 300 จำนวนตัวแปรที่ใช้ในการศึกษาเป็น 2, 4 และ 6 ตัวแปร โดยศึกษาทั้งในกรณีที่ตัวแปรอิสระต่อกัน และกรณีที่ตัวแปรมีความสัมพันธ์กัน โดยค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ สุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ค่าที่สุ่มอยู่ในช่วง (-1,1) และ (0,1) ซึ่งปัจจัยที่ใช้ศึกษาเหล่านี้ส่งผลต่อค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ที่แตกต่างกันในแต่ละกรณีที่ศึกษา ทางผู้วิจัยจึงได้พิจารณาค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของทุกกรณีที่กล่าวมาประกอบการสรุปผลดังนี้

ตารางที่ 5.3 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 4 กลุ่ม

Method	Mean			S.D.		
	Wilk	RMSSTD	Mis	Wilk	RMSSTD	Mis
SOM	0.2336	0.4973	0.4488	0.0395	0.0665	0.1797
SOM+K	0.1301	0.3571	0.0915	0.0272	0.0563	0.0308
SOM+F	0.1159	0.3372	0.0539	0.0330	0.0614	0.0353
Ward	0.2127	0.4539	0.4258	0.0199	0.0252	0.2295
Ward+K	0.1213	0.3532	0.0813	0.0158	0.0241	0.0243
Ward+F	0.1053	0.3305	0.0400	0.0204	0.0304	0.0289

จากตารางที่ 5.3 พบว่า สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 4 กลุ่ม การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี ward ร่วมกับ Fuzzy C-Means มีประสิทธิภาพในการจัดกลุ่มดีที่สุด โดยให้ค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate โดยเฉลี่ยในทุกกรณีต่ำที่สุด

กรณีที่ 4 การจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองเป็น 4 กลุ่ม

เนื่องจากขนาดตัวอย่างที่ศึกษาเป็น 25, 100 และ 300 จำนวนตัวแปรที่ใช้ในการศึกษาเป็น 2, 4 และ 6 ตัวแปร โดยศึกษาทั้งในกรณีที่ตัวแปรอิสระต่อกัน และกรณีที่ตัวแปรมีความสัมพันธ์กัน โดยค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ สุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ค่าที่สุ่มอยู่ในช่วง (-1,1) และ (0,1) ซึ่งปัจจัยที่ใช้ศึกษาเหล่านี้ส่งผลต่อค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ที่แตกต่างกันในแต่ละกรณีที่ศึกษา ทางผู้วิจัยจึงได้พิจารณาค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของทุกกรณีที่กล่าวมาประกอบการสรุปผลดังนี้

ตารางที่ 5.4 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองเป็น 4 กลุ่ม

Method	Mean			S.D.		
	Wilk	RMSSTD	Mis	Wilk	RMSSTD	Mis
SOM	0.3795	0.6574	0.3905	0.0350	0.0882	0.0329
SOM+K	0.2860	0.5738	0.1847	0.0414	0.0839	0.0615
SOM+F	0.2759	0.5434	0.1523	0.0470	0.0874	0.0733
Ward	0.3623	0.6136	0.3244	0.0446	0.0436	0.0386
Ward+K	0.2960	0.5555	0.1870	0.0507	0.0471	0.0675
Ward+F	0.2858	0.5454	0.1557	0.0555	0.0519	0.0787

จากตารางที่ 5.4 พบว่า สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 4 กลุ่ม การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means มีประสิทธิภาพในการจัดกลุ่มดีที่สุด โดยให้ค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate โดยเฉลี่ยในทุกกรณีต่ำที่สุด

กรณีที่ 5 การจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 6 กลุ่ม

เนื่องจากขนาดตัวอย่างที่ศึกษาเป็น 25, 100 และ 300 จำนวนตัวแปรที่ใช้ในการศึกษาเป็น 2, 4 และ 6 ตัวแปร โดยศึกษาทั้งในกรณีที่ตัวแปรอิสระต่อกัน และกรณีที่ตัวแปรมีความสัมพันธ์กัน โดยค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ สุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ค่าที่สุ่มอยู่ในช่วง (-1,1) และ (0,1) ซึ่งปัจจัยที่ใช้ศึกษาเหล่านี้ส่งผลต่อค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ที่แตกต่างกันในแต่ละกรณีที่ศึกษา ทางผู้วิจัยจึงได้พิจารณาค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของทุกกรณีที่กล่าวมาประกอบการสรุปผลดังนี้

ตารางที่ 5.5 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 6 กลุ่ม

Method	Mean			S.D.		
	Wilk	RMSSTD	Mis	Wilk	RMSSTD	Mis
SOM	0.1335	0.3920	0.3597	0.0346	0.0702	0.0918
SOM+K	0.0657	0.2731	0.1411	0.0085	0.0490	0.0281
SOM+F	0.0565	0.2431	0.0887	0.0109	0.0525	0.0414
Ward	0.1388	0.3639	0.3444	0.0143	0.0244	0.0298
Ward+K	0.0629	0.2600	0.1285	0.0098	0.0225	0.0215
Ward+F	0.0541	0.2395	0.0747	0.0111	0.0275	0.0344

จากตารางที่ 5.5 พบว่า สำหรับข้อมูลที่แต่ละกลุ่มไม่มีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 6 กลุ่ม การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี ward ร่วมกับ Fuzzy C-Means มีประสิทธิภาพในการจัดกลุ่มดีที่สุด โดยให้ค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate โดยเฉลี่ยในทุกกรณีต่ำที่สุด

กรณีที่ 6 การจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองเป็น 6 กลุ่ม

เนื่องจากขนาดตัวอย่างที่ศึกษาเป็น 25, 100 และ 300 จำนวนตัวแปรที่ใช้ในการศึกษาเป็น 2, 4 และ 6 ตัวแปร โดยศึกษาทั้งในกรณีที่ตัวแปรอิสระต่อกัน และกรณีที่ตัวแปรมีความสัมพันธ์กัน โดยค่าความสัมพันธ์ระหว่างตัวแปรแต่ละคู่ สุ่มมาจากการแจกแจงแบบยูนิฟอร์ม (Uniform Distribution) ค่าที่สุ่มอยู่ในช่วง (-1,1) และ (0,1) ซึ่งปัจจัยที่ใช้ศึกษาเหล่านี้ส่งผลต่อค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ที่แตกต่างกันในแต่ละกรณีที่ศึกษา ทางผู้วิจัยจึงได้พิจารณาค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของทุกกรณีที่กล่าวมาประกอบการสรุปผลดังนี้

ตารางที่ 5.6 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการจัดกลุ่มแบบ 2 ขั้นตอน สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน 40% โดยจำนวนกลุ่มที่จำลองเป็น 6 กลุ่ม

Method	Mean			S.D.		
	Wilk	RMSSTD	Mis	Wilk	RMSSTD	Mis
SOM	0.2056	0.4614	0.4046	0.0296	0.0683	0.0577
SOM+K	0.1403	0.3874	0.2306	0.0217	0.0636	0.0571
SOM+F	0.1336	0.3778	0.1884	0.0238	0.0650	0.0791
Ward	0.2146	0.4724	0.4108	0.0294	0.0274	0.0440
Ward+K	0.1436	0.3956	0.2355	0.0227	0.0321	0.0601
Ward+F	0.1363	0.3850	0.1929	0.0251	0.0366	0.0824

จากตารางที่ 5.6 พบว่า สำหรับข้อมูลที่แต่ละกลุ่มมีการซ้อนทับกัน โดยจำนวนกลุ่มที่จำลองเป็น 6 กลุ่ม การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means มีประสิทธิภาพในการจัดกลุ่มดีที่สุด โดยให้ค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate โดยเฉลี่ยในทุกกรณีต่ำที่สุด

จากตารางที่ 5.1-5.6 แสดงค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ของการจัดกลุ่มแบบ 2 ขั้นตอน โดยพิจารณาที่ข้อมูลแต่ละกลุ่มมีการซ้อนทับกัน 40% (40% Overlapping Clusters) และข้อมูลแต่ละกลุ่มไม่มีการซ้อนทับกัน (Nonoverlapping Clusters) จำนวนกลุ่มที่จำลองเป็น 2 กลุ่ม, 4 กลุ่ม และ 6 กลุ่ม พบว่าการจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means และวิธี ward ร่วมกับ Fuzzy C-Means ให้ประสิทธิภาพในการจัดกลุ่มดีที่สุด โดยให้ค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate โดยเฉลี่ยในทุกขนาดตัวอย่าง ทุกจำนวนตัวแปร และทุกลักษณะความสัมพันธ์ของตัวแปร ต่ำที่สุด โดยการจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means ให้ประสิทธิภาพที่ดีเมื่อข้อมูลที่จำลองมีจำนวนกลุ่มเป็น 2 กลุ่ม แต่เมื่อมีการจำลองจำนวนกลุ่มเพิ่มมากขึ้น กล่าวคือ เมื่อจำลองจำนวนกลุ่มเป็น 4 กลุ่ม และ 6 กลุ่ม การจัดกลุ่มแบบ 2 ขั้นตอนด้วยวิธี SOM ร่วมกับ Fuzzy C-Means จะมีประสิทธิภาพดีเมื่อข้อมูลมีการซ้อนทับกัน และการจัดกลุ่มด้วยวิธี ward ร่วมกับ Fuzzy C-Means จะมีประสิทธิภาพดีเมื่อข้อมูลไม่มีการซ้อนทับกัน นอกจากนี้ค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate มีแนวโน้มที่จะลดลงเมื่อมีการจำลองจำนวนกลุ่มเพิ่มมากขึ้น และในกรณีที่ข้อมูลแต่ละกลุ่มมีการซ้อนทับกัน ค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate จะมีค่ามากกว่ากรณีข้อมูลแต่ละกลุ่มไม่มีการซ้อนทับกัน

ตารางสรุปผล

ลักษณะการซ้อนทับกันของข้อมูล	จำนวนกลุ่ม		
	k=2	k=4	k=6
Nonoverlapping Clusters	SOM+Fuzzy	Ward+Fuzzy	Ward+Fuzzy
40% Overlapping Clusters	SOM+Fuzzy	SOM+Fuzzy	SOM+Fuzzy

5.2 **ข้อเสนอแนะ**

เนื่องจากการวิจัยในครั้งนี้ทำการศึกษาข้อมูลที่มีการแจกแจงแบบปกติหลายตัวแปร (Multivariate Normal Distribution) และพิจารณาเปรียบเทียบประสิทธิภาพค่า Wilk' s Lambda, ค่าความแตกต่างของข้อมูลภายในกลุ่ม (RMSSTD) และค่า Misclassification Rate ดังนั้นในการวิจัยต่อไปอาจทำการศึกษาข้อมูลที่มีการแจกแจงลักษณะอื่นๆ และอาจพิจารณาด้วยเกณฑ์วัดประสิทธิภาพอื่นๆ ซึ่งผลการวิจัยที่ได้อาจจะแตกต่างกัน

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. **การวิเคราะห์ข้อมูลหลายตัวแปร**. พิมพ์ครั้งที่ 4. จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพฯ, 2552.

ปิยธิดา รุจะศิริ. **การเปรียบเทียบเทคนิคการแบ่งกลุ่มข้อมูลสำหรับการวิเคราะห์กลุ่ม**. วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยเกษตรศาสตร์, 2552.

ศศิธร มงคลศรีพัฒนา. **การจัดกลุ่มศูนย์บริการและถ่ายทอดเทคโนโลยีทางการ เกษตรประจำ ตำบลในประเทศไทยโดยใช้อัลกอริทึม 2 ขั้นตอน คือ SOM และ Fuzzy C-Means**. วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยเกษตรศาสตร์, 2550.

อรนุช ชัยหมื่น. **การศึกษาเปรียบเทียบการแบ่งกลุ่มข้อมูลลูกค้าสินค้าหัตถกรรมไทยโดยใช้ วิธีการ 2 ขั้นตอนของ SOM กับ K-Means Algorithm และ Hierarchical กับ K-Means Algorithm**. วิทยานิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยเกษตรศาสตร์, 2548.

ภาษาอังกฤษ

Brian S. Everitt. **Cluster Analysis**. 5th edition. John Wiley and Sons, Ltd, 2011.

R.J. Kuo, L.M. Ho, C.M. Hu. **Integration of self-organizing feature map and K-means algorithm for market segmentation.** Computer & Operations Research 29 (2002) : 1475-1493.

R.J. Kuo, H.S. Wang, W.J. Chung. **Integration of self-organizing feature map neural network and genetic K-means algorithm for market segmentation.** Expert Systems with Applications 30 (2006) : 313-324.

Sueli A. Mingoti, Joab O. Lima. **Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms**. European Journal of Operational Research 174 (2006) : 1742-1759.

ภาคผนวก

โปรแกรม R ที่ใช้ในการจำลองข้อมูล

ในการวิจัยครั้งนี้ได้ใช้โปรแกรม R เวอร์ชัน 3.0.1 ในที่นี้จะขอแสดงเฉพาะกรณีที่จำลองกลุ่มเป็น 2 กลุ่ม (k=2) โดยขนาดตัวอย่าง n=25 และจำนวนตัวแปรเป็น 2 (p=2) ในการยกตัวอย่าง

1. การจำลองข้อมูลเป็น 2 กลุ่มโดยที่ข้อมูลแต่ละกลุ่มไม่ซ้อนทับกัน

```
n<-25
p<-2
xc<-c(gl(2,12),rep(2,1))
sd1<-runif(2,10,40)
mu1<-c(0,0)
sig1<-sd1*t(v2*sd1)
q<-runif(2,0.25,0.75)
sd2<-runif(2,10,40)
mu2<-(1.5*sd1)+q+(1.5*sd2)
sig2<-sd2*t(v2*sd2)
cl1<-mvrnorm(12,mu1,sig1)
cl2<-mvrnorm(13,mu2,sig2)
x<-rbind(cl1,cl2)
```

2. การจำลองข้อมูลเป็น 2 กลุ่มโดยที่ข้อมูลแต่ละกลุ่มมีการซ้อนทับกัน

```
n<-25
p<-2
xc<-c(gl(2,12),rep(2,1))
```

```
sd1<-runif(2,10,40)
mu1<-c(0,0)
sig1<-sd1*t(v2*sd1)
sd2<-runif(2,10,40)
mu2<-(0.3*sd1)+(1.5*sd2)
sig2<-sd2*t(v2*sd2)
cl1<-mvrnorm(12,mu1,sig1)
cl2<-mvrnorm(13,mu2,sig2)
x<-rbind(cl1,cl2)
```

3. การจัดกลุ่มข้อมูลโดยใช้ SOM, SOM+K-Means และ SOM+Fuzzy C-Means

```
x.std<-scale(x)
x.som<-som(x.std,xdim=2,ydim=3,topol="hexa",neigh="gaussian")
k<-sum(x.som$code.sum$nobs>0)
cl<-factor(x.som$visual$x):factor(x.som$visual$y)
cl<-factor(cl)
levels(cl)<-1:nlevels(cl)
cl<-as.integer(cl)
ind<-cluster.stats(d,cl,)
        ssw<-ind$within.cluster.ss
        ssb<-ind$ch*(ssw/(n-k))*(k-1)
        sst<-ssw+ssb
        w_som<-c(w_som,ssw/sst)
rmsstdsom[count]<-sqrt(ssw/sum(p*(table(cl)-1)))
aa<-table(cl,xc)
pred<-1-sum(apply(aa,1,max))/n
```

```
predictsom[count]<-pred
k<-sum(x.som$code.sum$nobs>0)
kb[count]<-k
        clk<-kmeans(x.std,k)
        ind<-cluster.stats(d,clk$cluster)
        ssw<-ind$within.cluster.ss
        ssb<-ind$ch*(ssw/(n-k))*(k-1)
        sst<-ssw+ssb
        wilk_lambdasom_k<-c(wilk_lambdasom_k,ssw/sst)
rmsstdsom_k[count]<-sqrt(ssw/sum(p*(clk$size-1)))
aa<-table(clk$cluster,xc)
pred<-1-sum(apply(aa,1,max))/n
predictsom_k[count]<-pred
clf<-cmeans(x.std,k)
        ind<-cluster.stats(d,clf$cluster)
        ssw<-ind$within.cluster.ss
        ssb<-ind$ch*(ssw/(n-k))*(k-1)
        sst<-ssw+ssb
        wilk_lambdasom_f<-c(wilk_lambdasom_f,ssw/sst)
rmsstdsom_f[count]<-sqrt(ssw/sum(p*(clf$size-1)))
aa<-table(clf$cluster,xc)
pred<-1-sum(apply(aa,1,max))/n
predictsom_f[count]<-pred
```

4. การจัดกลุ่มข้อมูลโดยใช้ Ward, Ward+K-Means และ Ward+Fuzzy C-Means

```
x.std<-scale(x)
dist.x<-dist(x.std,method = "euclidean")
```

```
x.ward<-hclust(dist.x,method="ward")
wk<-rep(Inf,3)
wn<-c()
rmsn<-c()
misn<-c()
for(kk in 2:6){
w<-w.fun(kk)
wn<-c(wn,w[1])
rmsn<-c(rmsn,w[3])
misn<-c(misn,w[4])
if(w[3]<wk[3]){
wk<-w}
}
wwn[count,]<-wn
rmsstdwn[count,]<-rmsn
miswn[count,]<-misn
kb[count]<-k<-wk[2]
clk<-kmeans(x.std,k)
ind<cluster.stats(d,clk$cluster)
ssw<-ind$within.cluster.ss
ssb<-ind$ch*(ssw/(n-k))*(k-1)
sst<-ssw+ssb
wilk_lambdaward_k<-c(wilk_lambdaward_k,ssw/sst)
rmsstdward_k[count]<-sqrt(ssw/sum(p*(clk$size-1)))
aa<-table(clk$cluster,xc)
misward_k[count]<-1-sum(apply(aa,1,max))/n
```

```
clf<-cmeans(x.std,k)
ind<-cluster.stats(d,clf$cluster)
ssw<-ind$within.cluster.ss
ssb<-ind$ch*(ssw/(n-k))*(k-1)
sst<-ssw+ssb
wilk_lambdaward_f<-c(wilk_lambdaward_f,ssw/sst)
rmsstdward_f[count]<-sqrt(ssw/sum(p*(clf$size-1)))
aa<-table(clf$cluster,xc)
pred<-1-sum(apply(aa,1,max))/n
misward_f[count]<-pred
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาว กัณฐิกา พรมมา เกิดวันพุธที่ 19 ตุลาคม พ.ศ. 2531 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต (วท.บ.) สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2553 และเข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม.) สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2554