วิธีการเลือกฐานหลักสำหรับเทคนิคการลดมิติ

นายพีรธรรม วิริยธรรมภูมิ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2552

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

AN APPROACH TO BASIS SELECTION FOR DIMENSIONAL REDUCTION TECHNIQUES

Mr.Peratham Wiriyathammabhum

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering Program in Computer Engineering
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2009

Copyright of Chulalongkorn University

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

หัวข้อวิทยานิพนธ์	วิธีการเลือกฐานหลักสำหรับเทคนิคการลดมิติ
โดย	นายพีรธรรม วิริยธรรมภูมิ
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

..... กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.ชลวิช นัทธี)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

พีรธรรม วิริยธรรมภูมิ : วิธีการเลือกฐานหลักสำหรับเทคนิคการลดมิติ. (AN APPROACH TO BASIS SELECTION FOR DIMENSIONAL REDUCTION TECHNIQUES) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ศ.ดร.บุญเสริม กิจศิริกุล, 82 หน้า.

การลดมิติถือเป็นงานสำคัญในงานการเรียนรู้ของเครื่องและการรู้จำ เนื่องจากในปัจจุบันนี้เรามีความสามารถในการเก็บข้อมูลทุกๆคุณลักษณะที่เราพอจะนึกถึงได้ ซึ่งนำไปสู่ข้อมูลขนาดใหญ่ที่มีคุณลักษณะมากเกินกว่าที่จะประมวลผลได้ เทคนิคการลดมิติช่วยลดความซ้ำซ้อนระหว่างคุณลักษณะและช่วยขจัดความห่างมากในชุดข้อมูล การเลือกฐานหลักถือเป็นปัญหาเปิดว่าเราจะสามารถลดมิติเพื่อให้ได้เซตย่อยของข้อมูลที่เหมาะที่สุดอย่างอัตโนมัติได้อย่างไร ในงานรู้จำบนชุดข้อมูลรูปภาพนั้นเทคนิคการลดมิติได้รับการพิสูจน์ว่าเป็นตัวประมวลผลก่อนที่สำคัญ เทคนิคแบบสองมิติและสองมิติสองทิศทางเป็นส่วนขยายของเทคนิคการลดมิติสำหรับชุดข้อมูลรูปภาพโดยเฉพาะ เทคนิคแบบสองมิติและสองมิติสองทิศทางให้ค่าความแม่นที่สูงกว่าและมีปริมาณการคำนวณที่น้อยกว่าเมื่อเทียบกับเทคนิคดั้งเดิม เราได้โต้แย้งว่ากระบวนการเลือกฐานหลักในเทคนิคแบบสองมิตินั้นไม่เหมาะที่สุดและไม่ชัดเจนในเทคนิคแบบสองมิติสองทิศทาง เราได้นำเสนอเทคนิคตัวกรองแบบคะแนนเพื่อช่วยทำให้การเลือกฐานหลักเหมาะที่สุดโดยการทำให้เป็นบรรทัดฐานและจัดเรียงฐานหลักใหม่ที่ให้ผลเป็นค่าความแม่นที่สูงขึ้นในการจำแนก นอกจากนี้เรายังนำเสนอวิธีการเลือกฐานหลักใหม่ในเทคนิคแบบสองมิติสองทิศทางสองวิธีได้แก่ วิธีใช้เทคนิคคัดเลือกคุณลักษณะแบบผสมและวิธีอัลกอริทึมเชิงละโมบที่ใช้ควบคู่กับการตัดออก ซึ่งทั้งสองวิธีนี้จะใช้ปริมาณการคำนวณน้อยกว่าวิธีดั้งเดิมที่ใช้การตรวจสอบไขว้ทั้งหมด อีกทั้งยังให้ค่าความแม่นในการจำแนกสูงกว่าวิธีตัดออกโดยอ้างอิงค่าลักษณะเฉพาะที่มักจะติดในหุบเขาของปริภูมิค่าความแม่น เราได้ประยุกต์ใช้แนวคิดกับงานในโลกจริงได้แก่ งานรู้จำใบหน้าและงานรู้จำลายมือ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต นายพีรธรรม

สาขาวิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก

ปีการศึกษา 2552

5270424521 : MAJOR COMPUTER ENGINEERING

KEYWORDS : DIMENSIONAL REDUCTION / SUBSPACE METHOD / FACE RECOGNITION / HAND WRITTEN RECOGNITION

PERATHAM WIRIYATHAMMABHUM : AN APPROACH TO BASIS SELECTION FOR DIMENSIONAL REDUCTION TECHNIQUES. THESIS ADVISOR : PROF. BOONSERM KIJSIRIKUL, Ph.D., 82 pp.

Dimensional Reduction is an important task in Machine Learning and Pattern Recognition. Nowadays, we have gained an ability to collect every feature in our imaginable aspect which leads to massive data that have too many features to process. Dimensional Reduction techniques reduce the redundancy among features and conquer the sparseness of the data set. Basis Selection is an open problem of how much we should reduce the dimension to be an optimal dataset automatically. In Pattern Recognition tasks of image data sets, techniques for Dimensional Reduction have been proven to be an important preprocessing step. 2D and $(2D)^2$ techniques are the extensions of Dimensional Reduction techniques tailored for image data sets. 2D and $(2D)^2$ techniques provide more classification accuracy with less computational cost compared to original techniques. We argue that the step in Basis Selection of 2D techniques is not optimal and the method of Basis Selection is not obvious in $(2D)^2$ techniques. We propose the score-based filter techniques to further optimize the Basis Selection by performing a normalizing and reordering step which results in higher classification accuracy. Moreover, we also introduce a novel Basis Selection method in $(2D)^2$ techniques based on a greedy algorithm and score metrics which requires less computational cost compared to cross validation based techniques and provides more classification accuracy than eigenvalue-based cut-off techniques which stuck in the valley of accuracy space. We have applied our proposed method to several real world applications, face recognition and hand written recognition.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

Department : Computer Engineering Student's Signature นาย พีรธรรม

Field of Study : Computer Engineering Advisor's Signature

Academic Year : 2009

กิตติกรรมประกาศ

ในความสำเร็จลุล่วงในวิทยานิพนธ์นี้ ข้าพเจ้าขอกราบขอบพระคุณท่านศาสตราจารย์ ดร. บุญเสริม กิจศิริกุล ผู้เป็นอาจารย์ที่ปรึกษา ผู้ให้ความรู้ ผู้คอยแนะนำและชี้แนะแนวทางอันเป็นประโยชน์ ผู้ช่วยเหลือดูแลและเป็นกำลังใจตลอดมา

ขอขอบคุณภาควิชาวิศวกรรมศาสตร์คอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ขอขอบคุณคณะวิศวกรรมศาสตร์สำหรับทุนยกเว้นค่าเล่าเรียนสำหรับนิสิตเก่าชั้นปริญญาบัณฑิตที่ต้องการศึกษาต่อในชั้นมหาบัณฑิตปีการศึกษา 2552

ข้าพเจ้าขอกราบขอบพระคุณบิดามารดาผู้มีพระคุณ ผู้ให้การสนับสนุนและกำลังใจในการดำเนินงานนี้โดยตลอด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฎ

สารบัญภาพ ..... ฏ

บทที่ 1 1

บทนำ 1

1.1 ปัญหาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ 3

1.3 ขอบเขตการดำเนินงาน 4

1.4 ขั้นตอนการดำเนินงาน 4

1.5 ประโยชน์ที่คาดว่าจะได้รับ 4

1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์ 5

บทที่ 2 6

ทฤษฎีที่เกี่ยวข้อง 6

2.1 เทคนิคประเภทคัดเลือกคุณลักษณะ (Feature Selection) 6

2.1.1 เทคนิคประเภทตัวกรอง (Filter) 6

2.1.2 เทคนิคประเภทตัวหุ้ม (Wrapper) 7

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.1.3 เทคนิคประเภทผสม (Hybrid) 8

2.1.4 เทคนิคประเภทฝังตัว (Embedded) 8

2.1.5 ขั้นตอนร่วมในเทคนิคประเภทคัดเลือกคุณลักษณะ (General Procedure for Feature Selection) 8

2.2 เทคนิคประเภทสกัดคุณลักษณะ (Feature Extraction) 9

2.2.1 เทคนิคประเภทสกัดคุณลักษณะแบบไม่สอน (Unsupervised Feature Extraction Techniques) 10

2.2.1.1 เทคนิคแบบคอนเวกซ์ 11

2.2.1.1 เทคนิคแบบไม่คอนเวกซ์ 11

2.2.2 เทคนิคประเภทสกัดคุณลักษณะแบบสอน (Supervised Feature Extraction Techniques) 12

2.2.2.1 เปรียบเทียบเทคนิคประเภทสกัดคุณลักษณะแบบสอนและเทคนิคประเภทสกัดคุณลักษณะแบบไม่สอน 12

2.2.2.2 ตัวอย่างของเทคนิคประเภทสกัดคุณลักษณะแบบสอน 14

2.2.3 การแบ่งเทคนิคประเภทสกัดคุณลักษณะตามความเป็นเชิงเส้น 15

บทที่ 3 17

งานวิจัยที่เกี่ยวข้อง 17

3.1 การวิเคราะห์องค์ประกอบสำคัญหรือพีซีเอ (PCA: Principal Component Analysis) 17

3.2 การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นหรือแอลดีเอ (LDA: Linear Discriminant Analysis) 21

3.3 การฉายรักษาสมบัติเฉพาะที่หรือแอลพีพี (LPP: Locality Preserving Projection) 25

3.4 ตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์หรือดีเอ็นอี (DNE: Discriminative Neighborhood Embedding) 28

3.5 การวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่หรือแอลเอสดีเอ (LSDA: Locality Sensitive Discriminant Analysis).....30

3.6 เทคนิคการลดมิติแบบสองมิติ (2D Methods) 33

3.6.1 การวิเคราะห์องค์ประกอบสำคัญแบบสองมิติหรือพีซีเอแบบสองมิติ (2DPCA: 2-D Principal Component Analysis).....33

3.6.2 การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติหรือแอลดีเอแบบสองมิติ (2DLDA: 2-D Linear Discriminant Analysis).....36

3.6.3 การฉายรักษาสมบัติเฉพาะที่แบบสองมิติหรือแอลพีพีแบบสองมิติ (2DLPP: 2-D Locality Preserving Projection) 38

3.6.4 ตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติหรือดีเอ็นอีแบบสองมิติ (2DDNE: 2-D Discriminant Neighborhood Embedding).....40

3.6.5 การวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติหรือแอลเอสดีเอแบบสองมิติ (2DLSDA: 2-D Locality Sensitive Discriminant Analysis) 41

3.7 เทคนิคประเภทคัดเลือกคุณลักษณะที่เกี่ยวข้อง.....42

3.7.1 คะแนนของฟิชเชอร์ (Fisher Score) 42

3.7.2 คะแนนลาปลาเซียน (Laplacian Score) 43

3.8 การแยกไบแอส-แวเรียนซ์ (Bias-Variance Decomposition).....44

3.9 อัลกอริทึมเพื่อนบ้านใกล้สุด *k* ตัว (k-Nearest Neighbors).....45

บทที่ 4 47

การเลือกฐานหลักสำหรับเทคนิคแบบสองมิติและสองมิติสองทิศทางด้วย.....47

เทคนิคการคัดเลือกคุณลักษณะ 47

4.1 แรงจูงใจ.....47

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

4.2 แนวคิด 49

4.3 การปรับปรุงเทคนิคแบบสองมิติสองทิศทาง 51

4.4 การปรับปรุงเทคนิคแบบสองมิติ 52

4.5 การเลือกฐานหลักและประมาณมิติเป้าหมาย 54

บทที่ 5 60

ผลการทดลอง 60

5.1 ชุดข้อมูล 60

5.2 การทดลองและผลการทดลอง 62

บทที่ 6 74

สรุปผลการวิจัยและข้อเสนอแนะ 74

6.1 สรุปผลการวิจัย 74

6.2 ข้อเสนอแนะ 75

รายการอ้างอิง 76

ประวัติผู้เขียนวิทยานิพนธ์ 82

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่ 1 ขั้นตอนวิธีการเลือกฐานหลักในเทคนิคแบบสองมิติแบบสองทิศทาง 52

ตารางที่ 2 ขั้นตอนวิธีการเลือกฐานหลักในเทคนิคแบบสองมิติ 53

ตารางที่ 3 ขั้นตอนวิธีในการประมาณมิติเป้าหมายด้วยเทคนิคคัดเลือกคุณลักษณะประเภทผสม ..55

ตารางที่ 4 ขั้นตอนวิธีในการประมาณมิติเป้าหมายด้วยอัลกอริทึมเชิงละโมบและการตัดออก.....59

ตารางที่ 5 ผลลัพธ์ค่าความแม่นแบบไม่กำหนดมิติเป้าหมายของเทคนิคแบบสองมิติสองทิศทาง..63

ตารางที่ 6 ผลลัพธ์ค่าความแม่นแบบไม่กำหนดมิติเป้าหมายของเทคนิคแบบสองมิติ 65

ตารางที่ 7 ผลลัพธ์ค่าความแม่นเมื่อกำหนดมิติเป้าหมายของเทคนิคแบบสองมิติสองทิศทาง 67

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

รูปที่ 1 ข้อมูลหลายตัวแปร 3 มิติที่ต้องการจำแนก....2

รูปที่ 2 ตัวอย่างการลดมิติข้อมูลหลายตัวแปรจาก 3 มิติลงมาเป็น 1 มิติ....2

รูปที่ 3 แผนภาพขั้นตอนการดำเนินงาน....4

รูปที่ 4 อนุกรมวิธานของเทคนิคประเภทคัดเลือกคุณลักษณะ....7

รูปที่ 5 ขั้นตอนร่วมของเทคนิคประเภทคัดเลือกคุณลักษณะ....9

รูปที่ 6 อนุกรมวิธานของเทคนิคประเภทสกัดคุณลักษณะ....10

รูปที่ 7 ตัวอย่างความแตกต่างของเทคนิคสำหรับการลดมิติประเภทสกัดลักษณะแบบสอนและไม่สอน....13

รูปที่ 8 กรณีที่พีซีเอเหมาะสมในการช่วยจำแนกมากกว่าแอลดีเอ....13

รูปที่ 9 ผลลัพธ์จากการสเกลข้อมูลด้วยเมตริกซ์ระยะทาง....14

รูปที่ 10 กรณีข้อมูลมีการกระจายตัวแบบหลายฐานซึ่งการเรียนรู้เมตริกซ์ระยะทางของมหาลาโนบิสโดยทั่วไปทำได้ยาก....15

รูปที่ 11 การประมาณข้อมูลด้วยเส้นตรงของ Pearson....18

รูปที่ 12 การส่งไปยังปริภูมิอุดมคติสำหรับการจำแนก....22

รูปที่ 13 จุดข้อมูลถูกดูดและผลักตามเพื่อนบ้านใกล้สุดที่อยู่ในประเภทเดียวกันและเพื่อนบ้านใกล้สุดที่อยู่ต่างประเภทกันตามลำดับ....28

รูปที่ 14 แนวคิดของเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่....31

รูปที่ 15 บริเวณจำแนกของกฎเพื่อนบ้านใกล้สุด 1 ตัวเป็นเทสเซลเลชันแบบโวโรนอยด์....46

รูปที่ 16 กระบวนการใช้เทคนิคประเภทคัดเลือกคุณลักษณะเข้าร่วมในเทคนิคแบบสองมิติแบบสองทิศทาง....51

รูปที่ 17 กระบวนการใช้เทคนิคประเภทคัดเลือกคุณลักษณะเข้าร่วมในเทคนิคแบบสองมิติ.....53

รูปที่ 18 ปริภูมิค่าความแม่นมีความกวัดแกว่งมาก.....55

รูปที่ 19 ตัวอย่างค่าลักษณะเฉพาะของเมตริกซ์การฉายแนวแถว (ซ้าย) และแนวสดมภ์ (ขวา) จากเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง 56

รูปที่ 20 ตัวอย่างค่าคะแนนของเมตริกซ์การฉายแนวแถว (ซ้าย) และแนวสดมภ์ (ขวา) จากเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทางที่นำเสนอ.....57

รูปที่ 21 คะแนนของฟิชเชอร์ที่เปลี่ยนไปในกระบวนการทำให้เป็นบรรทัดฐานได้แก่ คะแนนดั้งเดิม (แถวแรก) หลังทำให้เป็นบรรทัดฐาน 1 ครั้ง (แถวที่สอง) หลังทำให้เป็นบรรทัดฐาน 2 ครั้ง (แถวที่สาม) ของคะแนนจากเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทางที่นำเสนอ 58

รูปที่ 22 ตัวอย่างใบหน้าจากฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล 60

รูปที่ 23 ตัวอย่างใบหน้าจากฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล 61

รูปที่ 24 ตัวอย่างลายมือจากฐานข้อมูลไบนารีอัลฟา.....61

รูปที่ 25 ตัวอย่างลายมือตัวเลขจากฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นเอสไอที 62

รูปที่ 26 ค่าความผิดพลาดทั่วไป (แถวบน) ไบแอส (แถวกลาง) แวเรียนซ์ (แถวล่าง) ที่ประมาณด้วยวิธีของของ Bauer และ Kohavi สำหรับเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) ในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (สดมภ์ซ้าย) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (สดมภ์ขวา).....70

รูปที่ 27 ค่าความผิดพลาดทั่วไป (แถวบน) ไบแอส (แถวกลาง) แวเรียนซ์ (แถวล่าง) ที่ประมาณด้วยวิธีของของ Bauer และ Kohavi สำหรับเทคนิคการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) ในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (สดมภ์ซ้าย) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (สดมภ์ขวา).....71

รูปที่ 28 ค่าความผิดพลาดทั่วไป (แถวบน) ไบแอส (แถวกลาง) แวเรียนซ์ (แถวล่าง) ที่ประมาณด้วยวิธีของของ Bauer และ Kohavi สำหรับเทคนิคตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติ

สองทิศทาง ($(2D)^2$DNE) ในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (สดมภ์ซ้าย) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (สดมภ์ขวา).....72

รูปที่ 29 ค่าความผิดพลาดทั่วไป (แถวบน) ไบแอส (แถวกลาง) แวเรียนซ์ (แถวล่าง) ที่ประมาณด้วยวิธีของของ Bauer และ Kohavi สำหรับเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LSDA) ในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (สดมภ์ซ้าย) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (สดมภ์ขวา).....73

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ปัญหาและความสำคัญของปัญหา

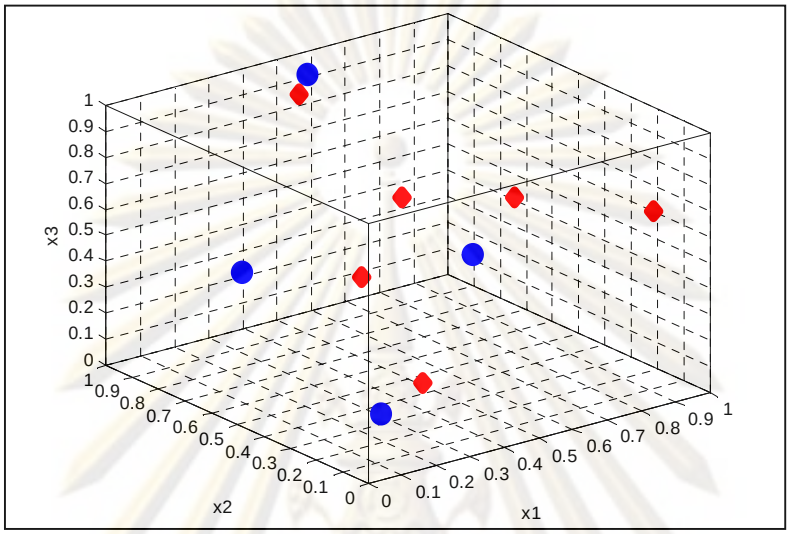
ในสังคมปัจจุบันมีแนวโน้มที่จะนำการประมวลผลของคอมพิวเตอร์ที่รวดเร็วและรองรับปริมาณข้อมูลมากๆได้ มาใช้กับงานต่างๆในชีวิตประจำวันมากขึ้น เพื่อการอำนวยความสะดวกในการคิดและตรวจสอบทางวิชาการต่างๆ รวมถึงใช้ทำงานแทนมนุษย์ อย่างไรก็ตามคอมพิวเตอร์ในปัจจุบันยังมีความเร็วในการประมวลผลที่จำกัดไม่สามารถรับงานที่เป็นข้อมูลขนาดใหญ่ (Massive Data) ที่มีจำนวนข้อมูลหรือมีมิติมากๆ อันเป็นผลจากความสามารถในการเก็บข้อมูลได้อย่างมีประสิทธิภาพ ในงานวิจัยนี้จะเน้นศึกษาวิธีที่ช่วยให้คอมพิวเตอร์สามารถประมวลผลข้อมูลที่มีมิติมากๆได้ดีขึ้น

ข้อมูลมิติสูงๆนั้นสามารถพบได้มากในชีวิตประจำวัน ยกตัวอย่างเช่นข้อมูลรูปภาพ ข้อมูลของยีน ข้อมูลจากตัวรับรู้ (Sensor Data) ข้อมูลทางการเงิน ข้อมูลเสียงพูด เป็นต้น ข้อมูลเหล่านี้มักจะมีจำนวนข้อมูลไม่เพียงพอต่อจำนวนมิติ ทำให้ข้อมูลมีความห่างมาก (Sparsity) และเมื่อเพิ่มมิติปริมาตรของข้อมูลจะเพิ่มขึ้นแบบชี้กำลัง เป็นเหตุให้การวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลทำได้ยาก ใช้เวลานาน และไม่มีประสิทธิภาพหากมีจำนวนมิติมากเกินไป ปัญหานี้เรียกว่าปัญหาคำสาปของมิติ (Curse of Dimensionality)

การลดมิติ (Dimensional Reduction) ช่วยแก้ปัญหาคำสาปของมิติ โดยจะลดจำนวนมิติหรือตัวแปรที่ซ้ำซ้อนที่ไม่จำเป็นต่อการวิจัยทั้งเนื่องจากสัญญาณรบกวน (noise) หรือตัวแปรนั้นไม่มีนัยสำคัญ ทำให้เกิดการลดรูปของปัญหาจากข้อมูลที่เล็กลง และยังคงความแม่นจากตัวแปรที่จำเป็นต่อการวิเคราะห์อย่างครบถ้วน ทำให้ประสิทธิภาพสูงขึ้นทั้งด้านความเร็วในการประมวลผลและความแม่นในการวิเคราะห์เมื่อเทียบกับชุดข้อมูลเดิม แต่มีจำนวนมิติหรืออาจเรียกว่าคุณลักษณะ (Feature) น้อยลง

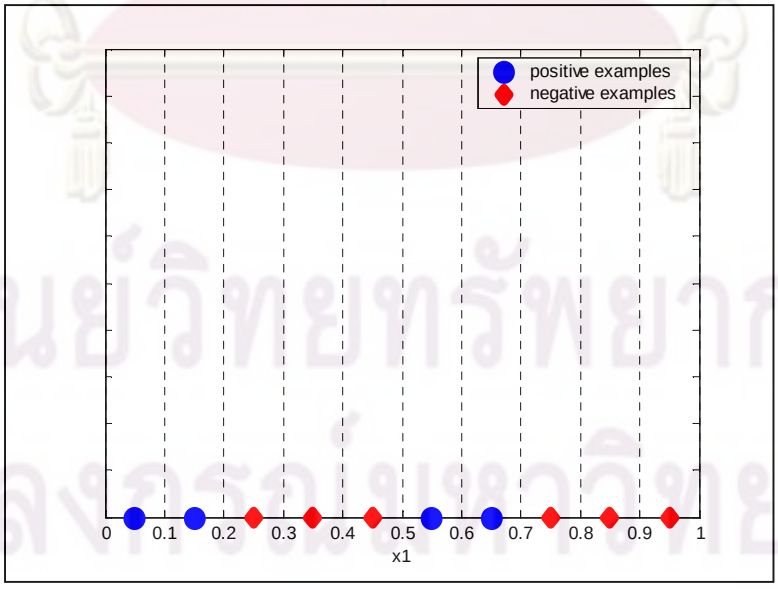
ตัวอย่างของงานที่การลดมิติช่วยในการประมวลผล ในที่นี้ขอยกตัวอย่างในปัญหาการจำแนก สมมติว่ามีฟาง 2 สี สีดำกับสีขาวที่นำมามัดรวมกันและมีความต้องการจำแนกฟาง 2 สีออกจากกันโดยให้ชี้สีของฟางให้ถูกต้อง การจำแนกจะทำได้ยากใน 3 มิติที่จะชี้ว่าฟางชิ้นไหนเป็นสีอะไร แต่หากมองที่ปลายด้านหนึ่งที่มี 2 มิติโดยจะเห็นแต่ปลายฟาง จะสามารถจำแนกได้ง่ายขึ้นมาก ในกรณีนี้ทำให้เห็นว่าการเปลี่ยนมุมมองมาที่มิติที่เล็กลงสามารถช่วยให้เกิดการลดรูปของปัญหาได้จริง

อีกตัวอย่างหนึ่งดังรูปที่ 1 มีข้อมูลหลายตัวแปร (Multivariate) ที่อยู่ในปริภูมิ 3 มิติและมีลักษณะคล้ายกลุ่มเมฆของจุด (Point Cloud) โดยมี 2 ประเภท (Class) ได้แก่สีแดงและสีน้ำเงิน โดยจะเห็นว่าข้อมูลแต่ละจุดมีความห่างมาก และเป็นการยากที่แบ่งข้อมูลเป็น 2 ประเภทให้ชัดเจนได้



รูปที่ 1 ข้อมูลหลายตัวแปร 3 มิติที่ต้องการจำแนก

แต่เมื่อเราลดมิติแล้ว จะเห็นว่าข้อมูลนั้นสามารถจำแนกได้ง่ายแม้ด้วยตาเปล่า ดังรูปที่ 2 แสดงว่าแท้จริงแล้วเราสามารถวิเคราะห์ข้อมูลได้ใน 1 มิติ ไม่มีความจำเป็นที่จะต้องสิ้นเปลืองปริมาณการคำนวณไปกับข้อมูลที่มีมิติสูงกว่าเลย



รูปที่ 2 ตัวอย่างการลดมิติข้อมูลหลายตัวแปรจาก 3 มิติลงมาเป็น 1 มิติ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การลดมิตินั้นได้รับการพิสูจน์ว่าช่วยในการเรียนรู้ของเครื่องและการประยุกต์ใช้หลายอย่างเช่น ปัญหาการจำแนกประเภท (Classification) การสร้างจินตทัศน์ (Visualization) การบีบอัดข้อมูล (Data Compression) เป็นต้น

ทั้งนี้การลดมิติมักจะถูกกล่าวถึงควบคู่ไปกับการเรียนรู้ข้อมูลท่อร่วม (Manifold Learning) ซึ่งเป็นการเรียนรู้ข้อมูลที่แท้จริงที่อยู่ในรูปของรูปแบบที่ซ่อนอยู่ (Hidden Pattern) ในมิติในตัว (Intrinsic Dimension) แนวคิดนี้สามารถยกตัวอย่างได้จากกระดาษทั่วไป โดยที่กระดาษนั้นจะอยู่ในโลกจริงที่มี 3 มิติประกอบด้วยความกว้าง ความยาว และความหนา ทำให้เมื่อเราเก็บข้อมูลกระดาษมาจากโลกจริงข้อมูลที่ได้จะมีลักษณะเป็น 3 มิติไปด้วย โดยจะเกิดกรณีพิเศษเป็นกรณีที่ไม่เป็นเชิงเส้นได้เช่นกระดาษม้วนหรือกระดาษเป็นคลื่นซึ่งไม่ใช่รูปร่างแท้จริงของกระดาษที่แท้จริงแล้วกระดาษมีมิติเพียง 2 มิติเท่านั้นคือความกว้างและความยาว การเรียนรู้รูปร่างแท้จริงของกระดาษก็เปรียบได้กับการเรียนรู้ข้อมูลท่อร่วมนั่นเอง ทั้งนี้จะถือว่ามิติที่เกินมาเป็นมิติที่ไม่จำเป็นต่อการวิเคราะห์สามารถตัดทิ้งไปได้ ทำให้เหลือแต่มิติที่เป็นตัวแปรที่จำเป็นต่อการวิเคราะห์จริงๆ เท่านั้น ซึ่งโดยทั่วไปมิติดังกล่าวจะอยู่บนข้อมูลท่อร่วมจึงเป็นจุดร่วมของการเรียนรู้ข้อมูลท่อร่วมและการลดมิตินั่นเอง

การลดมิติทั่วไปจะมีสัญกรณ์ดังนี้ กำหนดให้มีชุดข้อมูล X ประกอบด้วยเวกเตอร์ x_i ตามแนวแถวหรือแนวสดมภ์ $[x_1,x_2,\ldots,x_n]$ อยู่ในมิติขนาด D ซึ่ง D มีค่ามากๆ ต้องการหาเมตริกซ์การฉาย (Projection Matrix) P ที่จะส่งชุดข้อมูล X ไปยังชุดข้อมูลใหม่ซึ่งอยู่บนข้อมูลท่อร่วม Z ที่มีมิติต่ำๆ d โดยทั่วไปมักใช้สมมติฐานว่า d มีค่าน้อยกว่า D มากๆ ($d << D$) ในการลดมิตินั้นจะใช้หลักการของการฉาย โดยทั่วไปจะลดมิติด้วยสมการ $Z = P^TX$ อย่างไรก็ตามการลดมิตินั้นเราไม่ทราบว่าค่า d ที่ควรจะเป็นนั้นมีค่าเท่าใดและเราก็ไม่ทราบลักษณะของข้อมูลท่อร่วมที่จะเกิดขึ้นมาก่อนเช่นกัน

1.2 วัตถุประสงค์

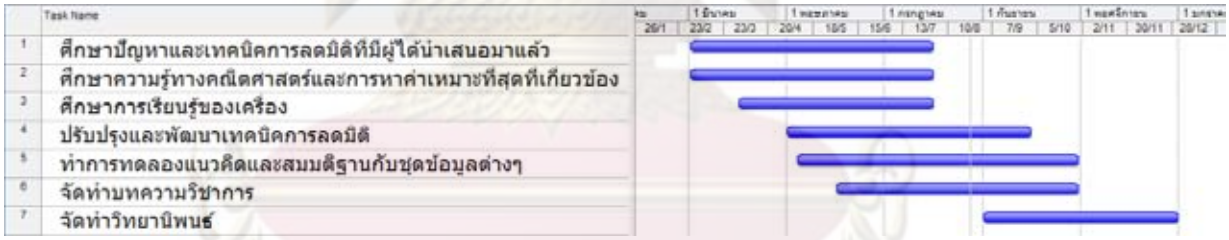
- เพื่อศึกษาแนวคิดและเทคนิคการลดมิติซึ่งเป็นหัวข้อที่น่าสนใจในปัจจุบันเพื่อนำไปใช้ให้เกิดประโยชน์
- เพื่อปรับปรุงเทคนิคประเภทสองมิติสองทิศทางให้มีประสิทธิภาพอย่างที่ควรจะเป็น
- เพื่อปรับปรุงเทคนิคสำหรับการลดมิติที่ยังคงมีข้อจำกัดบางประการให้มีประสิทธิภาพสูงขึ้นเช่น ทำงานได้รวดเร็วขึ้น มีความแม่นมากขึ้น เหมาะสมต่อการใช้งานจริง

1.3 ขอบเขตการดำเนินงาน

- ศึกษาและทำการปรับปรุงเทคนิคแบบไม่สอนหรือแบบสอน
- ในการทดสอบเทคนิคการลดมิติสำหรับการประมวลผลภาพจะทดสอบบนชุดข้อมูลใบหน้าเป็นหลัก และใช้วิธีการตรวจสอบไขว้ 10 พับ (10-Fold Cross Validation)
- จะใช้ชุดข้อมูลใบหน้าในการทดสอบโดยจะใช้อย่างน้อย 2 ชุดข้อมูลได้แก่ ฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล

1.4 ขั้นตอนการดำเนินงาน

1. ศึกษาปัญหาและเทคนิคการลดมิติที่มีผู้ได้นำเสนอมาแล้ว
2. ศึกษาความรู้ทางคณิตศาสตร์และการหาค่าเหมาะที่สุดที่เกี่ยวข้อง
3. ศึกษาการเรียนรู้ของเครื่อง
4. ปรับปรุงและพัฒนาเทคนิคการลดมิติ
5. ทำการทดลองแนวคิดและสมมติฐานกับชุดข้อมูลต่างๆ
6. จัดทำบทความวิชาการ
7. จัดทำวิทยานิพนธ์



รูปที่ 3 แผนภาพขั้นตอนการดำเนินงาน

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- ได้ศึกษาเทคนิคสำหรับการลดมิติซึ่งเป็นหัวข้อวิจัยที่เป็นที่สนใจอย่างกว้างขวางในปัจจุบัน และเป็นการเผยแพร่ความรู้เหล่านี้สู่สังคมไทยผ่านทางวิทยานิพนธ์และสิ่งตีพิมพ์
- เป็นการศึกษาการประมวลผลข้อมูลขนาดใหญ่ซึ่งเหมาะสมในยุคปัจจุบันที่เป็นยุคของข้อมูล

- เป็นการขยายขอบเขตของการใช้เทคนิคสำหรับการลดมิติให้ใช้ได้อย่างมีประสิทธิภาพสูงขึ้นทั้งชุดข้อมูลที่มีขนาดเล็กและใหญ่
- เป็นการปรับปรุงเทคนิคสำหรับการรู้จำใบหน้าซึ่งมีความจำเป็นในระบบรักษาความปลอดภัยในปัจจุบันให้มีประสิทธิภาพสูงขึ้น

1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตีพิมพ์เป็นบทความทางวิชาการจำนวน 2 บทความ ดังนี้

- บทความในงานประชุมวิชาการระดับนานาชาติ

P. Wiriyathammabhum and B. Kijsirikul. Basis selection for 2DLDA-based face recognition using fisher score. In Proc. International Conference on Neural Information Processing. (2009)

- บทความในงานประชุมวิชาการระดับชาติ

พีรธรรม วิริยธรรมภูมิ และ บุญเสริม กิจศิริกุล. การเลือกฐานหลักสำหรับเทคนิคในกลุ่มการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสำหรับการรู้จำใบหน้าด้วยคะแนนของฟิชเชอร์. การประชุมวิชาการวิทยาการคอมพิวเตอร์และวิศวกรรมคอมพิวเตอร์แห่งชาติ. (2009)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

เทคนิคสำหรับการลดมิตินั้นสามารถแบ่งได้เป็น 2 ประเภทใหญ่ๆ คือเทคนิคประเภทคัดเลือกคุณลักษณะ (Feature Selection) และเทคนิคประเภทสกัดคุณลักษณะ (Feature Extraction)

2.1 เทคนิคประเภทคัดเลือกคุณลักษณะ (Feature Selection)

เทคนิคประเภทคัดเลือกคุณลักษณะนั้นเป็นที่ได้รับการสนใจมาก่อนเทคนิคประเภทสกัดคุณลักษณะ และได้รับการพิสูจน์ว่าทำให้ตัวจำแนกใช้งานได้ดีขึ้นโดยทำให้เร็วขึ้นและมีความแม่นยำสูงขึ้นเมื่อใช้ในการประมวลผลก่อน (Preprocessing) เทคนิคประเภทนี้ช่วยในงานที่สำคัญต่างๆ เช่น การจำแนกเว็บไซต์ การทำเหมืองข้อความ (Text Mining) และการรู้จำเสียงพูด (Speech Recognition) นอกจากนี้ยังมีการนำไปประยุกต์ใช้จนได้ผลดีในศาสตร์อีกหลายแขนงเช่น ในด้านดาราศาสตร์ (Astronomy)[1] ชีวสารสนเทศศาสตร์ (Bioinformatics)[2] และการตรวจจับการบุกรุก (Intrusion Detection)[3] อีกด้วย

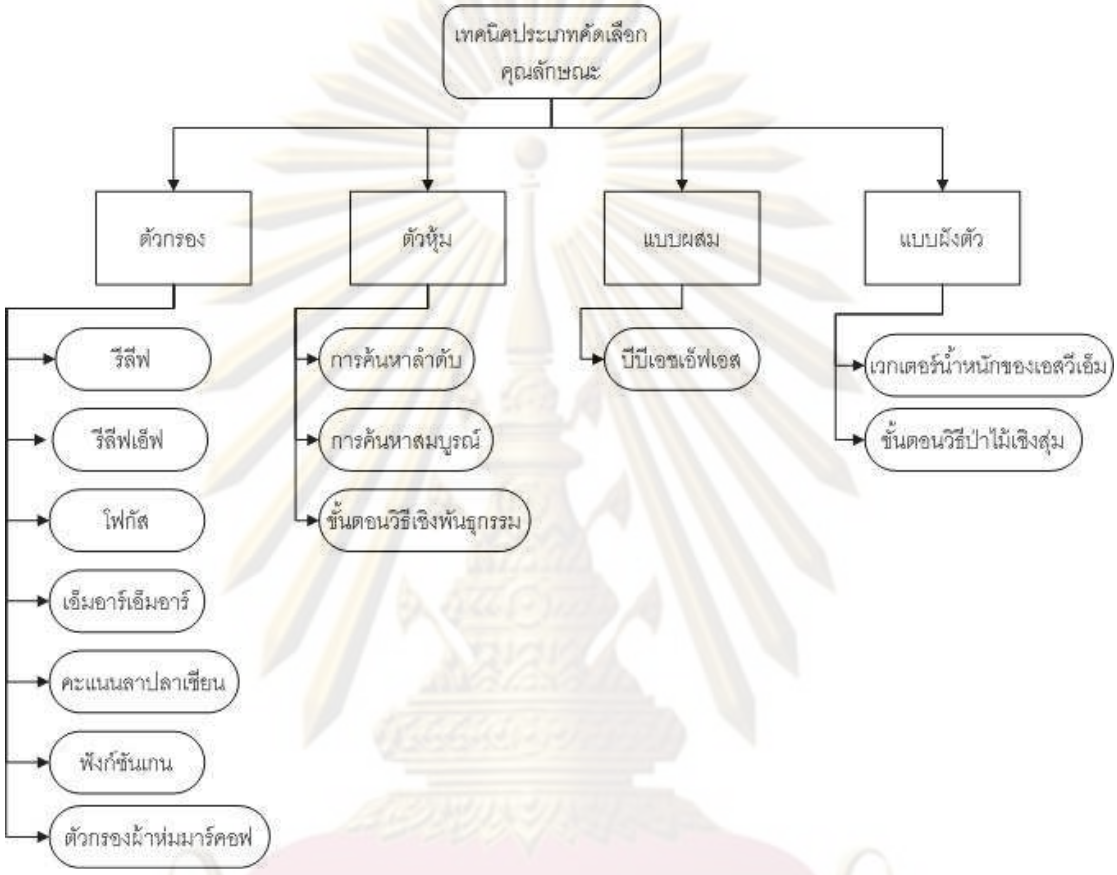
เทคนิคประเภทคัดเลือกลักษณะจะเลือกคุณลักษณะ (Feature) ให้ได้เซตย่อยที่เล็กที่สุดและตัดคุณลักษณะที่ซ้ำซ้อน (Redundant Feature) ออกให้เหลือแต่คุณลักษณะที่จำเป็น เนื่องจากคุณลักษณะที่ซ้ำซ้อนไม่ได้ให้แนวคิดการเรียนรู้เพิ่มเติม สามารถตัดทิ้งไปได้เพื่อให้เวลาการทำงานลดลง เทคนิคประเภทนี้ยังสามารถแบ่งย่อยได้อีก 3 ประเภท[4]ตามลักษณะของการเลือกเซตย่อย ได้แก่ประเภทตัวกรอง (Filter) ประเภทตัวหุ้ม (Wrapper) และประเภทผสม (Hybrid)

ทั้งนี้เราสามารถเขียนสรุปเทคนิคประเภทคัดเลือกคุณลักษณะออกมาเป็นอนุกรมวิธานของได้ดังรูปที่ 4

2.1.1 เทคนิคประเภทตัวกรอง (Filter)

เทคนิคประเภทตัวกรองนั้นจะกรองคุณลักษณะที่ไม่จำเป็นออกไปโดยจะใช้มาตรวัดที่เป็นอิสระต่อตัวเรียนรู้เช่น มาตรวัดระยะทาง (Distance Measure) มาตรวัดสหสัมพันธ์ (Correlation Measure) มาตรวัดความต้องกัน (Consistency Measure) เป็นต้น เพื่อวัดค่าความเหมาะสมของคุณลักษณะที่ควรจะเก็บไว้ ตัวอย่างเทคนิคประเภทตัวกรองเช่น รีลีฟ (Relief)[5] รีลีฟเอ็ฟ (Relief-F)[6] โฟกัส (FOCUS)[7] ฟังก์ชันเกน (Information Gain) คะแนนลาปลาเซียน (Laplacian Score)[8] ตัวกรองผ้าห่มมาร์คอฟ (Markov Blanket Filter)[9] เอ็มอาร์เอ็มอาร์ (mRMR: minimum-

Redundancy-Maximum-Relevance)[10] เป็นต้น เทคนิคประเภทนี้มักจะเร็ว รองรับข้อมูลขนาดใหญ่ๆ คำนวณครั้งเดียวและไม่ซับซ้อนในการคำนวณ นอกจากนี้ยังเป็นเทคนิคที่ไม่ขึ้นกับตัวเรียนรู้อีกด้วย แต่ก็นับเป็นข้อเสียเช่นกันที่อาจทำให้ขาดการปฏิสัมพันธ์กับตัวเรียนรู้ไป



รูปที่ 4 อนุกรมวิธานของเทคนิคประเภทคัดเลือกคุณลักษณะ

2.1.2 เทคนิคประเภทตัวหุ้ม (Wrapper)

เทคนิคประเภทตัวหุ้มจะใช้แนวคิดการเลือกตัวแบบ (Model Selection) และการตรวจสอบสมมติฐาน (Hypothesis Testing) โดยจะนำมาใช้ในการหาเซตย่อยของคุณลักษณะในลักษณะของการแบ่งชุดข้อมูลเป็นชุดฝึกและชุดทดสอบสำหรับการตรวจสอบสมมติฐานด้วยวิธีต่างๆเช่น การตรวจสอบไขว้ k พับ (K-Fold Cross Validation) การตรวจสอบไขว้แบบดึงออกหนึ่งตัว (Leave-one-out Cross Validation) แล้วจึงให้คะแนนตามค่าความแม่น (Accuracy) ต่อไป ทั้งนี้มักจะใช้เทคนิคการค้นหาเข้าร่วมด้วยเพื่อหาค่าความแม่นที่เหมาะที่สุดเช่น การค้นหาแบบแนวกว้างก่อน (Breadth First Search) การค้นหาเชิงละโมบ (Greedy Search) การค้นหาด้วยขั้นตอนวิธีอบเหนียวจำลอง (Simulated Annealing) การค้นหาแบบสุ่ม (Randomized Search) เป็นต้น เทคนิคประเภทนี้

จะให้เซตย่อยที่น่าเชื่อถือที่สุดในด้านของความแม่นที่พึงจะมีได้ ไม่ขึ้นกับลักษณะพิเศษใดๆของคุณลักษณะ รวมถึงมีปฏิสัมพันธ์กับตัวเรียนรู้แต่ก็เป็นข้อเสียที่อาจทำให้ขึ้นกับตัวเรียนรู้เกินไปเช่นกัน อีกทั้งยังมีความเสี่ยงที่จะปรับเหมาะเกินไปกับชุดฝึก (Overfit) และเทคนิคนี้มีข้อเสียที่ช้าเนื่องจากการตรวจสอบสมมติฐานอีกด้วย ตัวอย่างของเทคนิคประเภทตัวหุ้มเช่น การค้นหาลำดับ (Sequential Search) การค้นหาสมบูรณ์ (Complete Search) ขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm) เป็นต้น

2.1.3 เทคนิคประเภทผสม (Hybrid)

เทคนิคประเภทผสมจะรวมเทคนิคประเภทตัวกรองและตัวหุ้มเข้าด้วยกัน โดยจะนำลำดับคะแนนที่ได้จากตัวกรองมาลดสเปซค้นหาของเทคนิคประเภทตัวหุ้มลง มีข้อดีคือจะเร็วกว่าเทคนิคประเภทตัวหุ้มและให้ผลลัพธ์ที่น่าพอใจมากกว่าเทคนิคประเภทตัวกรอง แต่อย่างไรก็ตามยังคงมีข้อเสียที่ขึ้นกับตัวเรียนรู้ที่ใช้ในการตรวจสอบสมมติฐานเหลืออยู่ ตัวอย่างของเทคนิคประเภทผสม เช่น บีบีเอชเอฟเอส (BBHFS: Boosting-Based Hybrid Feature Selection)[11] เป็นต้น

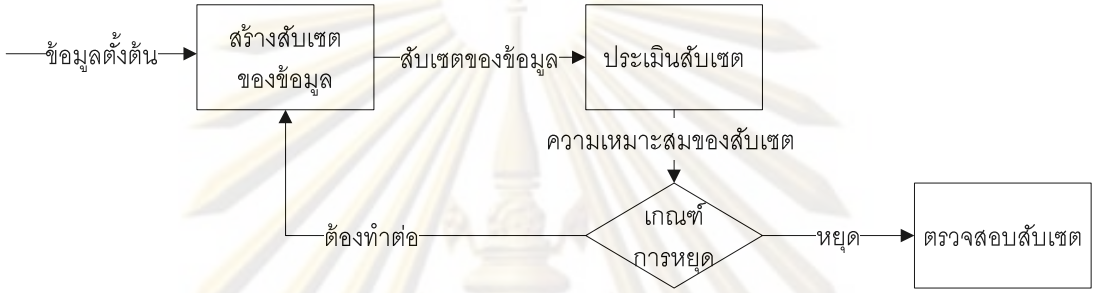
2.1.4 เทคนิคประเภทฝังตัว (Embedded)

เทคนิคประเภทฝังตัวจะมีลักษณะคล้ายกับเทคนิคประเภทตัวหุ้ม โดยจะหาเซตย่อยคุณลักษณะที่ดีที่สุดเช่นกัน แต่จะใช้แนวคิดของตัวเรียนรู้ซึ่งอาจเป็นตัวจำแนกเข้าร่วมในการหาเซตย่อยที่เหมาะที่สุด ทำให้สเปซการค้นหาไม่มากเท่าเทคนิคประเภทตัวหุ้ม อย่างไรก็ตามยังคงมีการแบ่งชุดข้อมูลเป็นชุดฝึกและชุดทดสอบสำหรับการตรวจสอบสมมติฐานอยู่ มีข้อดีคือจะเร็วกว่าเทคนิคประเภทตัวหุ้ม แต่ยังคงข้อเสียที่ประสิทธิภาพจะขึ้นกับตัวเรียนรู้ที่ใช้ในกระบวนการค้นหาเช่นกัน ตัวอย่างของเทคนิคประเภทฝังตัวเช่น การใช้เวกเตอร์น้ำหนักของซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ขั้นตอนวิธีป่าไม้เชิงสุ่ม (Random Forest) เป็นต้น

2.1.5 ขั้นตอนร่วมในเทคนิคประเภทคัดเลือกคุณลักษณะ (General Procedure for Feature Selection)

อย่างไรก็ตามเทคนิคทั้ง 3 ประเภทข้างต้นก็ยังคงมีขั้นตอนร่วม[12]ที่สามารถอธิบายได้ดังรูปที่ 5 โดยจะเริ่มจากการสร้างเซตย่อยของคุณลักษณะ (Subset Generation) ซึ่งจะแจกแจงเซตย่อยที่เป็นไปได้ออกมาโดยทั่วไปจะใช้วิธีการค้นหา (Searching Strategy) ซึ่งจะประกอบด้วย 2 ขั้นตอน คือการกำหนดจุดเริ่มต้นและการกำหนดวิธีการค้นหา จากนั้นจะนำเซตย่อยที่ได้มาประเมิน (Subset Evaluation) โดยจะใช้เกณฑ์ให้สามารถวัดค่าความเหมาะสมของเซตย่อยได้ ในขั้นตอนนี้อาจใช้

มาตรวัดที่เป็นอิสระต่อตัวเรียนรู้แบบที่ใช้ในเทคนิคประเภทตัวกรอง หรือจะใช้มาตรวัดที่ขึ้นกับตัวเรียนรู้แบบเทคนิคประเภทตัวหุ้มก็ได้ จากนั้นจึงจะนำค่าความเหมะสมของเซตย่อยมาตรวจสอบกับเกณฑ์การหยุด (Stopping Criteria) เพื่อยับยั้งกระบวนการ ทั้งนี้เกณฑ์การหยุดสามารถนิยามได้หลายแบบเช่น กระบวนการค้นหาเสร็จสิ้น กระบวนการครบจำนวนรอบ ค่าความเหมาะสมไม่เปลี่ยนแปลงเป็นระยะเวลานานเมื่อเทียบกับจำนวนคุณลักษณะ เป็นต้น จากนั้นจึงนำเซตย่อยของคุณลักษณะที่ได้มาตรวจสอบอีกครั้งว่าเมื่อลดมิติแล้วสามารถช่วยในการเรียนรู้ของเครื่องได้จริง



รูปที่ 5 ขั้นตอนร่วมของเทคนิคประเภทคัดเลือกคุณลักษณะ

2.2 เทคนิคประเภทสกัดคุณลักษณะ (Feature Extraction)

เทคนิคประเภทสกัดคุณลักษณะจะแตกต่างจากเทคนิคประเภทคัดเลือกคุณลักษณะโดยจะทำการหาผลรวมเชิงเส้น (Linear Combination) ที่เหมาะสมจากคุณลักษณะเดิมมาใช้เป็นคุณลักษณะใหม่ที่มีจำนวนมิติน้อยลงมากๆ ทั้งนี้แต่ละเทคนิคก็จะแตกต่างกันไปตามเงื่อนไขบังคับ (Constraint) ที่จะต้องการรักษาข้อมูลใดไว้ตามข้อสันนิษฐานว่าเมื่อรักษาคุณสมบัติทางเรขาคณิตในข้อมูลเดิมไว้เมื่อลดมิติแล้วจะพบกับข้อมูลท่อร่วมที่ฝังตัวอยู่

เทคนิคประเภทสกัดคุณลักษณะนั้นยังสามารถแบ่งย่อยได้อีก 2 ประเภท[13]ตามลักษณะของการเรียนรู้คือ แบบไม่สอน (Unsupervised) และแบบสอน (Supervised)

ทั้งนี้เราสามารถเขียนสรุปเทคนิคประเภทสกัดคุณลักษณะออกมาเป็นอนุกรมวิธานของได้ดังรูปที่ 6

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รูปที่ 6 อนุกรมวิธานของเทคนิคประเภทสกัดคุณลักษณะ

2.2.1 เทคนิคประเภทสกัดคุณลักษณะแบบไม่สอน (Unsupervised Feature Extraction Techniques)

เทคนิคประเภทสกัดคุณลักษณะแบบไม่สอนจะเรียนรู้เหมือนการเรียนรู้ของสมองคนเราปกติที่ไม่รู้เฉลยมาก่อนจึงดูเป็นธรรมชาติมากกว่า โดยแต่ละเทคนิคจะกำหนดข้อสันนิษฐานมาใช้เป็นเงื่อนไขบังคับแตกต่างกัน โดยเทคนิคที่เป็นแบบแผนที่สุดคือ การวิเคราะห์องค์ประกอบสำคัญหรือพีซีเอ (PCA: Principal Components Analysis)[14] ซึ่งสันนิษฐานว่าเราสามารถเรียนรู้ข้อมูลท่อร่วมที่มีมิติต่ำกว่าได้ด้วยการตามทิศทางที่มีความแปรปรวนสูงที่สุดของข้อมูลตั้งต้น

ต่อมาได้มีผู้แบ่งเทคนิคประเภทนี้ที่ใช้กันมากไว้เป็นหมวดหมู่ได้หลายแบบ ในที่นี้ขอยกตัวอย่างแบบวิเคราะห์ตามความเป็นคอนเวกซ์ (Convex) ของเทคนิค

2.2.1.1 เทคนิคแบบคอนเวกซ์

การแบ่งตามความเป็นคอนเวกซ์นั้นจะแบ่งได้เป็นเทคนิคแบบคอนเวกซ์และเทคนิคแบบไม่คอนเวกซ์ตามฟังก์ชันวัตถุประสงค์ (Objective Function) ว่ามีค่าเหมาะสมเฉพาะที่ (Local Optima) หรือไม่ เทคนิคแบบคอนเวกซ์นั้นจะมีฟังก์ชันจุดประสงค์ที่มีค่าเหมาะสมโดยรวม (Global Optima) กับค่าเหมาะสมเฉพาะที่เป็นค่าเดียวกัน เทคนิคในปัจจุบันส่วนมากจะอยู่ในประเภทนี้ เพราะสามารถนำไปใช้ได้ง่ายและรับประกันผลลัพธ์ตามข้อสันนิษฐาน เทคนิคประเภทนี้ส่วนมากจะใช้เทคนิคการหาค่าเหมาะที่สุดด้วยวิธีสเปกตรัม (Spectral Method) เนื่องจากฟังก์ชันจุดประสงค์สามารถเขียนให้อยู่ในรูปของปัญหาสัมประสิทธิ์ในปริภูมิลักษณะของเรย์ลี (Rayleigh Coefficient in $\mathcal{H}$) หรือผลหารเรย์ลี (Rayleigh Quotient) นั่นเอง ซึ่งปัญหานี้สามารถแก้ได้โดยการแก้ปัญหาค่าลักษณะเฉพาะแบบทั่วไป (Generalized Eigenvalue Problem) ซึ่งสามารถแบ่งได้อีกว่าลักษณะของเมตริกซ์ที่จะแยกนั้นมีความห่างมาก (Sparse) หรือไม่

เทคนิคประเภทคอนเวกซ์ที่เมตริกซ์เป็นแบบเต็ม (Full Matrix) เป็นเมตริกซ์ที่แน่นไปด้วยข้อมูลซึ่งอาจเป็นข้อมูลความแปรปรวนร่วม (Covariance) ระหว่างมิติหรือเป็นค่าความคล้าย (Similarity) ระหว่างจุดข้อมูล ตัวอย่างของเทคนิคประเภทนี้เช่น การวิเคราะห์องค์ประกอบสำคัญหรือพีซีเอ (PCA: Principal Components Analysis) การสเกลหลายมิติหรือเอ็มดีเอส (MDS: Multi Dimensional Scaling)[15] ไอโซแมป (Isomap)[16] การวิเคราะห์องค์ประกอบสำคัญแบบเคอร์เนลหรือเคอร์เนลพีซีเอ (Kernel PCA)[17] เป็นต้น

เทคนิคประเภทคอนเว็กซ์ที่เมตริกซ์เป็นแบบห่างมากมักจะเน้นการรักษาสมบัติเฉพาะที่ (Local Property) ของข้อมูล โดยทั่วไปมักจะใช้กราฟเพื่อนบ้านใกล้สุด k ตัว (K-Nearest Neighbor) เพื่อที่จะคงโครงสร้างทางเรขาคณิต ในขณะเดียวกันจะใช้การคำนวณไม่มากในการแก้ปัญหาเนื่องจากเมตริกซ์มีลักษณะพิเศษคือมีความห่างมากนั่นเอง ตัวอย่างของเทคนิคประเภทนี้เช่น ตัวฝังตัวเชิงเส้นเฉพาะที่หรือแอลแอลอี (LLE: Locally Linear Embedding)[18] กราฟไอเกนแบบลาปลาเชียนหรือแอลอี (LE: Laplacian Eigenmaps)[19] เป็นต้น

2.2.1.1 เทคนิคแบบไม่คอนเวกซ์

ส่วนเทคนิคประเภทไม่คอนเวกซ์นั้นจะไม่รับประกันว่าผลลัพธ์จะได้ตามสมมติฐานเต็มที่ เนื่องจากปัญหาไม่คอนเวกซ์จึงหาค่าที่เหมาะที่สุดได้ยาก จึงไม่ได้รับความสนใจมากนัก แต่ยังคงมีความน่าสนใจเชิงทฤษฎีอยู่ และเมื่อนำไปใช้งานก็ได้รับผลลัพธ์ที่ดีในระดับหนึ่ง ตัวอย่างของ

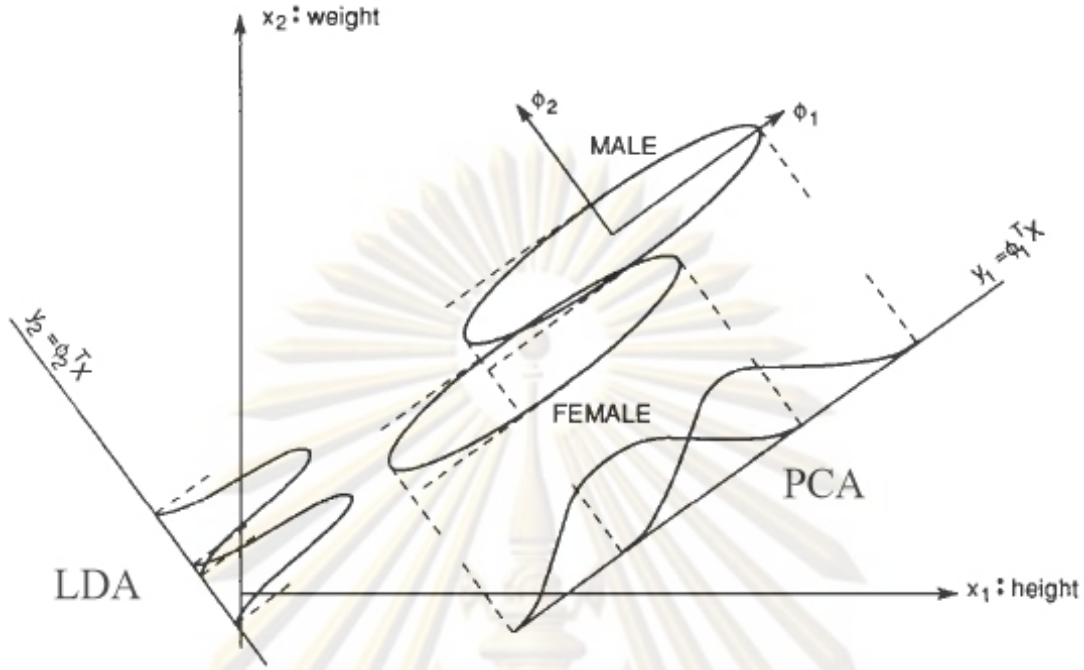
เทคนิคประเภทนี้เช่น ตัวเข้ารหัสอัตโนมัติหลายชั้น (Multilayer Autoencoders)[20] การประสานงานเชิงเส้นเฉพาะที่หรือแอลแอลซี (LLC: Locally Linear Coordination)[21]

2.2.2 เทคนิคประเภทสกัดคุณลักษณะแบบสอน (Supervised Feature Extraction Techniques)

เทคนิคประเภทสกัดคุณลักษณะแบบสอนนั้นนอกจากการเรียนรู้ของเครื่องแล้วมักจะนำไปใช้ในงานจำพวกรู้จำรูปแบบ (Pattern Recognition) การประมวลผลภาพ (Image Processing) อีกด้วย โดยเทคนิคประเภทนี้โดยทั่วไปจะทำงานได้ดีกว่าเทคนิคแบบไม่สอน เนื่องจากนำข้อมูลประเภทที่ทราบอยู่ก่อนมาใช้ในขั้นตอนการฝึก ทำให้ตัวแบบ (Model) ที่ได้มีความแม่นมากขึ้นและสามารถเรียนรู้ข้อมูลท่อร่วมที่เป็นไปตามประเภทได้ดีกว่า แต่อย่างไรก็ตามหากไม่มีข้อมูลประเภทแล้วเทคนิคแบบไม่สอนก็ยังคงจำเป็นอยู่เช่นเดิม

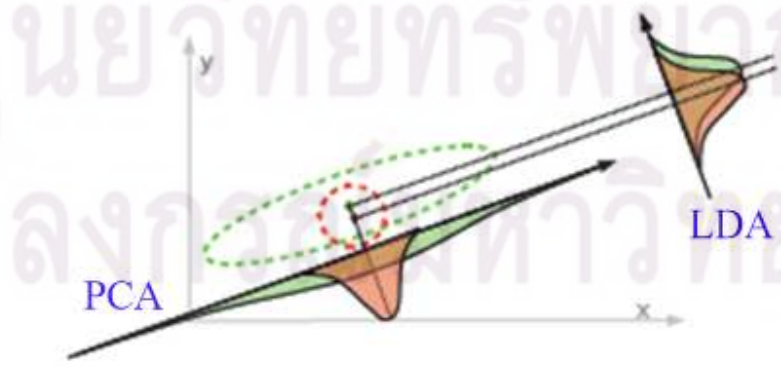
2.2.2.1 เปรียบเทียบเทคนิคประเภทสกัดคุณลักษณะแบบสอนและเทคนิคประเภทสกัดคุณลักษณะแบบไม่สอน

ยกตัวอย่างกรณีเปรียบเทียบโดยเริ่มจากเทคนิคการวิเคราะห์องค์ประกอบสำคัญหรือพีซีเอที่มักจะหาการฉายในมุมมองที่รักษาค่าความแปรปรวนมากที่สุด แต่หากการกระจายตัวของข้อมูลไม่ได้อยู่บนทิศทางของความแปรปรวน พีซีเอก็จะไม่เหมาะสมกับชุดข้อมูลนั้นทันที โดยเทคนิคประเภทมีผู้สอนที่เป็นที่นิยมที่เทียบเคียงกับการวิเคราะห์องค์ประกอบสำคัญคือ การวิเคราะห์ดิส-คริมิแนนต์เชิงเส้นหรือแอลดีเอ (LDA: Linear Discriminant Analysis)[22] ซึ่งมีอีกชื่อหนึ่งว่าดิส-คริมิแนนต์เชิงเส้นของฟิชเชอร์หรือเอฟแอลดี (FLD: Fisher's Linear Discriminant) บางครั้งอาจเรียกว่า การวิเคราะห์คุณลักษณะจำแนกหรือดีเอฟเอ (DFA: Discriminative Feature Analysis) เพื่อไม่ให้ซ้ำซ้อนกับเทคนิคการจำแนกที่ชื่อเหมือนกัน แอลดีเอจะหาทิศทางการฉายที่ทำให้เมื่อฉายข้อมูลลงไปแล้วอัตราส่วนของความแปรปรวนระหว่างประเภทต่อความแปรปรวนภายในประเภทสูงสุด หรือสามารถมองได้ว่าในมิติหลังการฉายจะพยายามให้ค่ามัชฌิมของแต่ละประเภทอยู่ห่างกันมากที่สุดก็ได้เช่นกัน จากรูปที่ 7 ซึ่งข้อมูลอยู่ในสเปซ 2 มิติมีลักษณะเป็นกลุ่ม (Cluster) โดยแต่ละกลุ่มจะมีข้อมูลที่มีประเภทเดียวกันอยู่ด้วยกัน เมื่อลดมิติให้เหลือ 1 มิติจะเห็นได้ว่าแอลดีเอสามารถคงข้อมูลที่แสดงถึงอำนาจจำแนก (Discriminative Power) ได้ดีกว่าพีซีเอ เนื่องจากนำข้อมูลประเภทเข้าร่วมในการประมวลผลด้วยนั่นเอง ในขณะที่พีซีเอตามทิศทางความแปรปรวนซึ่งทำให้อำนาจจำแนกลดลงอย่างเห็นได้ชัด



รูปที่ 7 ตัวอย่างความแตกต่างของเทคนิคสำหรับการลดมิติประเภทสกัดลักษณะแบบสอนและไม่สอน[23]

แต่อย่างไรก็ตามยังคงมีบางกรณีที่พีซีเออาจให้อำนาจจำแนกมากกว่าแอลดีเอ จากรูปที่ 8 ข้อมูล 2 ประเภทกระจายตัวแบบฐานเดี่ยว (Unimodal) ทำให้ประเภทหนึ่งอยู่ทับบนอีกประเภทหนึ่งที่มีการกระจายตัวสูงกว่า แอลดีเอจะได้การฉายที่ค่ามัชฌิมอยู่ห่างกันมากกว่าของพีซีเอตามสมมติฐานว่าความห่างของค่ามัชฌิมระหว่างประเภทที่มากและการกระจายตัวภายในประเภทที่น้อยจะทำให้มีอำนาจจำแนกสูง แต่กลับไม่เป็นเช่นนั้นเมื่ออำนาจจำแนกอยู่ในทิศทางของความแปรปรวนซึ่งพีซีเอจะทำงานได้ดีกว่า อย่างไรก็ดีในโลกความเป็นจริงโดยส่วนใหญ่แล้วแอลดีเอรวมถึงเทคนิคแบบสอนมักจะทำงานดีกว่าพีซีเอและเทคนิคแบบไม่สอนเสมอในสถานการณ์ที่สามารถหาข้อมูลประเภทได้



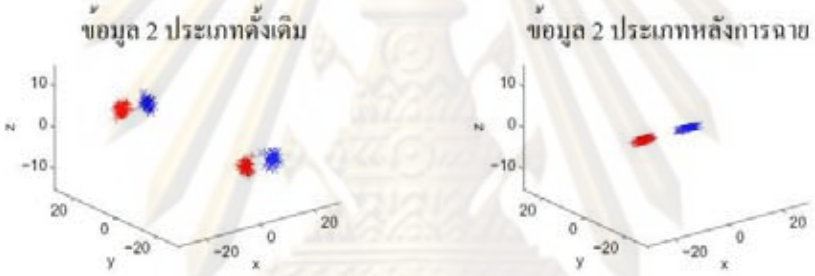
รูปที่ 8 กรณีที่พีซีเอเหมาะสมในการช่วยจำแนกมากกว่าแอลดีเอ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

2.2.2.2 ตัวอย่างของเทคนิคประเภทสกัดคุณลักษณะแบบสอน

เทคนิคประเภทสกัดคุณลักษณะแบบสอนนั้นมีผู้สร้างเทคนิคออกมาได้ผลในการทดลองจริงที่ดีมากมาย ซึ่งเทคนิคประเภทนี้มีผู้จำแนกออกเป็น 2 ประเภทใหญ่ๆ[24, 25]ได้แก่ ประเภทเรียนรู้โดยรวม (Global Learning) และประเภทเรียนรู้เฉพาะที่ (Local Learning)

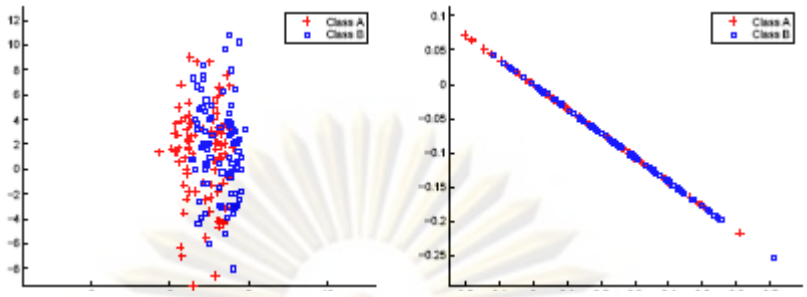
เทคนิคประเภทเรียนรู้สมบัติโดยรวมนั้นจะเรียนรู้ระยะทางของมหาลาโนบิส (Mahalanobis Distance) ที่ทำให้จุดข้อมูลที่มีประเภทเดียวกันอยู่ใกล้กันและจุดข้อมูลที่มีประเภทต่างกันอยู่ห่างกันมากที่สุด ตัวอย่างเช่นดังรูปที่ 9 ซึ่งมีข้อมูล 2 ประเภทที่มีลักษณะเป็นกลุ่ม (Cluster) เมื่อฉายเมตริกซ์ระยะทางทำให้ลักษณะการรวมกลุ่มชัดเจนขึ้น



รูปที่ 9 ผลลัพธ์จากการสเกลข้อมูลด้วยเมตริกซ์ระยะทาง

ตัวอย่างของเทคนิคประเภทเรียนรู้สมบัติโดยรวมเช่น การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นหรือแอลดีเอ (LDA: Linear Discriminant Analysis)[22] การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบเคอร์เนลหรือเคอร์เนลแอลดีเอ (KLDA: Kernel Linear Discriminant Analysis)[26] เป็นต้น

อย่างไรก็ตามยังคงมีกรณีซึ่งการเรียนรู้เมตริกซ์ระยะทางทำได้ยาก ตัวอย่างของกรณีเหล่านั้นเช่น กรณีที่ข้อมูลมีการกระจายตัวแบบหลายฐาน (Multimodal) ดังรูปที่ 10 มีข้อมูล 2 ประเภท โดยแต่ละประเภทมี 2 ฐาน การเรียนรู้สมบัติโดยรวมไม่สามารถทำให้ข้อมูลที่มีประเภทเดียวกันอยู่ใกล้กันและข้อมูลคนละประเภทอยู่ห่างกันได้ จึงมีผู้ปรับปรุงเทคนิคเดิมให้รักษาสมบัติเฉพาะที่ด้วยการเพิ่มเงื่อนไขบังคับเฉพาะที่ (Local Constraint) เพื่อรองรับกรณีเหล่านี้



รูปที่ 10 กรณีข้อมูลมีการกระจายตัวแบบหลายฐานซึ่งการเรียนรู้เมตริกซ์ระยะทางของมหาลาโนบิสโดยทั่วไปทำได้ยาก

เทคนิคประเภทเรียนรู้สมบัติเฉพาะที่จะใช้การเรียนรู้ระยะทางเฉพาะที่ (Local Distance Learning) เพิ่มเข้าไปเป็นเงื่อนไขบังคับในการเรียนรู้ระยะทางของมหาลาโนบิสของเทคนิคประเภทเรียนรู้สมบัติโดยรวม โดยจะเรียนรู้ระยะทางจากแต่ละจุดข้อมูลแทนระยะทางจากการสเกลสเปซโดยปกติ ทั้งนี้มักใช้เมตริกซ์ของเพื่อนบ้านใกล้สุด k ตัว (K-Nearest Neighbors) หรือเมตริกซ์สัมพรรค (Affine Matrix) เป็นเงื่อนไขบังคับเฉพาะที่

เทคนิคประเภทสกัดคุณลักษณะแบบสอนที่พัฒนาขึ้นเร็วๆนี้มักเป็นประเภทเรียนรู้เฉพาะที่ ตัวอย่างเช่น การวิเคราะห์ขอบแบบฟิชเชอร์หรือเอ็มเอฟเอ (MFA: Marginal Fisher Analysis)[27] ตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์หรือดีเอ็นอี (DNE: Discriminative Neighborhood Embedding)[28] การหาองค์ประกอบเพื่อนบ้านหรือเอ็นซีเอ (NCA: Neighborhood Component Analysis)[29] การวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่หรือแอลเอสดีเอ (LSDA: Locality Sensitive Discriminant Analysis)[30] เป็นต้น

2.2.3 การแบ่งเทคนิคประเภทสกัดคุณลักษณะตามความเป็นเชิงเส้น

นอกจากนี้เทคนิคประเภทสกัดคุณลักษณะยังสามารถแบ่งได้ตามความเป็นเชิงเส้นเป็นแบบเชิงเส้นและไม่เชิงเส้นได้อีกด้วย โดยเทคนิคประเภทเชิงเส้นนั้นจะมีสมมติฐานว่าข้อมูลท่อร่วมนั้นเป็นเชิงเส้นหรือในกรณีที่ข้อมูลท่อร่วมไม่เป็นเชิงเส้นจะประมาณให้ใกล้เคียงมากที่สุดเท่าที่ความเป็นเชิงเส้นจะทำได้ เทคนิคประเภทนี้ได้รับความนิยมมาประยุกต์ใช้งานจริง เนื่องจากให้ประสิทธิภาพที่ดีและปริมาณการคำนวณไม่มากเกินไป ตัวอย่างของเทคนิคประเภทเชิงเส้นได้แก่ การวิเคราะห์องค์ประกอบสำคัญหรือพีซีเอ (PCA: Principal Components Analysis) ซึ่งเป็นเทคนิคแบบไม่สอน ส่วนตัวอย่างเทคนิคประเภทสกัดคุณลักษณะแบบสอนนั้นโดยมากมักเป็นแบบเชิงเส้นอยู่แล้วตัวอย่างเช่น การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นหรือแอลดีเอ (LDA: Linear

Discriminant Analysis) ตัวฝังตัวที่จำแนกเพื่อนบ้านหรือดีเอ็นอี (DNE: Discriminative Neighborhood Embedding) เป็นต้น

นอกจากนี้ยังมีเทคนิคแบบเชิงเส้นที่เป็นอนุพันธ์ (Derive) มาจากเทคนิคแบบไม่เชิงเส้น เช่น การฉายรักษาสมบัติเฉพาะที่หรือแอลพีพี (LPP: Locally Preserving Projection)[31] ซึ่งเป็นอนุพันธ์ของกราฟไอเกนแบบลาปลาเชียนหรือแอลอี (LE: Laplacian Eigenmaps)[19] ตัวฝังตัวรักษาเพื่อนบ้านหรือเอ็นพีอี (NPE: Neighborhood Preserving Embedding)[32] ซึ่งเป็นอนุพันธ์ของตัวฝังตัวเชิงเส้นเฉพาะที่หรือแอลแอลอี (LLE: Locally Linear Embedding)[18] เป็นต้น เทคนิคประเภทนี้ประมาณแนวคิดแบบไม่ใช่เชิงเส้นด้วยความเป็นเชิงเส้นซึ่งให้ผลลัพธ์ที่ดีเช่นกัน

ส่วนเทคนิคแบบไม่เชิงเส้นจะเรียนรู้ข้อมูลท่อร่วมโดยไม่ได้ตั้งสมมติฐานว่าข้อมูลท่อร่วมจะมีลักษณะอย่างไร แต่จะเน้นเรียนรู้คุณสมบัติทางเรขาคณิตแทน บางเทคนิคจะใช้อุบายเคอร์เนล (Kernel Trick) เพื่อทำให้เทคนิคแบบเชิงเส้นสามารถทำงานบนข้อมูลที่ไม่เป็นเชิงเส้นได้ ซึ่งเทคนิคแบบไม่เชิงเส้นนี้ถูกสร้างขึ้นมาด้วยสมมติฐานว่าข้อมูลในโลกจริงส่วนมากเป็นแบบไม่เป็นเชิงเส้น ดังนั้นเทคนิคประเภทนี้ควรจะให้ผลลัพธ์ที่ดีกว่าเทคนิคแบบเชิงเส้น อย่างไรก็ตามเทคนิคประเภทนี้ยังมีความซับซ้อนทางการคำนวณที่สูงเกินกว่าจะใช้งานจริงได้ดีเมื่อข้อมูลมีจำนวนมากๆ และผลลัพธ์เบื้องต้นในปัจจุบันพบว่าเทคนิคประเภทนี้ทำงานได้ดีในชุดข้อมูลสังเคราะห์มากกว่าข้อมูลโลกจริง ตัวอย่างของเทคนิคประเภทนี้เช่น การวิเคราะห์องค์ประกอบสำคัญแบบเคอร์เนลหรือเคอร์เนลพีซีเอ (KPCA: Kernel Principal Components Analysis) การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบเคอร์เนลหรือเคอร์เนลแอลดีเอ (KLDA: Kernel Linear Discriminant Analysis) ตัวฝังตัวเชิงเส้นเฉพาะที่หรือแอลแอลอี (LLE: Locally Linear Embedding) กราฟไอเกนแบบลาปลาเชียนหรือแอลอี (LE: Laplacian Eigenmaps) เป็นต้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

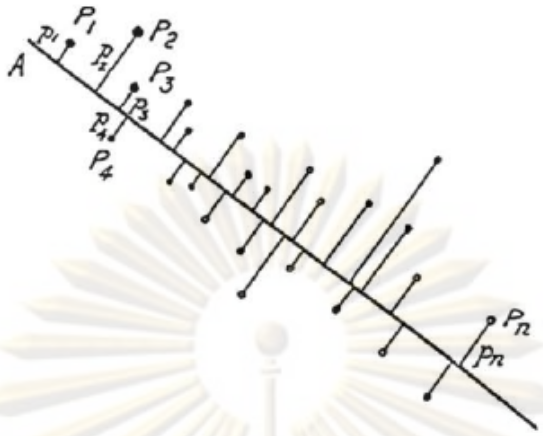
บทที่ 3

งานวิจัยที่เกี่ยวข้อง

3.1 การวิเคราะห์องค์ประกอบสำคัญหรือพีซีเอ (PCA: Principal Component Analysis)

เดิมทีนั้นมีผู้นำเสนอเทคนิคสำหรับการลดมิติในอดีตมานานแล้วคือ การวิเคราะห์-องค์ประกอบสำคัญหรือพีซีเอ (PCA: Principal Component Analysis) ซึ่งเสนอโดย Pearson ในปี ค.ศ.1901 ต่อมาเทคนิคนี้ได้รับการพัฒนาเพิ่มเติมโดย Hotelling ในปีค.ศ.1933 โดย Pearson ได้เสนอต่อวงการว่าปัญหาการประมาณข้อมูล (Data Approximation) นั้นสำคัญ สวยงาม และต่างจากปัญหาการถดถอย จากนั้นจึงนำเสนอการประมาณด้วยเส้นตรงและระนาบ[33]

Pearson ได้อธิบายเหตุผลไว้ 2 ข้อ [14] คือ ในการทำวิจัยด้านฟิสิกส์ สถิติ และชีววิทยา การประมาณข้อมูลด้วยเส้นหรือระนาบที่เหมาะสมที่สุดเป็นที่น่าพึงพอใจกว่าในการวิเคราะห์โดยทั่วไปมักใช้วิธีกำลังสองน้อยสุด (Least Square) คือหาค่าสัมประสิทธิ์ที่ดีที่สุดสำหรับผลรวมเชิงเส้นของตัวแปร ซึ่งค่าของผลรวมเชิงเส้นที่ประมาณจากตัวแปรอิสระชุดหนึ่ง กับการประมาณค่าของตัวแปรอิสระเมื่อรู้ผลรวมเชิงเส้นค่าหนึ่งนั้นไม่เท่ากัน และเหตุผลอีกข้อหนึ่งคือตัวแปรตามจะขึ้นกับค่าเบี่ยงเบนและค่าความผิดพลาดพอๆกับตัวแปรอิสระ เราไม่ได้ทราบค่าตัวแปรอิสระที่แม่นยำแล้วจึงหาฟังก์ชันไปหาตัวแปรตาม แต่เราได้ค่าเป็นชุดๆจากการทดลองหรือการสังเกต เพื่อหาฟังก์ชันความสัมพันธ์ระหว่างกัน ซึ่งตัวแปรที่สังเกตได้จะมีความผิดพลาดร่วมอยู่เสมอ ดังนั้นเพื่อที่จะทำให้เส้นตรงหรือระนาบเหมาะสมที่สุด เราอาจมีหลายวิธีที่จะทำซึ่งเราสามารถหาได้จากวิธีกำลังสองน้อยสุดที่พยายามให้ผลรวมของระยะตั้งฉากจากแต่ละจุดมาหาเส้นตรงหรือระนาบน้อยสุด ดังรูปที่ 11 ระบบถูกแทนด้วยจุดข้อมูล p_1 ถึง p_n ซึ่งแต่ละจุดสามารถแทนระบบทั้งหมดได้ด้วยเส้นตรง A ซึ่งเป็นผลรวมเชิงเส้นของแต่ละจุดข้อมูล ซึ่งนี่เป็นวิธีช่วยลดสัญญาณรบกวนได้อย่างดีอีกด้วย นี่คือข้อเปรียบเทียบที่ว่าการประมาณข้อมูลดีกว่าการถดถอยซึ่งเป็นการแปลงข้อมูลทางคู่อันดับ วิธีนี้คือการวิเคราะห์องค์ประกอบสำคัญนั่นเองหรืออีกชื่อหนึ่งคือการแยกแบบตั้งฉากถูกต้อง (Proper Perpendicular Decomposition) หรือชื่อที่เป็นที่รู้จักกันมากในภายหลังคือการแปลงของคาฮูเน็น-เลิฟ (Karhunen–Loève Transform)



รูปที่ 11 การประมาณข้อมูลด้วยเส้นตรงของ Pearson [14]

การวิเคราะห์องค์ประกอบสำคัญจะฉายข้อมูลตั้งต้น X โดยหาเมตริกซ์การฉาย P ซึ่งประกอบด้วยเวกเตอร์ $[\varphi_1, \varphi_2,..., \varphi_n]$ ทำให้ทิศทางที่ทำให้ข้อมูลที่ถูกฉาย Z มีความแปรปรวนมากที่สุด สามารถเขียนฟังก์ชันจุดประสงค์ (Objective Function) ได้ 2 แบบ[34, 35]โดยแบบแรกคือ

$$\max_{P^TP=1} P^TCP \qquad (3.1)$$

โดย C แทนเมตริกซ์ความแปรปรวนร่วม โดยมีเงื่อนไขบังคับ (Constraint) ว่าเมตริกซ์การฉาย P นั้นถูกทำให้เป็นบรรทัดฐาน (Normalized) ซึ่งมองได้ว่าข้อมูลแต่ละจุด x_i ถูกฉายไปเป็นข้อมูลที่ถูกฉาย $z_i = P^T x_i$ โดยมีค่ามัชฌิม $z_m = P^T x_m$ ความแปรปรวนของข้อมูลที่ถูกฉายคือ

$$\frac{1}{N}\sum_{i=1}^{N}(P^T x_i - P^T x_m)^2 = P^TCP \qquad (3.2)$$

จากนั้นเราจะใช้ตัวคูณลากรานจ์ (Lagrange Multiplier) เพื่อทำให้เป็นการหาค่าเหมาะที่สุดแบบไม่มีเงื่อนไขบังคับ (Unconstraint Optimization)

$$P^TCP + \lambda(1 - P^TP) \qquad (3.3)$$

เมื่อหาอนุพันธ์บางส่วนเทียบกับ P แล้วจับเท่ากับ 0 เพื่อหาจุดนิ่ง (Stationary Point) จะได้

$$CP - \lambda P = 0 \qquad (3.4)$$

เมื่อย้ายข้างจะได้ในรูปของสมการค่าลักษณะเฉพาะ (Eigenvalue Problem)

$$CP = \lambda P \qquad (3.5)$$

ดังนั้นเมตริกซ์ P สามารถหาได้จากการแก้สมการค่าลักษณะเฉพาะจากเมตริกซ์ความแปรปรวนร่วม C นั่นเอง โดยเวกเตอร์การฉาย φ จะรักษาความแปรปรวนของข้อมูลหลังการฉายตามขนาดของค่าลักษณะเฉพาะ λ ซึ่งผลรวมของค่าลักษณะเฉพาะจะเท่ากับค่าความแปรปรวนของข้อมูลดั้งเดิม ดังนั้นผลรวมของค่าลักษณะเฉพาะ λ ที่คู่กับเวกเตอร์การฉาย φ จำนวน d ตัว สามารถใช้เป็นเกณฑ์ในการคงปริมาณข้อมูลเมื่อลดมิติแล้วได้

การเขียนฟังก์ชันจุดประสงค์อีกแบบจะมองว่าเราจะหาการฉายที่ทำให้ค่าความผิดพลาด (Error) น้อยที่สุด สมมติว่าข้อมูลแต่ละจุดสามารถแทนได้ด้วยผลรวมเชิงเส้นของเวกเตอร์ฐานหลัก ในที่นี้ขอนำสัญกรณ์จากฟังก์ชันจุดประสงค์แบบแรกมาใช้จะได้

$$x_n = \sum_{i=1}^{D} \alpha_{ni} \varphi_i \qquad (3.6)$$

เมื่อสัมประสิทธิ์ α จะมีค่าเท่ากับ $x_n^T \varphi$ ซึ่งคือผลคูณจุดของ x_n และ φ นั่นเอง โดยสามารถมองเป็นขนาดของการฉายของข้อมูลลงมาแต่ละแกนได้ ซึ่งหากเราต้องการประมาณ x_n ด้วย x_n' จะได้จากการเลือกเวกเตอร์ฐานหลักบางตัวที่จะใช้

$$x_n' = \sum_{i=1}^{d} p_{ni} \varphi_i + \sum_{i=d+1}^{D} q_{ni} \varphi_i \qquad (3.7)$$

ดังนั้นฟังก์ชันจุดประสงค์จะเป็นการทำให้ค่าการบิดเบี้ยว (Distortion) เมื่อทำการลดมิติต่ำสุดคือ

$$\min_{P^T P=1} \frac{1}{N} \sum_{n=1}^{N} \|x_n - x_n'\|^2 \qquad (3.8)$$

เมื่อหาอนุพันธ์บางส่วนในสมการ (3.7) เทียบกับ p_{ni} จะได้ $p_{ni} = x_n^T \varphi$ เมื่อ n มีค่าจาก 1 ถึง d และเมื่อหาอนุพันธ์บางส่วนเทียบกับ q_{ni} จะได้ $q_{ni} = x_n^T \varphi$ เมื่อ n มีค่าจาก d+1 ถึง D ดังนั้นจะได้

$$x_n - x_n' = \sum_{i=d+1}^{D} \{(x_n - x_n')^T \varphi_i\} \varphi_i \qquad (3.9)$$

เมื่อนำไปแทนค่าในฟังก์ชันจุดประสงค์จะได้

$$\min_{P^T P=1} \frac{1}{N} \sum_{n=1}^{N} \left(\sum_{i=d+1}^{D} \{(x_n - x_n')^T \varphi_i\} \varphi_i \right)^2 \qquad (3.10)$$

หรือเมื่อกระจายแล้วจะกลายเป็นเมตริกซ์ความแปรปรวนร่วม C ทำให้ฟังก์ชันจุดประสงค์กลายเป็น

$$\min_{P^TP=1} \sum_{i=d+1}^{D} \varphi_i{}^T C \varphi_i \qquad (3.11)$$

ซึ่งมีความหมายว่าทำให้ค่าความผิดพลาดหรือความบิดเบี้ยวที่ตัดออกเมื่อลดมิติน้อยสุดนั่นเอง ทั้งนี้ยังคงมีเงื่อนไขบังคับ $P^TP = 1$ เพื่อป้องกันผลเฉลยชัด (Trivial Solution) คือ $\varphi_i = 0$ อันเป็นผลเฉลยที่ไม่ต้องการ จะเห็นว่าสามารถแปลงเป็นปัญหาค่าลักษณะเฉพาะเพื่อหาค่าเหมาะที่สุดได้ในทำนองเดียวกันกับการเขียนฟังก์ชันจุดประสงค์แบบแรก โดยเลือกเวกเตอร์ฐานหลักตามค่าลักษณะเฉพาะที่น้อยสุดออก

ในการเขียนฟังก์ชันจุดประสงค์ทั้ง 2 แบบ ผลเฉลยจะเกิดจากการแก้ปัญหาค่าลักษณะเฉพาะจากเมตริกซ์ความแปรปรวนร่วมแล้วเลือกเวกเตอร์ลักษณะเฉพาะ (Eigenvector) มาเป็นเวกเตอร์การฉายหรือเวกเตอร์ฐานหลักตามขนาดของค่าลักษณะเฉพาะ (Eigenvalue) นั่นเอง

สำหรับงานด้านการประมวลผลภาพ (Image Processing) หรือการรู้จำใบหน้า (Face Recognition) ข้อมูลมักจะมีจำนวนมิติสูงกว่าจำนวนข้อมูลมากๆ (D >> N) ซึ่งค่าลักษณะเฉพาะอย่างน้อย D-N+1 มีค่าเป็นศูนย์ ทำให้การคำนวณตามปกติไม่มีประสิทธิภาพเท่าที่ควร Turk และคณะได้นำเสนอการแก้ปัญหาแบบใหม่เรียกว่า ใบหน้าลักษณะเฉพาะหรือใบหน้าไอเกน (Eigenface) [36] ซึ่งจะใช้อุบายกับการคำนวณเมตริกซ์ความแปรปรวนร่วมเพื่อให้สามารถคำนวณได้อย่างรวดเร็วและแม่นยำ

เทคนิคใบหน้าลักษณะเฉพาะทำการคำนวณเมตริกซ์ความแปรปรวนด้วยสูตร

$$C' = \frac{1}{N} X X^T \qquad (3.12)$$

จากนั้นก็จะแก้ปัญหาค่าลักษณะเฉพาะของเมตริกซ์ความแปรปรวนร่วม C' จนได้เวกเตอร์ลักษณะเฉพาะ γ_i ออกมา ซึ่งเมื่อจะนำไปใช้เป็นเวกเตอร์การฉายต้องแปลงก่อนด้วยสมการ

$$\varphi_i = \frac{1}{(N\lambda_i)^{\frac{1}{2}}} X^T \gamma_i \qquad (3.13)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ซึ่งเทคนิคนี้มาจากการนำเมตริกซ์ข้อมูล X คูณทั้ง 2 ข้างของสมการ (3.5) โดยแปลงรูปเมตริกซ์ความแปรปรวนร่วมเนื่องจากเมตริกซ์ความแปรปรวนร่วม C สามารถเขียนในรูปการคูณของเมตริกซ์ข้อมูล X^TX ได้

$$X(X^TX)P = X\lambda P \quad (3.14)$$

เนื่องจาก λ เป็นสเกลาร์จึงสลับที่กับเมตริกซ์ข้อมูล X ได้ จากนั้นสลับวงเล็บจะได้

$$(XX^T)(XP) = \lambda(XP) \quad (3.15)$$

เมื่อให้เมตริกซ์การฉาย Q ประกอบด้วยเวกเตอร์ $[\gamma_1, \gamma_2, ..., \gamma_n]$ ทำให้สมการเปลี่ยนเป็นสมการของปัญหาค่าลักษณะเฉพาะอีกครั้ง

$$C'Q = \lambda Q \quad (3.16)$$

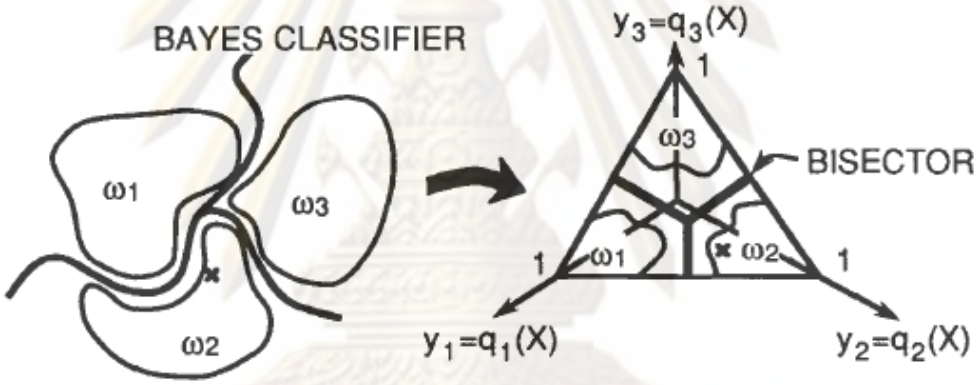
ซึ่งสามารถแก้สมการเพื่อหาค่าเหมาะที่สุดด้วยวิธีเดียวกันกับสมการ (3.5) จากนั้นจึงนำเวกเตอร์ลักษณะเฉพาะมาแปลงให้เป็นเวกเตอร์การฉายด้วยสมการ (3.13) ซึ่งสัมประสิทธิ์ด้านหน้าคือการทำให้เวกเตอร์เป็นบรรทัดฐาน (Normalized) ด้วยการหารด้วยรากที่สองของผลคูณระหว่างจำนวนข้อมูลและค่าลักษณะเฉพาะที่คู่กับเวกเตอร์นั้นๆ

เทคนิคนี้ได้รับการพิสูจน์ว่าสามารถใช้งานได้ผลลัพธ์ดีมากจนได้รับยกย่องเป็นเทคนิคที่ทันสมัย (State of the art Technique) ในการรู้จำใบหน้าและเป็นเทคนิคที่ทำให้วงการเกิดความสนใจที่จะนำเทคนิคสำหรับการลดมิติมาใช้ในการรู้จำใบหน้ารวมถึงการรู้จำรูปแบบต่างๆมากขึ้น ทั้งนี้เทคนิคสำหรับการลดมิติที่ใช้ในการรู้จำรูปแบบมักใช้ชื่อว่า วิธีปริภูมิย่อย (Subspace Method)

3.2 การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นหรือแอลดีเอ (LDA: Linear Discriminant Analysis)

ในการสกัดคุณลักษณะเมื่อทราบข้อมูลประเภทนั้น เราต้องการให้แต่ละประเภทสามารถแยกกันได้อย่างดีที่สุดมากกว่าใช้แทนข้อมูลเท่านั้น ทั้งนี้เกณฑ์ของการแยกประเภทจะไม่ขึ้นกับระบบคู่อันดับและต่างจากเกณฑ์โดยทั่วไปของการประมวลผลสัญญาณ (Signal Processing) ทั้งนี้ถึงแม้ว่าข้อมูลจะมีการกระจายเหมือนกัน ชุดของคุณลักษณะอาจจะมีความเหมาะสมขึ้นกับตัวจำแนกอีกด้วย ดังนั้นจึงควรที่จะตั้งสมมติฐานตามค่าความผิดพลาดของเบย์ (Bayes Error) อย่างไรก็ตามยังคงเป็นการยากที่จะหาตัวแบบทางคณิตศาสตร์ที่จะแสดงแนวคิดนี้ได้อย่างชัดเจน [23]

ในตัวจำแนกประเภทเบย์นั้นหากใช้จำแนกข้อมูลที่มีประเภทจำนวน c ประเภท ก็จะใช้ความน่าจะเป็น $q_1(X),q_2(X),\ldots,q_c(X)$ ในการจำแนกตามความควรจะเป็นสูงสุด (Maximum Likelihood) แต่เนื่องจากผลรวมของความน่าจะเป็น $q_i(X)$ นั้นเท่ากับ 1 ดังนั้นจะมีความน่าจะเป็นเพียง c-1 ตัวเท่านั้นที่เป็นอิสระเชิงเส้นต่อกัน ซึ่งหมายความว่าหากคุณลักษณะอุดมคติเป็นฟังก์ชันความน่าจะเป็นนี้ จำนวนคุณลักษณะที่จะใช้จำแนกข้อมูลได้ต้องมีขนาดอย่างน้อย c-1 เช่นกัน ดังนั้นหากเราแปลงข้อมูลด้วยฟังก์ชันความน่าจะเป็น $q_i(X)$ โดยที่ i มีค่าตั้งแต่ 1 ถึง c-1 เราสามารถแปลงข้อมูลจาก D มิติมายัง c-1 มิติโดยที่ไม่มีข้อมูลที่จำเป็นต่อการจำแนกสูญเสียไป ดังนั้น $q_1(X),q_2(X),\ldots,q_c(X)$ คือชุดข้อมูลเล็กสุดที่เหมาะที่สุดสำหรับการจำแนกนั่นเอง จากรูปที่ 12 ไม่ว่าการกระจายของข้อมูลในปริภูมิดั้งเดิมจะเป็นอย่างไรในปริภูมิอุดมคติจะกระจายในปริภูมิขนาด c-1 มิติตามสมบัติที่ว่าฟังก์ชันความน่าจะเป็นมีค่ารวมกันเป็น 1 และการจำแนกสามารถจำแนกได้ด้วยตัวแบ่งครึ่งระหว่างแต่ละประเภทได้ง่าย



รูปที่ 12 การส่งไปยังปริภูมิอุดมคติสำหรับการจำแนก [23]

เทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นถูกคิดค้นโดย Sir R.A.Fisher ในปีค.ศ. 1936 [22] โดยใช้แนวคิดของเมตริกซ์การกระจาย (Scatter Matrix) และเกณฑ์ความแบ่งแยกกันได้ (Seperability Criteria) ซึ่งประกอบด้วยเมตริกซ์การกระจายภายในประเภท S_w (Within-class Scatter Matrix) เมตริกซ์การกระจายระหว่างประเภท S_b (Between-class Scatter Matrix) และเมตริกซ์ผสม S_t (Mixture Matrix) ของเมตริกซ์ทั้งสองแบบแรก

โดยกำหนดให้เวกเตอร์ข้อมูล x_i มีค่ามัชฌิมโดยรวม m มีค่ามัชฌิมภายในประเภท i เป็น m_i จำนวนข้อมูลในประเภท i คือ N_i เมตริกซ์ S_w และ S_b สามารถคำนวณได้ดังนี้

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (x_i - m_i)(x_i - m_i)^T \qquad (3.17)$$

$$S_b = \sum_{i=1}^{c} (m_c - m)\,(m_c - m)^T \qquad (3.18)$$

ซึ่งใช้ในฟังก์ชันจุดประสงค์

$$\max \frac{P^T S_b P}{P^T S_w P} \qquad (3.19)$$

โดยที่ P คือเมตริกซ์การฉาย ทั้งนี้สามารถใช้เมตริกซ์ผสม S_t ซึ่งเป็นผลรวมของเมตริกซ์ S_b และ S_w โดยมีค่าเท่ากับเมตริกซ์ความแปรปรวนร่วมของข้อมูลแทนเมตริกซ์การกระจายระหว่างประเภท S_w ได้เช่นกัน

ฟังก์ชันจุดประสงค์นี้สามารถหาค่าเหมาะที่สุดของเมตริกซ์การฉาย P ได้ด้วยการใช้ตัวคูณลากรานจ์ โดยเราจะตั้งสมมติฐานว่าพจน์ $P^T S_w P$ มีค่าเป็น 1 เนื่องจากในฟังก์ชันจุดประสงค์จะไม่ขึ้นกับจำนวนเท่าของเมตริกซ์การฉาย กล่าวคือจะให้ค่าเท่าเดิมไม่ว่าจะแทน P ด้วย kP ที่มีค่าเท่าใดก็ตาม เมื่อ k แทนสัมประสิทธิ์การสเกลเมตริกซ์ P ดังนั้นฟังก์ชันจุดประสงค์จะกลายเป็น

$$\min_{P^T S_w P = 1} -\frac{P^T S_b P}{2} \qquad (3.20)$$

เมื่อใช้ตัวคูณลากรานจ์จะได้สมการในการหาจุดนิ่ง

$$-\frac{P^T S_b P}{2} + \frac{\lambda(P^T S_w P - 1)}{2} = 0 \qquad (3.21)$$

หาอนุพันธ์เทียบกับ P จะได้

$$S_b P = \lambda S_w P \qquad (3.22)$$

ซึ่งเป็นสมการปัญหาค่าลักษณะเฉพาะแบบทั่วไป (Generalized Eigenvalue Problem) ซึ่งถ้าคูณทั้งสองข้างของสมการด้วยตัวผกผันของ S_w จะได้สมการที่คล้ายปัญหาค่าลักษณะเฉพาะ

$$(S_w^{-1} S_b) P = \lambda P \qquad (3.23)$$

อย่างไรก็ตามเมตริกซ์ที่จะแก้ปัญหาค่าลักษณะเฉพาะนี้กลับไม่สมมาตร แต่เมตริกซ์การกระจายระหว่างประเภท S_b นั้นสมมาตรและเป็นบวกจำกัดเขต (Symmetric Positive Definite) ดังนั้นจึงสามารถกระจายออกมาเป็น

$$S_b = U \Delta U^T \qquad (3.24)$$

ซึ่งสามารถหารากที่สองได้เป็น

$$S_b^{\frac{1}{2}} = U\Delta^{\frac{1}{2}}U^T \quad (3.25)$$

เมื่อเรานำรากที่สองของ S_b มาคูณทั้งสองข้างของสมการ (3.23) จะทำให้

$$S_b^{\frac{1}{2}}(S_w^{-1}S_b)P = \lambda S_b^{\frac{1}{2}}P \quad (3.26)$$

เราสามารถแตกรากที่สองของ S_b และนิยามเมตริกซ์การฉายใหม่ P'

$$(S_b^{\frac{1}{2}}S_w^{-1}S_b^{\frac{1}{2}})P' = \lambda P' \quad (3.27)$$

ครั้งนี้เมตริกซ์ที่ต้องการหาค่าลักษณะเฉพาะนั้นสมมาตรดังนั้นเราสามารถแก้สมการนี้เหมือนกับการแก้ปัญหาค่าลักษณะเฉพาะได้ทันทีแล้วจากนั้นจึงหาเมตริกซ์การฉาย P จาก

$$P = S_b^{-\frac{1}{2}}P' \quad (3.28)$$

โดยจะเลือกเวกเตอร์การฉายตามค่าลักษณะเฉพาะที่คู่กับเวกเตอร์ลักษณะเฉพาะจากมากไปน้อย ซึ่งสามารถพิสูจน์ได้จากการแทนค่าในตัวคูณลากรานจ์ในสมการที่ (3.21) ด้วย P^TS_wP มีค่าเป็น 1 และ P^TS_bP มีค่าเป็นส่วนกลับของค่าลักษณะเฉพาะซึ่งพิสูจน์ได้จากการนำตัวสลับเปลี่ยน (Transpose) ของเมตริกซ์การฉาย P หรือ P^T คูณทั้งสองข้างของสมการ (3.22) ซึ่งทำให้ได้ตัวคูณลากรานจ์เป็นตัวติดลบของครึ่งหนึ่งของค่าลักษณะเฉพาะ หมายความว่าค่าลักษณะเฉพาะยิ่งมากยิ่งทำให้ฟังก์ชันจุดประสงค์ (3.20) น้อยนั่นเอง

นอกจากวิธีนี้เรายังสามารถแก้ปัญหาค่าลักษณะเฉพาะแบบทั่วไปได้ด้วยอีกวิธีหนึ่งคือวิธีกำลัง (Power Method) อีกด้วย อย่างไรก็ตามเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นกลับพบปัญหาเมื่อนำมาใช้ในงานประมวลผลภาพ เนื่องจากชุดข้อมูลที่มีจำนวนข้อมูลน้อยแต่มีจำนวนมิติสูงทำให้เมตริกซ์การกระจายภายในประเภท S_w เป็นเอกฐาน (Singular) ทำให้หาตัวผกผันไม่ได้ และหากสามารถหาเวกเตอร์ลักษณะเฉพาะออกมาได้ จะมีเพียง c-1 ตัวเท่านั้นที่มีค่าลักษณะเฉพาะไม่เป็น 0 ปัญหานี้เรียกว่าปัญหาจำนวนตัวอย่างน้อยเกินไป (Small Sample Size Problem)

Belhumeur และคณะได้เสนอวิธีการแก้ปัญหานี้ โดยได้นำการวิเคราะห์องค์ประกอบสำคัญเข้าร่วมด้วย เนื่องจากจำนวนมิติต่ำสุดที่มีอำนาจจำแนก (Discriminative Power) สำหรับข้อมูล c ประเภทคือ c-1 ดังนั้นจึงควรที่จะลดมิติด้วยการวิเคราะห์องค์ประกอบสำคัญให้เหลือ c-1 มิติก่อนเพื่อเป็นการขจัดปริภูมิสู่ศูนย์ (Null Space) ออกไปก่อน ทำให้เมตริกซ์การกระจายภายในประเภท

S_w ไม่เป็นเอกฐานและมีค่าลำดับชั้น (Rank) เป็น c-1 จากนั้นจึงค่อยใช้เทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นกับข้อมูลที่ลดมิติแล้วอีกครั้งหนึ่งในการลดมิติ เทคนิคนี้ให้ชื่อว่าใบหน้าฟิชเชอร์ (Fisherface) [37]

ใบหน้าของฟิชเชอร์จะมีฟังก์ชันจุดประสงค์เป็น

$$\max \frac{P^T(P_{PCA}^T S_b P_{PCA})P}{P^T(P_{PCA}^T S_w P_{PCA})P} \qquad (3.29)$$

เมื่อ P_{PCA} คือเมตริกซ์การฉายที่ได้จากการวิเคราะห์องค์ประกอบสำคัญกับข้อมูลตั้งต้น ซึ่งเราจะนิยามเมตริกซ์การฉาย P' ว่าเป็นผลคูณของเมตริกซ์การฉายของการวิเคราะห์องค์ประกอบสำคัญและการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น $P' = P_{PCA}P$ เพื่อที่จะนำไปใช้ในการลดมิติต่อไปนั่นเอง ทั้งนี้จะลดมิติได้มากที่สุดถึง c-1 มิติเท่านั้น

งานวิจัยชิ้นนี้ถือเป็นการนำเสนอเทคนิคที่ทันสมัยและเป็นงานวิจัยที่ได้รับการอ้างอิงสูงที่สุดในวงการในช่วงคริสทศวรรษที่ 90 ทีเดียว เนื่องจากได้ผลลัพธ์ที่ดีกว่าใบหน้าลักษณะเฉพาะอย่างมีนัยสำคัญ อีกทั้งยังทนต่อสัญญาณรบกวนประเภทแสงเงา (Illumination) ได้ดีเป็นพิเศษอีกด้วย

3.3 การฉายรักษาสมบัติเฉพาะที่หรือแอลพีพี (LPP: Locality Preserving Projection)

เทคนิคการฉายรักษาสมบัติเฉพาะที่หรือแอลพีพี (LPP: Locally Preserving Projection) ถูกนำเสนอโดย He และ Niyogi ในปีค.ศ. 2003 [31] เป็นเทคนิคซึ่งเป็นอนุพันธ์จากกราฟไอเกนแบบลาปลาเชียนหรือแอลอี (LE: Laplacian Eigenmaps) ซึ่งการฉายรักษาสมบัติเฉพาะที่จะเน้นการรักษาโครงสร้างของบริเวณใกล้เคียง (Neighborhood Structure) โดยจะเป็นการประมาณแบบเชิงเส้นของตัวดำเนินการลาปลาซ-เบลทรามี (Laplace-Beltrami Operator) บนข้อมูลท่อร่วม ทำให้สามารถดึงสมบัติของข้อมูลท่อร่วมที่ไม่เป็นเชิงเส้นออกมาในลักษณะการประมาณแบบเชิงเส้นได้ นอกจากนี้การฉายรักษาสมบัติเฉพาะที่จะนิยามทุกๆที่ในปริภูมิโดยรอบ (Ambient Space) ไม่ได้จำกัดเฉพาะชุดฝึกเหมือนกับกราฟไอเกนแบบลาปลาเชียนทำให้เหมาะสมต่อการใช้งานกับชุดข้อมูลจริงอีกด้วย

การฉายรักษาสมบัติเฉพาะที่จะนำทฤษฎีกราฟเชิงสเปกตรัม (Spectral Graph Theory) มาใช้ โดยเป็นการสร้างกราฟน้ำหนักแล้วหาการฉายที่ทำให้แต่ละบัพที่มีเส้นเชื่อมกันอยู่ใกล้กันมากที่สุด ซึ่งจะสามารถเขียนเป็นฟังก์ชันจุดประสงค์ดังนี้

$$\min \sum_{ij} (z_i - z_j)^2 W_{ij} \quad (3.30)$$

โดย z_i และ z_j แทนข้อมูลที่ลดมิติแล้ว เมตริกซ์น้ำหนัก W สามารถสร้างได้ด้วยการสร้างเมตริกซ์ประชิด (Adjacency Matrix) โดยสร้างเส้นเชื่อมระหว่างบัพหากบัพคู่นั้นอยู่ใกล้กันตามหลักการของอัลกอริทึมเพื่อนบ้านใกล้สุด k ตัว (K-Nearest Neighbors) หรืออัลกอริทึมบริเวณใกล้เคียง $\in$ ($\in$-Neighborhoods) จากนั้นจึงให้น้ำหนักแบบเคอร์เนลของเกาส์ด้วยสมการดังนี้

$$W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}} \quad (3.31)$$

โดยกำหนดให้บัพของเวกเตอร์ข้อมูล x_i มีเส้นเชื่อมไปยังบัพของเวกเตอร์ข้อมูล x_j และ t เป็นพารามิเตอร์จำนวนจริงบวกใดๆ อีกทางเลือกหนึ่งคือการให้น้ำหนักแบบ 0-1 โดยให้น้ำหนักเป็น 1 เมื่อมีเส้นเชื่อมระหว่างบัพและให้น้ำหนักเป็น 0 เมื่อระหว่างบัพนั้นไม่มีเส้นเชื่อมระหว่างกัน

จากสมการ (3.30) จะเห็นว่าค่าน้ำหนักจะเป็นเหมือนตัวคูณที่ทำให้ฟังก์ชันจุดประสงค์มีค่าเพิ่มขึ้นมากหากระยะห่างระหว่างข้อมูลที่ลดมิติแล้วมีค่ามากแต่มีเส้นเชื่อมระหว่างกัน ซึ่งสมการ (3.30) สามารถเปลี่ยนรูปได้ดังนี้

$$\begin{aligned}
&\frac{1}{2}\sum_{ij}\left(z_i - z_j\right)^2 W_{ij} \\
&= \frac{1}{2}\sum_{ij}\left(p^T x_i - p^T x_j\right)^2 W_{ij} \\
&= \sum_i p^T x_i D_{ii} x_i^T p - \sum_{ij} p^T x_i W_{ij} x_j^T p \\
&= p^T X(D - W)X^T p \\
&= p^T X L X^T p
\end{aligned} \quad (3.32)$$

เมื่อ X คือเมตริกซ์ของข้อมูล D เป็นเมตริกซ์ดีกรี (Degree Matrix) ซึ่งเป็นเมตริกซ์ทแยงมุม (Diagonal Matrix) ซึ่งช่องในคอลัมน์ i และแถว i จะเป็นผลรวมของแถวหรือคอลัมน์ที่ i ในเมตริกซ์ W L คือเมตริกซ์ลาปลาเชียน จากนั้นจะเพิ่มเงื่อนไขบังคับเพื่อให้เวกเตอร์การฉายถูกทำให้เป็นบรรทัดฐานจนได้ฟังก์ชันจุดประสงค์ดังนี้

$$\min_{p^T X D X^T p = 1} p^T X L X^T p \quad (3.33)$$

โดยสามารถใช้ตัวคูณลากรานจ์ในการหาค่าเหมาะที่สุด ทำให้เป็นปัญหาค่าลักษณะเฉพาะแบบทั่วไปซึ่งสามารถแก้ได้ในทำนองเดียวกันกับเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นดังนี้

$$(XLX^T)p = \lambda(XDX^T)p \qquad (3.34)$$

แต่เนื่องจากเทคนิคการฉายรักษาสมบัติเฉพาะที่ทำการลดค่าฟังก์ชันจุดประสงค์จึงต้องเลือกเวกเตอร์ลักษณะเฉพาะเพื่อใช้เป็นฐานหลักโดยพิจารณาจากค่าลักษณะเฉพาะจากน้อยไปมาก

เทคนิคการฉายรักษาสมบัติเฉพาะที่เป็นเทคนิคแบบไม่สอนเช่นเดียวกับการวิเคราะห์องค์ประกอบสำคัญ แต่สามารถปรับปรุงเมตริกซ์น้ำหนักให้เทคนิคนี้นำข้อมูลประเภทมาประมวลผลแบบสอนได้ โดยนิยามความใกล้ของข้อมูลในการสร้างเส้นเชื่อมด้วยการอยู่ในประเภทเดียวกันแทน

สำหรับงานประมวลผลภาพนั้นบางครั้งเมตริกซ์ XLX^T อาจเป็นเอกฐานซึ่งเกิดจากปัญหาจำนวนตัวอย่างน้อยเกินไป He และคณะได้นำเสนอวิธีแก้ปัญหานี้เรียกว่าใบหน้าลาปลาเชียน (Laplacianface) [38] โดยใช้การวิเคราะห์องค์ประกอบสำคัญในการขจัดปริภูมิสู่ศูนย์ออกไปก่อนคล้ายกับใบหน้าฟิชเชอร์ ซึ่ง He และคณะได้เสนอว่าในการวิเคราะห์องค์ประกอบสำคัญเพื่อลดมิติควรใช้หลักการรักษาข้อมูล 98% เพื่อลดสัญญาณรบกวน โดยสามารถวัดได้จากอัตราส่วนของผลรวมค่าลักษณะเฉพาะของมิติที่จะลดไปหาต่อผลรวมค่าลักษณะเฉพาะทั้งหมดให้มีค่าไม่น้อยกว่า 0.98 นอกจากนี้ยังเสนอว่าในงานประมวลผลภาพนั้นการสร้างเมตริกซ์ประชิดควรใช้อัลกอริทึมเพื่อนบ้านใกล้สุด k ตัว เนื่องจากสามารถกำหนดพารามิเตอร์ได้ง่ายกว่าอัลกอริทึมบริเวณใกล้เคียง $\in$ จากนั้นจึงใช้เทคนิคการฉายรักษาสมบัติเฉพาะที่เพื่อหาเมตริกซ์การฉายต่อไป

ในขณะที่การวิเคราะห์องค์ประกอบสำคัญและการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นมองเฉพาะโครงสร้างระยะทางแบบยูคลิดเท่านั้น การฉายรักษาสมบัติเฉพาะที่ทำการหาตัวแบบของข้อมูลท่อร่วมด้วยแผนภาพเพื่อนบ้านใกล้สุด ทำให้สามารถรักษาสมบัติเฉพาะที่ได้โดยมีผลให้มีอำนาจจำแนกมากกว่าการวิเคราะห์องค์ประกอบสำคัญ

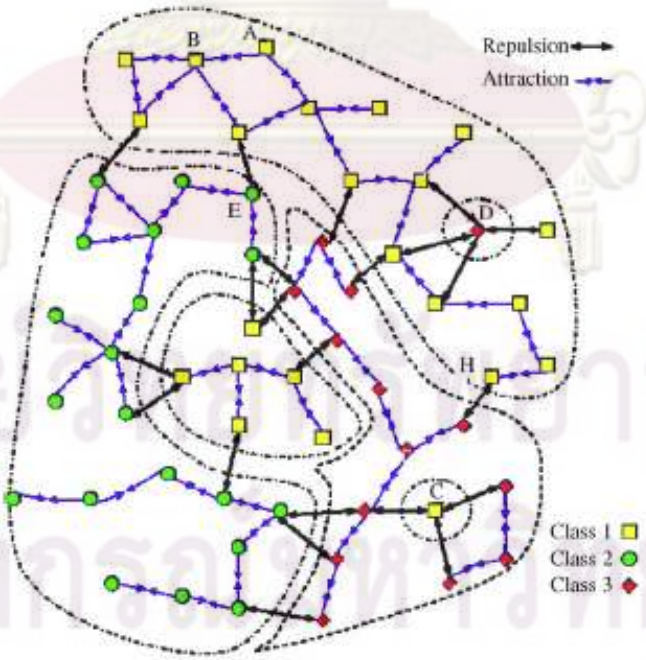
นอกจากนี้ยังสามารถพิสูจน์ได้อีกว่าการวิเคราะห์องค์ประกอบสำคัญและการวิเคราะห์-ดิสคริมิแนนต์เชิงเส้นมาจากหลักการของโครงสร้างแผนภาพในการฉายรักษาสมบัติเฉพาะที่ได้เช่นกัน [38, 39] โดยหากเราให้ทุกๆช่องของเมตริกซ์น้ำหนัก W เป็นส่วนกลับของจำนวนข้อมูลยกกำลังสองและเลือกเวกเตอร์ลักษณะเฉพาะตามค่าลักษณะเฉพาะจากมากไปน้อยจะเห็นว่าผลลัพธ์สมมูลกับการวิเคราะห์องค์ประกอบสำคัญ ทำให้เห็นว่าการเลือกฐานหลักด้วยค่าลักษณะเฉพาะจาก

ค่าน้อยไปมากจะเป็นการรักษาสมบัติเฉพาะที่และหากเลือกจากมากไปน้อยจะเป็นการรักษาสมบัติโดยรวมนั่นเอง

สำหรับการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นนั้น หากเราให้ค่าในแต่ละช่องของเมตริกซ์น้ำหนัก W คือ W_{ij} แทนด้วยส่วนกลับของจำนวนข้อมูลในประเภทหากข้อมูล i และ j อยู่ในประเภทเดียวกัน และให้ค่าเป็น 0 หากอยู่คนละประเภทกัน ทำให้เมตริกซ์ XLX^T สมมูลกับเมตริกซ์การกระจายภายในประเภท S_w และเมตริกซ์ความแปรปรวนร่วม C ลบด้วยเมตริกซ์ XLX^T จะสมมูลกับเมตริกซ์การกระจายระหว่างประเภท S_b อีกด้วย

3.4 ตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์หรือดีเอ็นอี (DNE: Discriminative Neighborhood Embedding)

เนื่องจากเทคนิคการลดมิติต่างๆมักมีความลำบากในการเลือกมิติเป้าหมาย (Target Dimension) ที่จะต้องเลือกเชิงทดลอง ดังนั้น Zhang และคณะจึงได้นำเสนอเทคนิคตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์ในปีค.ศ. 2006 [28, 40] ซึ่งเป็นเทคนิคที่ทำการประมาณค่ามิติเป้าหมายไปพร้อมๆกันกับการเรียนรู้ตัวฝังตัวที่เพื่อนบ้านใกล้สุดในประเภทเดียวกันมีระยะห่างระหว่างกันน้อยกว่าเพื่อนบ้านใกล้สุดที่อยู่คนละประเภทกัน ทั้งนี้จะมองว่าระยะห่างในประเภทเดียวกันเป็นเหมือนแรงดูดและระยะห่างต่างประเภทเป็นเหมือนแรงผลักดังรูปที่ 13



รูปที่ 13 จุดข้อมูลถูกดูดและผลักตามเพื่อนบ้านใกล้สุดที่อยู่ในประเภทเดียวกันและเพื่อนบ้านใกล้สุดที่อยู่ต่างประเภทกันตามลำดับ [40]

นิยามความกระชับภายในประเภท (Intra-class Compactness) $\Delta(P)$ ดังนี้

$$\Delta(P) = \sum_{ij} \left\| P^T x_i - P^T x_j \right\|^2 \quad (3.35)$$

โดยที่ x_i และ x_j เป็นเพื่อนบ้านใกล้สุดที่อยู่ในประเภทเดียวกัน และความแยกกันได้ระหว่างประเภท (Inter-class Seperability) $\delta(P)$ ดังนี้

$$\delta(P) = \sum_{ij} \left\| P^T x_i - P^T x_j \right\|^2 \quad (3.36)$$

โดยที่ x_i และ x_j เป็นเพื่อนบ้านใกล้สุดที่อยู่ต่างประเภทกัน และฟังก์ชันจุดประสงค์จะนิยามจากผลต่างของความกระชับภายในประเภท $\Delta(P)$ และความแยกกันได้ระหว่างประเภท $\delta(P)$ และสามารถแปลงรูปได้ดังนี้

$$\begin{aligned} & min\big(\Delta(P) - \delta(P)\big) \\ & = 2\sum_{ij} \left\| P^T x_i - P^T x_j \right\|^2 F_{ij} \\ & = 2\sum_{i} P^T x_i S_{ii} x_i^T P - \sum_{ij} P^T x_i F_{ij} x_j^T P \\ & = 2P^T X(S-F)X^T P \end{aligned} \quad (3.37)$$

เมื่อ F คือเมตริกซ์ประชิดดิสคริมิแนนต์ (Discriminant Adjacency Matrix) ซึ่งให้น้ำหนักเป็น 1 หาก x_i และ x_j เป็นเพื่อนบ้านใกล้สุดที่อยู่ในประเภทเดียวกัน ($Neig^I$) และให้น้ำหนักเป็น -1 หาก x_i และ x_j เป็นเพื่อนบ้านใกล้สุดที่อยู่ต่างประเภทกัน ($Neig^E$) นอกจากนั้นจะให้น้ำหนักเป็น 0

$$F_{ij} = \begin{cases} 1\,; x_i, x_j \epsilon Neig^I \\ -1\,;\ x_i, x_j \epsilon Neig^E \\ 0\,; otherwise \end{cases} \quad (3.38)$$

และ S เป็นเมตริกซ์ดีกรี (Degree Matrix) ซึ่งเป็นเมตริกซ์ทแยงมุม (Diagonal Matrix) ซึ่งช่องในคอลัมน์ i และแถว i จะเป็นผลรวมของแถวหรือคอลัมน์ที่ i ในเมตริกซ์ F เมื่อเพิ่มเงื่อนไขบังคับจะได้ฟังก์ชันจุดประสงค์ดังนี้

$$\min_{P_i^T P_i = 1, P_i^T P_j = 0 (i \neq j)} P^T X(S-F)X^T P \quad (3.39)$$

โดยเราสามารถหาเมตริกซ์การฉายที่เหมาะที่สุดได้จากการแก้ปัญหาค่าลักษณะเฉพาะของเมตริกซ์ $X(S-F)X^T$ อย่างไรก็ตามเมตริกซ์ $X(S-F)X^T$ นั้นสมมาตรแต่ไม่ได้เป็นบวกจำกัดเขต (Positive Definite) ทำให้มีค่าลักษณะเฉพาะที่เป็นทั้งบวกและลบหรือแม้แต่ค่าเป็น 0 ซึ่งจุดนี้สามารถมองได้ว่าค่าลักษณะเฉพาะนั้นแทนผลต่างของความกระชับภายในประเภท $\Delta(P)$ และความแยกกันได้ระหว่างประเภท $\delta(P)$ เมื่อแทนเมตริกซ์ $X(S-F)X^TP$ ในบรรทัดสุดท้ายของสมการ (3.37) ด้วย λP ตามสมการค่าลักษณะเฉพาะทำให้ได้ผลดังนี้

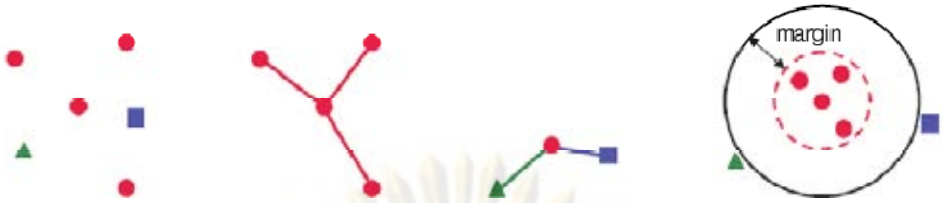
$$\begin{aligned} min\big(\Delta(P) - \delta(P)\big) \\ &= 2P^T X(S-F)X^T P \\ &= 2P^T \lambda P \\ &= 2\lambda \end{aligned} \qquad (3.40)$$

ซึ่งโดยปกติจะคาดหมายให้ความกระชับภายในประเภท $\Delta(P)$ มีค่าน้อยๆและความแยกกันได้ระหว่างประเภท $\delta(P)$ มีค่ามากๆเพื่อให้เหมาะสมต่อการจำแนก ดังนั้นเวกเตอร์ลักษณะเฉพาะที่ควรใช้เป็นฐานหลักควรมีค่าลักษณะเฉพาะเป็นลบโดยยิ่งมีค่าลบมากยิ่งดี ทั้งนี้สามารถคำนวณเป็นอัตราส่วนของผลรวมค่าลักษณะเฉพาะที่เป็นลบของมิติเป้าหมายต่อผลรวมค่าลักษณะเฉพาะที่เป็นลบทั้งหมดเพื่อกำหนดมิติเป้าหมายที่ต้องการได้

นอกจากนี้เทคนิคตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์ยังไม่พบปัญหาเอกฐานในงานประมวลผลภาพ เนื่องจากในการหาค่าเหมาะที่สุดได้เกี่ยวข้องเพียงปัญหาค่าลักษณะเฉพาะของเมตริกซ์เดียวเท่านั้น จึงไม่มีความกังวลเรื่องการหาตัวผกผันแต่อย่างใดอีกด้วย

3.5 การวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่หรือแอลเอสดีเอ (LSDA: Locality Sensitive Discriminant Analysis)

เนื่องจากการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นไม่ได้รักษาสมบัติเฉพาะที่ซึ่งจำเป็นในกรณีที่มีข้อมูลไม่มาก Cai และคณะได้นำเสนอเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่ในค.ศ. 2007 [30] ซึ่งเป็นเทคนิคที่รักษาสมบัติเฉพาะที่และใช้เกณฑ์ระยะขอบเฉพาะที่กว้างที่สุด (Maximum Local Margin Criteria) สำหรับข้อมูลต่างประเภทเพื่อรักษาข้อมูลการจำแนก ทำให้เมื่อส่งข้อมูลเข้าสู่ปริภูมิย่อยด้วยการฉายแล้ว เพื่อนบ้านใกล้สุดที่อยู่ในประเภทเดียวกัน ($Neig^I$) จะอยู่ใกล้ที่สุดและเพื่อนบ้านใกล้สุดที่อยู่ต่างประเภทกัน ($Neig^E$) จะอยู่ห่างออกไปให้มากที่สุด แนวคิดนี้สามารถแสดงได้ดังรูปที่ 14



รูปที่ 14 แนวคิดของเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่ [30]

เทคนิคนี้จะใช้ทฤษฎีกราฟเชิงสเปกตรัมเช่นเดียวกันกับเทคนิคการฉายรักษาสมบัติเฉพาะที่ โดยเริ่มจากการสร้างเมตริกซ์น้ำหนัก W เช่นกัน แต่การวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่จะสร้างเมตริกซ์น้ำหนัก 2 เมตริกซ์ได้แก่เมตริกซ์น้ำหนักภายในประเภท W_w และเมตริกซ์น้ำหนักระหว่างประเภท W_b ดังนี้

$$W_{w_{ij}} = \begin{cases} 1 \, ; x_i, x_j \epsilon Neig^I \\ 0 \, ; otherwise \end{cases} \quad (3.41)$$

$$W_{b_{ij}} = \begin{cases} 1 \, ; x_i, x_j \epsilon Neig^E \\ 0 \, ; otherwise \end{cases} \quad (3.42)$$

ทำให้เราได้ฟังก์ชันจุดประสงค์ 2 สมการที่ต้องการหาค่าเหมาะที่สุด

$$\min \sum_{ij} (z_i - z_j)^2 W_{w_{ij}} \quad (3.43)$$

$$\max \sum_{ij} (z_i - z_j)^2 W_{b_{ij}} \quad (3.44)$$

โดย z_i และ z_j แทนข้อมูลที่ลดมิติแล้ว จะทำการแปลงรูปสมการ (3.43) และ (3.44) เพื่อหารูปแบบมาตรฐานในการหาค่าเหมาะที่สุดดังนี้

$$\begin{aligned} &\frac{1}{2}\sum_{ij} \|z_i - z_j\|^2 W_{w_{ij}} \\ &= \frac{1}{2}\sum_{ij} \left(p^T x_i - p^T x_j\right)^2 W_{w_{ij}} \\ &= p^T X D_{w_{ij}} X^T p - p^T X W_{w_{ij}} X^T p \end{aligned} \quad (3.45)$$

$$\frac{1}{2}\sum_{ij} \|(z_i - z_j)\|^2 W_{b_{ij}} \quad (3.46)$$

$$= \frac{1}{2}\sum_{ij}\left(p^T x_i - p^T x_j\right)^2 W_{b_{ij}}$$

$$= p^T X(D_b - W_b)X^T p$$
$$= p^T X L_b X^T p$$

เมื่อ X คือเมตริกซ์ของข้อมูล D_w, D_b เป็นเมตริกซ์ดีกรี (Degree Matrix) ซึ่งเป็นเมตริกซ์ทแยงมุม (Diagonal Matrix) ซึ่งช่องในคอลัมน์ i และแถว i จะเป็นผลรวมของแถวหรือคอลัมน์ที่ i ใน เมตริกซ์ W_w, W_b ตามลำดับ L_b คือเมตริกซ์ลาปลาเชียน

จากนั้นเมื่อเพิ่มเงื่อนไขบังคับดังนี้

$$Z^T D_w Z = 1 \rightarrow P^T X D_w X^T P = 1 \quad (3.47)$$

ทำให้สมการ (3.45) สามารถแปลงรูปต่อได้เป็น

$$p^T X D_{w_{ij}} X^T p - p^T X W_{w_{ij}} X^T p = 1 - p^T X W_{w_{ij}} X^T p \quad (3.48)$$

มีผลให้ฟังก์ชันจุดประสงค์ (3.43) เปลี่ยนเป็น

$$\min(1 - p^T X W_{w_{ij}} X^T p) \rightarrow \max\ p^T X W_{w_{ij}} X^T p \quad (3.49)$$

เมื่อรวมสมการจุดประสงค์ทั้งสองที่เป็นผลจากสมการ (3.46) และ (3.49) เข้าด้วยกันจะได้สมการจุดประสงค์ดังนี้

$$\max_{P^T X D_w X^T P = 1}\ p^T X(\alpha L_b + (1-\alpha)W_{w_{ij}})X^T p \quad (3.50)$$

เมื่อ α เป็นพารามิเตอร์ที่มีค่าตั้งแต่ 0 ถึง 1 โดยสามารถใช้ตัวคูณลากรานจ์ในการหาค่าเหมาะที่สุด ทำให้เป็นปัญหาค่าลักษณะเฉพาะแบบทั่วไปซึ่งสามารถแก้ได้ในทำนองเดียวกันกับเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นดังนี้

$$(X(\alpha L_b + (1-\alpha)W_{w_{ij}})X^T)p = \lambda(X D_w X^T)p \quad (3.51)$$

เนื่องจากเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่ทำการเพิ่มค่าฟังก์ชันจุดประสงค์จึงต้องเลือกเวกเตอร์ลักษณะเฉพาะเพื่อใช้เป็นฐานหลักโดยพิจารณาจากค่าลักษณะเฉพาะจากมากไปน้อย และในงานประมวลผลภาพเทคนิคนี้ยังคงมีปัญหาเอกฐานอยู่และ Cai และคณะเสนอว่าปัญหา

นี้สามารถแก้ได้โดยใช้การวิเคราะห์องค์ประกอบสำคัญขจัดปริภูมิสู่ศูนย์ออกไปก่อนในทำนองเดียวกันกับใบหน้าฟิชเชอร์และใบหน้าลาปลาเชียน

3.6 เทคนิคการลดมิติแบบสองมิติ (2D Methods)

เทคนิคการลดมิติแบบสองมิตินั้นเป็นการปรับปรุงเทคนิคการลดมิติให้มีความเหมาะสมสำหรับงานประมวลผลภาพโดยเฉพาะ ซึ่งจะไม่แทนข้อมูลด้วยเวกเตอร์เนื่องจากจะทำให้สมบัติเฉพาะที่ของข้อมูลเสียไป แต่จะแทนข้อมูลรูปภาพด้วยเมตริกซ์ที่เหมือนกับข้อมูลรูปภาพดั้งเดิม ซึ่งจะลดมิติด้วยสมการ

$$Z = A^T X \qquad (3.52)$$

เมื่อ A เป็นเมตริกซ์การฉายที่ประกอบด้วยเวกเตอร์ฐานหลักสำหรับการลดมิติ โดย X คือเมตริกซ์ที่แทนข้อมูล 1 ตัว และ Z แทนเมตริกซ์ข้อมูลที่ลดมิติแล้ว

3.6.1 การวิเคราะห์องค์ประกอบสำคัญแบบสองมิติหรือพีซีเอแบบสองมิติ (2DPCA: 2-D Principal Component Analysis)

เทคนิคการลดมิติแบบสองมิติได้ถูกเสนอเป็นครั้งแรกโดย Yang และคณะ ซึ่งได้นำเสนอเทคนิคการวิเคราะห์องค์ประกอบสำคัญแบบสองมิติหรือพีซีเอแบบสองมิติ (2DPCA: 2-D Principal Component Analysis) ในปีค.ศ.2004 [41] โดยใช้การประมวลผลแบบเมตริกซ์ในการหาเมตริกซ์ความแปรปรวนร่วม G ซึ่งสามารถคำนวณได้ดังสมการ (3.53) เมื่อ M แทนเมตริกซ์มัชฌิมของชุดข้อมูล

$$G = \sum_{i=1}^{N} (X_i - M)\,(X_i - M)^T \qquad (3.53)$$

โดยฟังก์ชันจุดประสงค์ยังคงเป็นเช่นเดิมคือ

$$\max_{A^T A=1} A^T G A \qquad (3.54)$$

ทำให้สามารถหาเมตริกซ์การฉาย A จากการแก้ปัญหาค่าลักษณะเฉพาะของเมตริกซ์ G

$$GA = \lambda A \qquad (3.55)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ทั้งนี้เมื่อจะนำข้อมูลไปใช้ในการจำแนก ให้เมตริกซ์ข้อมูลที่ลดมิติแล้ว $Z = [z_1, z_2, \ldots, z_n]$ การวัดความคล้ายจะวัดด้วยระยะห่างแบบยูคลิดของแต่ละแถวดังนี้

$$d(Z^1, Z^2) = \sum_{k=1}^{d} \|z_k^1 - z_k^2\|_2 \qquad (3.56)$$

เทคนิคแบบสองมิตินี้ได้รับการพิสูจน์ว่าให้ผลลัพธ์ที่ดีกว่าเทคนิคแบบหนึ่งมิติอย่างชัดเจน ในกรณีที่มีข้อมูลน้อยๆ โดยได้ให้เหตุผลว่าการสร้างเมตริกซ์ความแปรปรวนร่วมด้วยวิธีแบบสองมิติให้การคำนวณที่ละเอียดกว่าและรักษาสมบัติเฉพาะที่ของใบหน้าได้มากกว่าถือเป็นการนำความรู้ก่อนหน้า (Prior Knowledge) มาใช้ โดยผลลัพธ์ได้แสดงถึงความแม่นในการจำแนกที่สูงกว่าเทคนิคแบบหนึ่งมิติอย่างมีนัยสำคัญ รวมถึงเวลาการฝึกที่น้อยกว่ามากและมีการใช้พารามิเตอร์ที่น้อยกว่าทำให้ใช้เนื้อที่หน่วยความจำในการประมวลผลและเก็บพารามิเตอร์น้อยลงอีกด้วย จากข้อได้เปรียบของเทคนิคแบบสองมิติที่มีต่อเทคนิคแบบหนึ่งมิติเหล่านี้ทำให้เกิดมุมมองว่าเทคนิคประเภทสองมิติจะเป็นแนวคิดใหม่สำหรับการลดมิติทีเดียว

อย่างไรก็ตามเทคนิคแบบสองมิติจะลดมิติเฉพาะในแนวแถวเท่านั้น ซึ่งเป็นผลจากสมการ (3.52) ดังนั้น Zhang และคณะจึงปรับปรุงสมการนี้ใหม่โดยเพิ่มเมตริกซ์การฉายอีก 1 เมตริกซ์ ทำให้ได้สมการการลดมิติใหม่เป็น

$$Z = U^T X V \qquad (3.57)$$

โดย U และ V คือเมตริกซ์การฉายโดยแต่ละสดมภ์เป็นฐานหลักที่แผ่ปริภูมิ L และ R ตามลำดับ ซึ่งเมตริกซ์การฉาย U และ V จะส่งเมตริกซ์ข้อมูล X ไปสู่ปริภูมิ $L \otimes R$ ที่เป็นปริภูมิที่เกิดจากผลคูณเทนเซอร์ของปริภูมิ L และ R

เมตริกซ์การฉาย U นั้นสามารถหาได้จากเทคนิคแบบสองมิติอยู่แล้ว Zhang ได้นำเสนอเทคนิคแบบสองมิติแบบสลับในการหาเมตริกซ์ V ที่จะช่วยลดมิติในแนวสดมภ์ คือเทคนิคการวิเคราะห์องค์ประกอบสำคัญแบบสองมิติแบบสลับ (Alternate 2DPCA) ในปีค.ศ.2005 [42] ซึ่งจะคำนวณเมตริกซ์ความแปรปรวนร่วมดังนี้

$$G = \sum_{i=1}^{N} (X_i - M)^T (X_i - M) \qquad (3.58)$$

ซึ่งฟังก์ชันจุดประสงค์ยังคงนำไปสู่การแก้ปัญหาค่าลักษณะเฉพาะของเมตริกซ์ความแปรปรวนร่วม G เช่นเดิม เมื่อได้เมตริกซ์การฉายทั้งสองแล้วนำมาลดมิติพร้อมกันทั้งแนวแถวและ

แนวสดมภ์จะสามารถลดมิติได้มากขึ้นซึ่งจะเรียกว่าเทคนิคการวิเคราะห์องค์ประกอบสำคัญแบบสองมิติสองทิศทาง ($(2D)^2PCA$) [42]

เทคนิคแบบสองมิติสองทิศทางมีพารามิเตอร์เพิ่มขึ้นจากเทคนิคแบบสองมิติเนื่องจากใช้เมตริกซ์การฉายสองเมตริกซ์ เป็นผลให้มีปริมาณการคำนวณและหน่วยความจำที่จำเป็นสูงขึ้น แต่ผลลัพธ์ที่ได้กลับมาคือความแม่นที่สูงกว่าเทคนิคแบบสองมิติและสามารถลดมิติได้มากกว่า ทำให้ปริมาณหน่วยความจำที่จะใช้เก็บข้อมูลที่ลดมิติแล้วลดลงมากกว่าเทคนิคแบบสองมิติ นอกจากนี้ด้วยความสามารถในการลดมิติที่ลดทั้งแนวแถวและแนวสดมภ์ทำให้แนวคิดของเทคนิคแบบสองมิติสองทิศทางดูเป็นเทคนิคที่สมบูรณ์แบบกว่าเทคนิคแบบสองมิติอีกด้วย

นอกจากแนวทางของ Yang และคณะและ Zhang และคณะแล้ว ยังมีงานวิจัยของ Ye และคณะที่ได้นำเสนอแนวทางในการคำนวณเทคนิคการวิเคราะห์องค์ประกอบสำคัญแบบสองมิติสองทิศทางด้วยวิธีสลับการฉาย (Alternate Projection) ในปีค.ศ.2004 [43, 44] โดยให้ชื่อว่าการประมาณเมตริกซ์ลำดับชั้นน้อยแบบทั่วไปหรือจีแอลอาร์เอเอ็ม (GLRAM: Generalized Low Rank Approximations of Matrices) ซึ่งใช้หลักการประมาณเมตริกซ์ที่มีค่าลำดับชั้นน้อยกว่าเมตริกซ์เดิมด้วยนอร์มของโฟรเบนิอุสเป็นฟังก์ชันจุดประสงค์ดังนี้

$$\min_{\substack{U^TU=I \\ V^TV=I}} \|X - U^TXV\|_F \qquad (3.59)$$

ซึ่งมีรูปคล้ายกับฟังก์ชันจุดประสงค์แบบทำให้ค่าการบิดเบี้ยว (Distortion) เมื่อทำการลดมิติต่ำสุด โดยสามารถแปลงรูปได้ในทำนองเดียวกันจนทำให้มีรูปแบบของการหาค่าเหมาะที่สุดเป็นการคำนวณหาเมตริกซ์การฉายแบบวนซ้ำจากการแก้ปัญหาค่าลักษณะเฉพาะของเมตริกซ์ความแปรปรวนร่วม 2 เมตริกซ์สลับกัน ซึ่งจะคำนวณเมตริกซ์การกระจายบนปริภูมิย่อยที่เกิดจากการฉายบนเมตริกซ์การฉายที่ไม่ได้ทำการหาค่าเหมาะที่สุดในขณะนั้นสลับกันไปจนถึงจำนวนรอบที่กำหนด ในการเริ่มคำนวณเราสามารถกำหนดค่าเริ่มต้นให้กับเมตริกซ์การฉายได้ ซึ่งโดยทั่วไปจะกำหนดให้เป็นเมตริกซ์เอกลักษณ์ I

ทั้งนี้จะคำนวณเมตริกซ์การฉาย U จากการแก้ปัญหาค่าลักษณะเฉพาะของเมตริกซ์ความแปรปรวนร่วมดังนี้

$$G_U = \sum_{i=1}^{N} (X_i - M)\, VV^T (X_i - M)^T \qquad (3.60)$$

สำหรับเมตริกซ์การฉาย V จะคำนวณจากเมตริกซ์ดังนี้

$$G_V = \sum_{i=1}^{N} (X_i - M)^T U^T U (X_i - M) \quad (3.61)$$

จะเห็นว่าเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติด้วยวิธีสลับการฉาย (Alternate Projection) เป็นการแก้ไขการหาค่าเหมาะที่สุดของเทคนิคแบบสองมิติสองทิศทางทำให้มีการคำนวณหลายรอบขึ้นเพื่อให้การคำนวณแม่นที่สุด โดยหวังว่าการวนซ้ำจะช่วยให้ค่าของฟังก์ชันจุดประสงค์จะลู่เข้าสู่ค่าที่ดีที่สุด เนื่องจากปัญหานี้ไม่มีรูปแบบมาตรฐานในการหาค่าเหมาะที่สุดนั่นเอง

3.6.2 การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติหรือแอลดีเอแบบสองมิติ (2DLDA: 2-D Linear Discriminant Analysis)

Li และคณะได้นำเสนอเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติหรือแอลดีเอแบบสองมิติ (2DLDA: 2-D Linear Discriminant Analysis) ในปีค.ศ.2005 [45] โดยใช้การแทนแบบเมตริกซ์ปรับปรุงเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นเดิม และได้รับผลลัพธ์ที่น่าพึงพอใจมากว่าสามารถแก้ปัญหาจำนวนตัวอย่างน้อยเกินไปได้ เนื่องจากเมตริกซ์การกระจายจะมีขนาดเล็กลงทำให้ต้องการจำนวนข้อมูลที่จะให้เมตริกซ์การกระจายภายในประเภทไม่เป็นเอกฐานน้อยลงมาก จึงเหมาะสมต่องานประเภทการรู้จำที่มีตัวอย่างค่อนข้างน้อย

การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติจะคำนวณเมตริกซ์การกระจายภายในประเภท G_w และเมตริกซ์การกระจายระหว่างประเภท G_b ดังนี้

$$G_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (X_i - M_i)(X_i - M_i)^T \quad (3.62)$$

$$G_b = \sum_{i=1}^{c} (M_c - M)\,(M_c - M)^T \quad (3.63)$$

เมื่อค่ามัชฌิมโดยรวมคือ M ค่ามัชฌิมภายในประเภท i คือ M_i จำนวนข้อมูลในประเภท i คือ N_i ซึ่งฟังก์ชันจุดประสงค์จะเหมือนกับเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น

$$\max \frac{A^T G_b A}{A^T G_w A} \quad (3.64)$$

โดยสามารถหาเมตริกซ์การฉาย A จากการแก้ปัญหาค่าลักษณะเฉพาะแบบทั่วไป

$$G_b A = \lambda G_w A \quad (3.65)$$

ต่อมา Noushath และคณะได้นำเสนอเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติแบบสลับ (Alternate 2DLDA) ในปีค.ศ.2006 [46] โดยใช้แนวคิดของการวิเคราะห์องค์ประกอบสำคัญแบบสองมิติแบบสลับในการปรับปรุงการหาเมตริกซ์การกระจายภายในประเภท G_w และเมตริกซ์การกระจายระหว่างประเภท G_b

$$G_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (X_i - M_i)^T (X_i - M_i) \quad (3.66)$$

$$G_b = \sum_{i=1}^{c} (M_c - M)^T (M_c - M) \quad (3.67)$$

ซึ่งฟังก์ชันจุดประสงค์ยังคงนำไปสู่การแก้ปัญหาค่าลักษณะเฉพาะแบบทั่วไปเช่นเดิม เมื่อได้เมตริกซ์การฉายทั้งสองแล้วนำมาลดมิติพร้อมกันทั้งแนวแถวและแนวสดมภ์จะสามารถลดมิติได้มากขึ้นซึ่งจะเรียกว่าเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) [46]

เทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นทั้งแบบสองมิติและแบบสองมิติสองทิศทางเป็นเทคนิคแบบสอนทำให้เหมาะสมต่องานประเภทรู้จำซึ่งต้องการใช้งานข้อมูลประเภทในการจำแนกมากกว่าการวิเคราะห์องค์ประกอบสำคัญแบบสองมิติและแบบสองมิติสองทิศทาง ถึงแม้ว่าจะใช้เวลาและปริมาณคำนวณในการฝึกสูงกว่าแต่จะให้ความแม่นซึ่งเป็นเป้าหมายของงานรู้จำสูงกว่าอย่างมีนัยสำคัญ

นอกจากแนวทางของ Li และคณะและ Noushath และคณะแล้ว ยังมีงานวิจัยของ Ye และคณะที่ได้นำเสนอแนวทางในการคำนวณเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทางด้วยวิธีสลับการฉาย (Alternate Projection) ในปีค.ศ.2004 [47] โดยจะคำนวณหาเมตริกซ์การฉายแบบวนซ้ำเช่นเดียวกันกับเทคนิคการประมาณเมตริกซ์ลำดับชั้นน้อยแบบทั่วไป

เมตริกซ์การกระจายสำหรับการคำนวณเมตริกซ์การฉาย U จะสามารถคำนวณได้ดังนี้

$$G_{U_w} = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (X_i - M_i) V V^T (X_i - M_i)^T \quad (3.68)$$

$$G_{U_b} = \sum_{i=1}^{c} (M_c - M)\, VV^T (M_c - M)^T \qquad (3.69)$$

สำหรับเมตริกซ์การกระจายสำหรับการคำนวณเมตริกซ์การฉาย V จะสามารถคำนวณได้ดังนี้

$$G_{V_w} = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (X_i - M_i)^T U^T U (X_i - M_i) \qquad (3.70)$$

$$G_{V_b} = \sum_{i=1}^{c} (M_c - M)^T U^T U (M_c - M) \qquad (3.71)$$

ซึ่งเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทางด้วยวิธีสลับการฉายจะเป็นเทคนิคต้นแบบที่ภายหลังมีผู้นำแนวคิดการหาค่าเหมาะที่สุดนี้มาประยุกต์ใช้ในการประดิษฐ์เทคนิคแบบสองมิติสองทิศทางต่อไป

3.6.3 การฉายรักษาสมบัติเฉพาะที่แบบสองมิติหรือแอลพีพีแบบสองมิติ (2DLPP: 2-D Locality Preserving Projection)

Chen และคณะได้นำเสนอเทคนิคการฉายรักษาสมบัติเฉพาะที่แบบสองมิติในปีค.ศ.2007 [48] โดยใช้การแทนแบบเมตริกซ์ในทำนองเดียวกันกับเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติ ซึ่งใช้การคำนวณระยะทางด้วยนอร์มของโฟรเบนิอุสในการคำนวณระยะทางเพื่อวัดความคล้าย และแปลงขนาดของเมตริกซ์การฉายที่จะได้จากสมการ (3.34) ของการฉายรักษาสมบัติเฉพาะที่ด้วยการคูณเทนเซอร์กับเมตริกซ์เอกลักษณ์จนได้สมการลักษณะเฉพาะใหม่ดังนี้

$$(X_U (L \otimes I) X_U{}^T)p = \lambda (X_U (D \otimes I) X_U{}^T)p \qquad (3.72)$$

โดยเมตริกซ์ X_U มีขนาด $(p*n)*q$ เกิดจากการคลี่แบบเทนเซอร์ (Tensor Unfolding) ของเทนเซอร์ข้อมูลที่เกิดจากการนำเมตริกซ์ข้อมูล X_i มาซ้อนกันเป็นเมตริกซ์ 3 มิติหรือเทนเซอร์สภาวะที่ 3 เมตริกซ์เอกลักษณ์ I ในสมการนี้มีขนาด $p*p$

สำหรับเทคนิคแบบสองมิติสองทิศทางของการฉายรักษาสมบัติเฉพาะที่นั้น He และคณะได้นำเสนอในแบบฉบับของวิธีสลับการฉายในปีค.ศ.2006 [49] โดยทำการหาค่าเหมาะที่สุดให้กับฟังก์ชันจุดประสงค์ดังนี้

$$\min \sum_{ij} \left\| U^T X_i V - U^T X_j V \right\|^2 W_{ij} \qquad (3.73)$$

จะเห็นว่าจะนำสมการลดมิติของเทคนิคแบบสองมิติสองทิศทางจากสมการ (3.57) มาแทนที่ในข้อมูลที่ลดมิติแล้วเดิมในสมการ (3.30) ฟังก์ชันจุดประสงค์ของการฉายรักษาสมบัติเฉพาะที่ ซึ่งสามารถแปลงคืนแล้วทำการแปลงรูปเพื่อหาค่าเหมาะที่สุดดังนี้

$$\begin{aligned}
&\frac{1}{2}\sum_{ij} \left\| U^T X_i V - U^T X_j V \right\|^2 W_{ij} \\
&\quad = \frac{1}{2}\sum_{ij} tr\left((Y_i - Y_j)(Y_i - Y_j)^T \right) \\
&\quad = tr\left(\sum_i D_{ii} Y_i Y_i^T - \sum_{ij} S_{ij} Y_i Y_i^T \right) \\
&\quad = tr\left(U^T \left(\sum_i D_{ii} X_i V V^T X_i^T - \sum_{ij} S_{ij} X_i V V^T X_j^T \right) U \right) \\
&\quad = tr(U^T (D_V - S_V) U)
\end{aligned} \qquad (3.74)$$

หรือสามารถแปลงได้ในอีกรูปแบบหนึ่งคือ

$$\begin{aligned}
&\frac{1}{2}\sum_{ij} \left\| U^T X_i V - U^T X_j V \right\|^2 W_{ij} \\
&\quad = \frac{1}{2}\sum_{ij} tr\left((Y_i - Y_j)^T (Y_i - Y_j) \right) \\
&\quad = tr\left(\sum_i D_{ii} Y_i^T Y_i - \sum_{ij} S_{ij} Y_i^T Y_i \right) \\
&\quad = tr\left(V^T \left(\sum_i D_{ii} X_i^T U U^T X_i - \sum_{ij} S_{ij} X_i^T U U^T X_j \right) V \right) \\
&\quad = tr(V^T (D_U - S_U) V)
\end{aligned} \qquad (3.75)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ซึ่งเกิดจากสมบัติของค่านอร์มที่สามารถหาได้ 2 แบบจากการคูณเมตริกซ์นั่นเอง He และคณะเสนอว่าควรจะทำการหาค่าเหมาะที่สุดของทั้งสองฟังก์ชันจุดประสงค์พร้อมๆกันและทำการเพิ่มเงื่อนไขบังคับเพื่อให้เวกเตอร์การฉายถูกทำให้เป็นบรรทัดฐานจะทำให้ได้ฟังก์ชันจุดประสงค์ 2 ฟังก์ชันที่จะใช้วิธีสลับการฉายในการหาค่าเหมาะที่สุดของปัญหาค่าลักษณะเฉพาะแบบทั่วไปต่อไปดังนี้

$$\min \frac{tr(U^T(D_V - S_V)U)}{tr(U^T D_V U)} \quad (3.76)$$

$$\min \frac{tr(V^T(D_U - S_U)V)}{tr(V^T D_U V)} \quad (3.77)$$

3.6.4 ตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติหรือดีเอ็นอีแบบสองมิติ (2DDNE: 2-D Discriminant Neighborhood Embedding)

เนื่องจากเทคนิคประเภทสองมิติสองทิศทางสามารถลดมิติได้มากกว่าเทคนิคแบบสองมิติภายหลังเมื่อมีการนำเสนอเทคนิคแบบสองมิติจะใช้เทคนิคประเภทสองมิติสองทิศทางซึ่งหาค่าเหมาะที่สุดด้วยวิธีสลับการฉายเป็นส่วนมาก

Zhang และคณะได้เสนอการแทนแบบเมตริกซ์กับเทคนิคตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์ในปีค.ศ.2007 [50] โดยใช้ฟังก์ชันจุดประสงค์แบบเดียวกันกับการฉายรักษาสมบัติเฉพาะที่แบบสองมิติในสมการ (3.73) แต่จะให้น้ำหนักเมตริกซ์ W ตามแบบฉบับของตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แทน ซึ่งสามารถแปลงรูปด้วยสมบัติของนอร์มได้ 2 แบบเช่นกันทำให้ได้ฟังก์ชันจุดประสงค์ 2 ฟังก์ชันดังนี้

$$\min tr(U^T(S_V - F_V)U) \quad (3.78)$$

$$\min tr(V^T(S_U - F_U)V) \quad (3.79)$$

โดยที่ S_V และ F_V คือ

$$S_V = \sum_i S_{ii} X_i V V^T X_i^T \quad (3.80)$$

$$F_V = \sum_{ij} F_{ij} X_i V V^T X_j^T \quad (3.81)$$

ส่วน S_U และ F_U สามารถหาได้ดังนี้

$$S_U = \sum_i S_{ii} X_i^T U U^T X_i \quad (3.82)$$

$$F_V = \sum_{ij} F_{ij} X_i^T U U^T X_j \quad (3.83)$$

ทั้งนี้การแปลงรูปนั้นจะใช้หลักการเดียวกันกับการแปลงของการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทางในสมการ (3.74) และ (3.75)

3.6.5 การวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติหรือแอลเอสดีเอแบบสองมิติ (2DLSDA: 2-D Locality Sensitive Discriminant Analysis)

Liu และคณะได้นำเสนอการแทนแบบเทนเซอร์กับเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่ในปีค.ศ.2008 [51] ซึ่งเสนอแนวทางสองมิติสองทิศทางแบบปกติแทนที่วิธีสลับการฉายที่มีปัญหาด้านการลู่เข้าในการหาค่าเหมาะที่สุด

กำหนดให้ Z_i และ Z_j แทนข้อมูลที่ลดมิติแล้วและ X_i และ X_j แทนเมตริกซ์ข้อมูล จะทำการแปลงรูปสมการ (3.43) และ (3.44) ด้วยสมการลดมิติแบบสองมิติสองทิศทาง (3.57) เพื่อหารูปแบบมาตรฐานในการหาค่าเหมาะที่สุดดังนี้

$$\begin{aligned}
&\sum_{ij} \|Z_i - Z_j\|^2 W_{w_{ij}} \\
&= \sum_{ij} \|U^T X_i V - U^T X_j V\|^2 W_{w_{ij}} \\
&= \sum_{ij} tr\left(U^T (X_i - X_j) V V^T (X_i - X_j)^T U\right) W_{w_{ij}} \\
&= tr(U^T S_w^V U)
\end{aligned} \quad (3.84)$$

$$\begin{aligned}
&\sum_{ij} \|Z_i - Z_j\|^2 W_{w_{ij}} \\
&= \sum_{ij} \|U^T X_i V - U^T X_j V\|^2 W_{w_{ij}} \\
&= \sum_{ij} tr\left(V^T (X_i - X_j)^T U U^T (X_i - X_j) V\right) W_{w_{ij}} \\
&= tr(V^T S_w^U V)
\end{aligned} \quad (3.85)$$

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สำหรับสมการ(3.44) เราสามารถแทนสมการลดมิติแบบสองมิติสองทิศทาง (3.57) เพื่อหารูปแบบมาตรฐานในการหาค่าเหมาะที่สุดดังนี้

$$\sum_{ij}\|Z_i - Z_j\|^2 W_{b_{ij}}$$

$$= \sum_{ij}\|U^T X_i V - U^T X_j V\|^2 W_{b_{ij}} \qquad (3.86)$$

$$= \sum_{ij} tr\left(U^T(X_i - X_j)VV^T(X_i - X_j)^T U\right) W_{b_{ij}}$$

$$= tr(U^T S_b^V U)$$

$$\sum_{ij}\|Z_i - Z_j\|^2 W_{b_{ij}}$$

$$= \sum_{ij}\|U^T X_i V - U^T X_j V\|^2 W_{b_{ij}} \qquad (3.87)$$

$$= \sum_{ij} tr\left(V^T(X_i - X_j)^T UU^T(X_i - X_j)V\right) W_{b_{ij}}$$

$$= tr(V^T S_b^U V)$$

จากนั้นจะหาเวกเตอร์ลักษณะเฉพาะสลับกันระหว่างการแก้ปัญหาค่าลักษณะเฉพาะแบบทั่วไปของเมตริกซ์ S_w^U และเมตริกซ์ S_b^U กับแก้ปัญหาค่าลักษณะเฉพาะแบบทั่วไปของเมตริกซ์ S_w^V และเมตริกซ์ S_b^V ทั้งนี้ Liu และคณะได้เสนออีกว่าหากเพิ่มเงื่อนไขบังคับให้เวกเตอร์การฉายในแต่ละสดมภ์ของเมตริกซ์ U และ V เป็นเชิงตั้งฉาก (Orthogonal) จะทำให้ใช้วิธีสองมิติสองทิศทางแบบปกติในการหาเวกเตอร์ลักษณะเฉพาะเพื่อใช้เป็นฐานหลักได้อย่างเหมาะที่สุด นอกจากนี้ยังเสนอในการใช้การแยกตัวประกอบแบบเชอร์ (Schur Decomposition) สำหรับการแก้ปัญหาค่าลักษณะเฉพาะอีกด้วย

3.7 เทคนิคประเภทคัดเลือกคุณลักษณะที่เกี่ยวข้อง

3.7.1 คะแนนของฟิชเชอร์ (Fisher Score)

คะแนนของฟิชเชอร์เป็นเทคนิคประเภทคัดเลือกคุณลักษณะแบบสอนประเภทตัวกรองที่จะคำนวณสหสัมพันธ์ (Correlation) ระหว่างคุณลักษณะและข้อมูลประเภทจากเกณฑ์ของฟิชเชอร์

เพื่อวัดความเหมาะสมในการจำแนกของคุณลักษณะนั้นๆ โดยจะมีฟังก์ชันจุดประสงค์เป็นเกณฑ์ของฟิชเชอร์ดังนี้

$$\max \frac{\sum_{i=1}^{c} n_i(\mu_i - \mu)^2}{\sum_{i=1}^{c} n_i \sigma_i^2} \quad (3.88)$$

โดย μ_i แทนค่ามัชฌิมในประเภท i ที่มีข้อมูลอยู่ n_i ตัวจากประเภททั้งหมด c ประเภท μ แทนค่ามัชฌิมโดยรวม σ_i แทนค่าเบี่ยงเบนมาตรฐานในประเภท i

3.7.2 คะแนนลาปลาเชียน (Laplacian Score)

He และคณะได้นำเสนอเทคนิคคะแนนลาปลาเชียนโดยเป็นเทคนิคประเภทคัดเลือกคุณลักษณะแบบไม่สอนประเภทตัวกรอง (Filter) โดยดัดแปลงจากเทคนิคการฉายรักษาสมบัติเฉพาะที่ในปีค.ศ.2005 [8] และให้แนวคิดว่าเทคนิคประเภทตัวหุ้ม (Wrapper) ซึ่งโดยมากจะเป็นเทคนิคแบบไม่สอนนั้นมักจะขึ้นกับตัวเรียนรู้ ส่วนเทคนิคประเภทตัวกรองนั้นมักจะเป็นเทคนิคแบบสอนและขึ้นกับฉลาก (Label) ที่เป็นข้อมูลประเภทมากเกินไป ในขณะที่การรักษาสมบัติเฉพาะที่ (Locality Preserving) สามารถปรับให้เป็นได้ทั้งเทคนิคแบบสอนหรือไม่สอน รวมถึงสมบัติเฉพาะที่ที่เหมาะสมต่อการจำแนกมากกว่าสมบัติโดยรวม

คะแนนลาปลาเชียนจะมีคะแนนที่ดีก็ต่อเมื่อทุกๆคู่จุดข้อมูลที่มีน้ำหนักระหว่างกันอยู่ใกล้กันซึ่งเป็นการเป็นไปตามโครงสร้างแผนภาพ จากแนวคิดนี้ทำให้ได้ฟังก์ชันจุดประสงค์ดังนี้

$$\min \frac{\sum_{ij}(f_{ri} - f_{rj})^2 W_{ij}}{Var(\boldsymbol{f}_r)} \quad (3.89)$$

เวกเตอร์ $\boldsymbol{f}_r$ แทนเวกเตอร์ข้อมูลขนาด $n \times 1$ โดยประกอบด้วยข้อมูลที่คุณลักษณะที่ r ของข้อมูลทุกตัว n แทนจำนวนข้อมูล คะแนนลาปลาเชียนสามารถคำนวณได้โดยกำหนดให้สร้างเมตริกซ์น้ำหนัก W ด้วยวิธีเดียวกันกับการฉายรักษาสมบัติเฉพาะที่ จากนั้นจึงคำนวณคะแนนลาปลาเชียนของคุณลักษณะที่ r ดังนี้

$$L_r = \frac{\tilde{f}_r^T L \tilde{f}_r}{\tilde{f}_r^T D \tilde{f}_r} \quad (3.90)$$

$$\tilde{f}_r = \boldsymbol{f}_r - \frac{\boldsymbol{f}_r^T D 1}{1^T D 1} 1 \quad (3.91)$$

เมื่อ L_r คือคะแนนลาปลาเชียนของคุณลักษณะที่ r เมตริกซ์ D เป็นเมตริกซ์ดีกรีซึ่งเป็นเมตริกซ์ทแยงมุมที่สร้างจากผลรวมในแนวแถวหรือแนวสดมภ์ของเมตริกซ์ W เมตริกซ์ L คือเมตริกซ์ลาปลาเชียนและเวกเตอร์ 1 แทนเวกเตอร์ขนาด $n\times1$ โดยประกอบด้วยค่า 1 ทุกช่อง

ทั้งนี้คะแนนลาปลาเชียนยังคงมีความเกี่ยวเนื่องกับคะแนนของฟิชเชอร์ในทำนองเดียวกับความเกี่ยวเนื่องระหว่างการฉายรักษาสมบัติเฉพาะที่และการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น โดยสามารถให้น้ำหนักเมตริกซ์น้ำหนักแบบสอนเพื่อให้ผลลัพธ์ของคะแนนลาปลาเชียนสมมูลกับคะแนนของฟิชเชอร์ได้เช่นเดียวกัน

3.8 การแยกไบแอส-แวเรียนซ์ (Bias-Variance Decomposition)

ในตัวแบบสำหรับการเรียนรู้ของเครื่องต่างๆนั้นมีความหลากหลายในความซับซ้อนมาก โดยตัวแบบที่มีความซับซ้อนมากก็จะมีความเสี่ยงที่จะปรับเหมาะเกินไปกับชุดฝึก (Overfit) ส่วนตัวแบบที่มีความซับซ้อนน้อยนั้นแม้จะมีความเสี่ยงต่อการปรับเหมาะเกินไปกับชุดฝึกน้อยกว่า แต่กลับมีความเสี่ยงที่จะทำให้ตัวแบบมีความยืดหยุ่นไม่เพียงพอที่จะเรียนรู้ข้อมูลสำคัญและแนวโน้มของข้อมูลเป็นการปรับไม่เหมาะกับชุดฝึก (Underfit) ทั้งนี้ในการที่จะหลีกเลี่ยงข้อเสียของตัวแบบต่างๆ เราจำเป็นที่จะเรียนรู้ถึงทฤษฎีความซับซ้อนของตัวแบบซึ่งคือการแลกเปลี่ยนไบแอส-แวเรียนซ์ (Bias-Variance Tradeoff)[34, 52]

ฟังก์ชันสูญเสีย (Loss Function) ในปัญหาการถดถอย (Regression) มักจะใช้การคาดหมายฟังก์ชันสูญเสียแบบกำลังสอง (Expected Square Loss Function) ซึ่งสามารถเขียนในรูปผลบวกของพจน์ 3 พจน์ได้แก่ ไบแอส แวเรียนซ์ และสัญญาณรบกวน (Noise)

$$E[Loss] = (Bias)^2 + Variance + Noise \quad (3.92)$$

$$(Bias)^2 = \int \{E_D[y(x;D)] - h(x)\}^2 p(x)dx \quad (3.93)$$

$$Variance = \int E_D[\{y(x;D) - E[y(x;D)]\}^2]p(x)dx \quad (3.94)$$

$$Noise = \iint \{h(x) - t\}^2 p(x,t)dxdt \quad (3.95)$$

เมื่อ $y(x)$ คือสมมติฐานที่เหมาะที่สุด $h(x)$ คือสมมติฐานที่ตัวแบบที่เหมาะที่สุด t คือประเภทของข้อมูล x และ D คือชุดข้อมูล ในการลดมิตินี้สามารถช่วยลดแวเรียนซ์และสัญญาณ

รบกวนได้ ดังนั้นการตรวจวัดประสิทธิภาพของการปรับปรุงเทคนิคการลดมิติ จึงสามารถวัดจากการเปลี่ยนแปลงของพจน์ต่างๆของฟังก์ชันสูญเสียแบบกำลังสอง

การประมาณการกระจายไบแอส-แวเรียนซ์มักจะใช้วิธีการชักตัวอย่าง (Resampling) ในการประมาณ โดยทั่วไปมักใช้เทคนิคมีดพก (Jackknife) หรือเทคนิคการปลุกเครื่อง (Bootstrap) ในการประมาณ

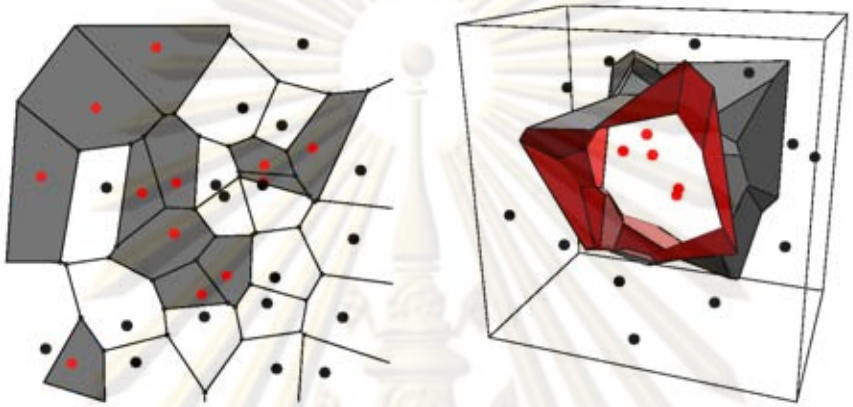
เทคนิคมีดพกเป็นวิธีการชักตัวอย่างเพื่อประมาณค่าทางสถิติโดยจะใช้วิธีนำตัวหนึ่งออกไป (Leave-one-out) ในการสร้างชุดของตัวประมาณ (Estimator) แล้วนำมาเฉลี่ยเพื่อสร้างตัวประมาณสุดท้ายที่เป็นคำตอบ ยกตัวอย่างเช่นการประมาณค่ามัชฌิมจะประมาณจากผลรวมของข้อมูลที่ถูกนำตัวหนึ่งออกไปซึ่งมีจำนวนชุดเท่าจำนวนข้อมูล จากนั้นจึงนำค่าที่ประมาณได้ในทุกๆชุดมาเฉลี่ยจะได้ค่าประมาณด้วยวิธีมีดพกออกมา ส่วนเทคนิคการปลุกเครื่องนั้นจะใช้วิธีสร้างชุดของตัวประมาณด้วยวิธีสุ่มหยิบแบบสับเปลี่ยน คือจะสุ่มเลือกข้อมูลขึ้นมาเป็นจำนวน B ชุดแต่ละชุดมีจำนวนข้อมูลเท่าๆกันแล้วจึงสร้างตัวประมาณเพื่อนำมาเฉลี่ยเป็นตัวประมาณสุดท้ายที่เป็นคำตอบ

ในปีค.ศ. 1998 Bauer และ Kohavi ได้นำเสนอวิธีประมาณการกระจายไบแอส-แวเรียนซ์ที่ถือว่าแม่นยำกว่าวิธีอื่นๆในขณะนั้น [53] โดยจะใช้กระบวนการชักตัวอย่างสองขั้นคือในขั้นแรกจะดึงข้อมูลบางส่วนออกไปเป็นชุดทดสอบก่อน จากนั้นจึงทำการสุ่มข้อมูลจำนวนครึ่งหนึ่งคล้ายเทคนิคการปลุกเครื่องเพื่อสร้างตัวประมาณค่าไบแอสและแวเรียนซ์ซึ่งจะทดสอบความแม่นบนชุดทดสอบ ทั้งนี้ Bauer และ Kohavi เสนอว่าควรสุ่มข้อมูล 10 ครั้งมาเฉลี่ยกัน แล้วจึงทำซ้ำทั้งกระบวนการใหม่อีกอย่างน้อย 3 ครั้งเพื่อความเสถียรของการประมาณ แต่อย่างไรก็ตามในการทดลองเราไม่สามารถวัดสัญญาณรบกวนได้เนื่องจากเราไม่ทราบสมมติฐานตัวแบบที่เหมาะที่สุด ดังนั้นจึงแทนสมมติฐานตัวแบบที่เหมาะที่สุดด้วยข้อมูลประเภททำให้พจน์ไบแอสและสัญญาณรบกวนถูกวัดขึ้นมาพร้อมกัน

3.9 อัลกอริทึมเพื่อนบ้านใกล้สุด *k* ตัว (k-Nearest Neighbors)

อัลกอริทึมเพื่อนบ้านใกล้สุด k ตัวเป็นตัวจำแนกที่เป็นแบบไม่อิงพารามิเตอร์ (Non-parametric) โดยจะนำข้อมูลทดสอบมาเทียบหาความคล้ายกับข้อมูลฝึกด้วยมาตรวัดต่างๆเช่น ระยะทางแบบมหาลาโนบิส ระยะทางแบบแมนฮัตตัน เป็นต้น และจะอนุมานว่าข้อมูลทดสอบมีประเภทเดียวกันกับข้อมูลฝึกที่คล้ายกันมากที่สุดกล่าวคือมีระยะห่างน้อยที่สุดนั่นเอง ดังนั้นหากใช้จำนวนเพื่อนบ้านเป็น k ตัวจะหมายถึงการหาฝ่ายข้างมาก (Majority) จากเพื่อนบ้านเหล่านั้นนั่นเอง

ทั้งนี้ค่า k จะเป็นตัวกำหนดความสม่ำเสมอของตัวจำแนกทำให้มีบริเวณที่จำแนกว่าเป็นประเภทเดียวกันน้อยลงหากค่า k สูงขึ้น นอกจากนี้ยังมีสมบัติที่น่าสนใจคือหากกำหนดให้ k มีค่าเป็น 1 บริเวณจำแนกจะเป็นเทสเซลเลชันแบบโวโรนอย (Voronoi Tesselation) ดังรูปที่ 15 และหากจำนวนตัวอย่างเข้าใกล้อนันต์แล้วอัตราความผิดพลาดของการจำแนกจะไม่มากกว่าสองเท่าของตัวจำแนกที่เหมาะสมที่สุดซึ่งคือตัวจำแนกที่ใช้การกระจายประเภทที่แท้จริง



รูปที่ 15 บริเวณจำแนกของกฎเพื่อนบ้านใกล้สุด 1 ตัวเป็นเทสเซลเลชันแบบโวโรนอยด์ [52]

อัลกอริทึมเพื่อนบ้านใกล้สุด 1 ตัวจะเป็นตัวจำแนกที่นิยมใช้ในการวัดประสิทธิภาพในการช่วยจำแนกของเทคนิคการลดมิติ เนื่องจากจะวัดที่ระยะห่างของข้อมูลที่ลดมิติแล้วโดยตรง ทำให้ใกล้เคียงกับการวัดประสิทธิภาพการช่วยจำแนกเมื่อลดมิติตามฟังก์ชันจุดประสงค์ของเทคนิคการลดมิตินั้นๆมากที่สุด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

การเลือกฐานหลักสำหรับเทคนิคแบบสองมิติและสองมิติสองทิศทางด้วยเทคนิคการคัดเลือกคุณลักษณะ

4.1 แรงจูงใจ

ถึงแม้ว่าในปัจจุบันเทคนิคการลดมิติสำหรับงานประมวลผลภาพที่เป็นแบบสองมิติได้รับความสนใจในวงการมากขึ้น แต่ก็ยังมีข้อเปรียบเทียบว่าแท้จริงแล้วนั้นเทคนิคแบบสองมิติยังคงมีข้อจำกัดหลายประการที่อาจทำให้ไม่สะดวกต่อการนำไปใช้งานจริงได้เหมือนเทคนิคแบบหนึ่งมิติ

Z. Wang และคณะ [54] ได้ทำการทดลองในประเด็นนี้โดยได้เปรียบเทียบเทคนิคการลดมิติแบบหนึ่งมิติและสองมิติซึ่งใช้ตัวจำแนกทั้งแบบหนึ่งมิติและสองมิติในการจำแนกโดยได้ผลลัพธ์ว่าการใช้เทคนิคการลดมิติแบบสองมิติและตัวจำแนกแบบสองมิติร่วมกันจะให้ผลลัพธ์ที่ดีมากสำหรับชุดข้อมูลซึ่งทราบโครงสร้างว่าเหมาะกับการแทนแบบเมตริกซ์อยู่แล้วเช่นข้อมูลรูปภาพ โดยได้ทดสอบทุกๆการจัดหมู่ใช้ร่วมกันของเทคนิคการลดมิติและตัวจำแนก

นอกจากนี้งานทดลองเทคนิคแบบสองมิติสองทิศทางได้พิสูจน์ทางทฤษฎีว่าเทคนิคประเภทสองมิติสองทิศทางเมื่อใช้กับอัลกอริทึมเพื่อนบ้านใกล้สุด 1 ตัว (1-NN: 1-Nearest Neighbor) จะให้ผลดีกว่าเทคนิคประเภทสองมิติ แต่ผลการทดลองกับตัวจำแนกแบบสองมิติกลับไม่เป็นไปตามข้อพิสูจน์นั้นโดยได้ให้เหตุผล 2 ข้อคือ

1. อัลกอริทึมเพื่อนบ้านใกล้สุด 1 ตัวที่นำมาใช้พิสูจน์นั้นใกล้เคียงกับการเป็นตัวจำแนกแบบมิติเดียว ซึ่งสมบัติเฉพาะที่มักไม่ได้ช่วยในการจำแนกมากนัก แต่ตัวจำแนกแบบสองมิติซึ่งไวต่อสมบัติเหล่านี้กลับแสดงผลออกมาว่าเทคนิคแบบสองมิติดีกว่าเทคนิคแบบสองมิติสองทิศทาง เนื่องจากยังคงสมบัติเชิงโครงสร้างของข้อมูลดั้งเดิมไว้มากกว่า ซึ่งสามารถสังเกตจากข้อมูลที่ลดมิติแล้วด้วยตาเปล่า
2. เงื่อนไขบังคับของตัวจำแนกแบบสองมิติต้องใช้การแยกแบบโครเน็กเกอร์ (Kroneckor product decomposition) 2 พับแต่ตัวจำแนกแบบหนึ่งมิติใช้เพียงพับเดียวทำให้รองรับการรักษาสมบัติเฉพาะที่น้อยกว่า

L. Wang และคณะ [55] ได้พิสูจน์ความสมมูลกันของเทคนิคประเภทหนึ่งมิติและสองมิติซึ่งได้อ้างว่าในปี 1994 ได้เคยมีผู้นำเสนอการใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov

Model) ในการแบ่งข้อมูลเป็นบล็อกย่อยๆแล้วจึงใช้การวิเคราะห์องค์ประกอบสำคัญ การวิเคราะห์องค์ประกอบสำคัญแบบสองมิตินั้นจึงนับเป็นกรณีเฉพาะหนึ่งของงานนี้โดยมองการทำให้เป็นสองมิติเป็นการแบ่งข้อมูลภาพเป็นบล็อกที่มีลักษณะเป็นเส้น (Line Block) จากนั้นจึงพิสูจน์ว่าการหาเมตริกซ์ความแปรปรวนร่วมนั้นได้เมตริกซ์เดียวกัน ซึ่งเป็นจริงในทำนองเดียวกันสำหรับการคำนวณเมตริกซ์การกระจายในเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น

Lieng และคณะ [56] ได้ทำการโต้แย้งว่าเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิตินั้นไม่ได้ดีกว่าเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบหนึ่งมิติและอำนาจจำแนกกลับต่ำกว่าแบบหนึ่งมิติเมื่อลดมิติไปยังมิติขนาดเท่ากัน อีกทั้งเมื่อทดลองในชุดข้อมูลขนาดใหญ่เทคนิคแบบสองมิติกลับให้ความแม่นที่น้อยกว่าและเทคนิคแบบหนึ่งมิติแบบปริภูมิสู่ศูนย์ (Nullspace) กลับใช้ได้ดีเทียบเท่ากับเทคนิคแบบสองมิติในชุดข้อมูลที่มีขนาดเล็กเช่นกันทำให้เกิดข้อสงสัยว่าการแก้ปัญหาจำนวนตัวอย่างน้อยเกินไป (Small Sample Size Problem) นั้นเทคนิคแบบสองมิติอาจไม่ได้แก้ปัญหานี้ได้ดีกว่าเทคนิคแบบหนึ่งมิติที่ใช้วิธีปริภูมิสู่ศูนย์

Zheng และคณะ [57] ได้ทำการทดลองทั้งทางทฤษฎีและปฏิบัติในการเปรียบเทียบเทคนิคการวิเคราะห์ดิสคริมิแนนต์แบบหนึ่งมิติและสองมิติซึ่งเป็นเทคนิคที่ได้ผลในการรู้จำอันเป็นที่นิยมที่สุด โดยได้ข้อสรุป 4 ข้อดังนี้

1. ในกรณีที่ข้อมูลกระจายตัวด้วยค่าความแปรปรวนไม่เท่ากัน (Heteroscedastic) เทคนิคแบบสองมิติจะได้รับผลกระทบมากกว่าเทคนิคแบบหนึ่งมิติ
2. เทคนิคแบบสองมิติทำให้ข้อมูลความแปรปรวนระหว่างแถวเสียไป ขณะที่เทคนิคแบบหนึ่งมิติยังคงข้อมูลเหล่านี้อยู่ อีกทั้งยังได้อำนาจจำแนกตามเกณฑ์ของฟิชเชอร์ (Fisher Criteria) สูงกว่าอีกด้วย
3. เทคนิคแบบสองมิติจะเหมาะที่สุดแบบเบย์ (Bayes Optimal) เมื่อแต่ละแถวในรูปภาพเป็นอิสระต่อกันเท่านั้น
4. เทคนิคแบบสองมิติอาจไม่ได้ดีกว่าในกรณีที่มีข้อมูลน้อยๆหรือลดมิติลงมาเหลือมิติต่ำๆจริง โดยเปรียบเทียบว่าเทคนิคแบบหนึ่งมิติที่ทำให้สม่ำเสมอ (Regularized) จะให้ความแม่นในการจำแนกสูงกว่า และการหาจำนวนมิติที่เหมาะที่สุดที่จะลดมิติลงมาสำหรับการจำแนกนั้นทำได้ยากกว่าในเทคนิคแบบสองมิติ

ด้วยปัญหาเหล่านี้ทำให้เกิดแนวคิดในการปรับปรุงเทคนิคประเภทสองมิติสองทิศทางขึ้นเนื่องจากเป็นเทคนิคที่ใช้เวลาฝึกไม่มากแต่ให้ความแม่นที่สูง ใช้การแทนแบบเมตริกซ์ที่สอดคล้องกับข้อมูลรูปภาพและสามารถลดมิติได้มากที่สุด จึงควรที่จะมีประสิทธิภาพสูงกว่าเทคนิคแบบสอง

มิติหรือแม้แต่แบบหนึ่งมิติ ซึ่งเทคนิคแบบสองมิติสองทิศทางนั้นได้เปรียบในด้านที่สามารถรองรับขนาดของภาพที่ใหญ่ขึ้นได้ดีกว่า แต่ในด้านจำนวนของข้อมูลที่สูงขึ้นยังคงเสียเปรียบวิธีที่เหลืออยู่

ทั้งนี้มีงานวิจัยที่นำเสนอในประเด็นเรื่องการเลือกฐานหลักอาจทำให้ความแม่นสูงขึ้นดังนี้

Nyugen และคณะ [58] ได้ศึกษาปรากฏการณ์ที่เกิดจากการสุ่มเลือกฐานหลักในการวิเคราะห์องค์ประกอบสำคัญแบบสองมิติในการจำแนกพบว่าเป็นถึงส่วนมากที่ค่าความแม่นในการจำแนกของการสุ่มเลือกฐานหลักจะสูงกว่าการเลือกตามค่าลักษณะเฉพาะที่แสดงถึงความแปรปรวน อย่างไรก็ตามผลลัพธ์จากงานวิจัยนี้อาจเป็นเพราะการวิเคราะห์องค์ประกอบสำคัญยังเป็นเทคนิคแบบไม่สอนจึงไม่แน่นอนที่ความแปรปรวนจะแสดงถึงอำนาจจำแนก แต่โดยทั่วไปไม่น่าจะเป็นไปได้ที่การอ้างอิงอำนาจจำแนกตามความแปรปรวนจากมากไปน้อยจะให้ผลน้อยกว่าการสุ่มเลือกฐานหลักโดยส่วนมาก

Belhumeur และคณะ [37] ยังได้กล่าวถึงประเด็นนี้สำหรับเทคนิคใบหน้าลักษณะเฉพาะ (Eigenface) ว่าเวกเตอร์ฐานหลักที่มีค่าลักษณะเฉพาะเป็นลำดับแรกๆอาจเกี่ยวเนื่องกับสัญญาณรบกวนประเภทแสงเงา และการตัดเวกเตอร์เหล่านั้นทิ้งก่อนจะทำให้ได้ผลลัพธ์ที่ดีขึ้น โดยในงานวิจัยได้ทดลองตัดเวกเตอร์ฐานหลักที่มีค่าลักษณะเฉพาะมากเป็น 3 อันดับแรกของการวิเคราะห์องค์ประกอบสำคัญออกไปก่อนจึงเลือกฐานหลักมาใช้พบว่าให้ความแม่นสูงขึ้นกว่าการเลือกฐานหลักตามค่าลักษณะเฉพาะ ซึ่งได้ให้สมมติฐานว่าฐานหลัก 3 อันดับแรกอาจเป็นข้อมูลของสัญญาณรบกวน แสงเงา ท่าทาง ทำให้ได้ผลดีเฉพาะในชุดฝึกเท่านั้น

ดังนั้นจึงมีแรงจูงใจที่จะศึกษาการเลือกฐานหลักในเทคนิคการลดมิติแบบสองมิติสองทิศทางว่าอาจมีวิธีที่สามารถประมาณลำดับของฐานหลักใหม่ให้ได้ความแม่นสูงขึ้นได้

4.2 แนวคิด

จากแรงจูงใจที่ได้กล่าวมาสามารถสรุปได้ว่าเทคนิคประเภทสองมิติสองทิศทางยังคงมีข้อด้อยที่สามารถแก้ไขให้ดีขึ้นได้ดังนี้

1. การหาค่าเหมาะที่สุดไม่เหมาะสม เนื่องจากเทคนิคประเภทสองมิติสองทิศทางนั้นจะใช้เมตริกซ์การฉายจำนวน 2 เมตริกซ์ แต่การคำนวณเมตริกซ์เหล่านั้นกลับมองข้อมูลเป็นบล็อกทำให้การคำนวณค่อนข้างหยาบ รักษาสมบัติเฉพาะที่ได้น้อยลง และเมื่อใช้การหาค่าเหมาะที่สุดกลับแยกกระบวนการหาเมตริกซ์การฉายไม่ได้หาไปด้วยกัน ไม่ว่าจะเป็นแบบใช้เทคนิคแบบสลับแล้วนำเมตริกซ์ที่ได้มาใช้งานทันที หรือใช้วิธีสลับการฉายที่เป็นกระบวนการทำซ้ำโดยตรึงเมตริกซ์ที่

ไม่ได้กำลังหาค่าเหมาะที่สุดเอาไว้ก่อนก็ตาม แนวคิดในการหาปริภูมิย่อยที่เกิดจากการฉายไปบนฐานหลัก 2 ฐานพร้อมกันนั้น กลับไม่ได้รับประกันว่าจะได้ค่าที่ดีที่สุดตามแนวคิดของเทคนิคที่ควรจะเป็นจริงๆ เพราะแต่ละครั้งเป็นการคำนวณเพื่อลดมิติด้วยผลรวมเชิงเส้นของข้อมูลที่เป็นบล็อกที่มีลักษณะเป็นเส้น แต่ทุกครั้งที่คำนวณกลับคำนวณเพื่อลดมิติโดยไม่คำนึงถึงความสัมพันธ์ของข้อมูลเฉพาะที่ และการคำนวณแบบสลับนั้นไม่แน่ว่าจะลู่เข้าสู่ค่าเหมาะที่สุดซึ่งทำให้ได้ค่าที่เหมาะที่สุดเฉพาะที่เท่านั้น

ในประเด็นนี้จะแก้ไขโดยนำเทคนิคการลดมิติประเภทคัดเลือกคุณลักษณะ (Feature Selection) มาร่วมด้วย โดยการทำเมตริกซ์การฉายทั้งสองให้เหมาะสมที่สุดถึงแม้ว่าเราจะยังไม่สามารถหารูปแบบปิดของการหาค่าเหมาะที่สุด (Closed Form Optimization) แต่เราสามารถทำให้ผลลัพธ์โดยรวมดีขึ้นได้โดยวัดสมบัติของข้อมูลให้ละเอียดขึ้นและจัดเรียงฐานหลักใหม่ให้เกิดการร่วมมือกันระหว่างฐานหลักที่ใช้ในการฉายในข้อมูลที่อยู่ในมิติที่ลดแล้วใหม่ มีผลให้การฉายช่วยเพิ่มความแม่นเป็นการปรับปรุงกระบวนการส่วนของการเลือกตัวแบบ (Model Selection) โดยใช้เทคนิคประเภทคัดเลือกคุณลักษณะที่เป็นอนุพันธ์จากคะแนนของฟังก์ชันจุดประสงค์แทนค่าลักษณะเฉพาะตามปกติ

2. ในการประมาณค่ามิติเป้าหมาย (Target Dimension Estimation) เพื่อกำหนดให้เป็นค่าคงที่เพื่อลดมิติในการใช้งานจริงไม่เหมาะสม โดยมีสมมติฐานว่าค่าลักษณะเฉพาะที่ใช้ในการประมาณค่ามิติเป้าหมายตามปกตินี้ไม่สามารถเป็นบรรทัดฐานที่แทนค่าของฟังก์ชันจุดประสงค์ที่ใช้ในการวัดได้ เนื่องจากการหาค่าเหมาะที่สุดไม่ได้ลู่เข้าค่าเหมาะที่สุดโดยรวม (Global Optimum) อีกข้อหนึ่งคือปัญหาด้านการวัดเนื่องจากค่าที่ใช้ในการวัดอำนาจจำแนกของคุณลักษณะมีน้อยกว่าจำนวนคุณลักษณะ โดยจะวัดคุณลักษณะเป็นกลุ่มซึ่งจะมีคุณลักษณะหนึ่งที่ถูกวัดด้วยค่าลักษณะเฉพาะ 2 ค่า ค่าลักษณะเฉพาะจะบ่งชี้ถึงอำนาจจำแนกเป็นกลุ่มของข้อมูลเป็นบล็อกที่มีลักษณะแบบเส้นแล้วเลือกฐานหลักที่จะคงไว้ในการลดมิติตามความสามารถในการให้อำนาจจำแนกในกลุ่มข้อมูลนี้ ซึ่งขัดกับวิธีการเลือกตัวแบบโดยทั่วไปของข้อมูลที่มีในเทคนิคแบบหนึ่งมิติที่ค่าลักษณะเฉพาะจะมีเท่าจำนวนคุณลักษณะทำให้ค่าลักษณะเฉพาะวัดคุณลักษณะที่เป็นอิสระต่อกันและจะคัดคุณลักษณะที่มีค่าลักษณะเฉพาะขนาดน้อยๆออก ทำให้เหลือแต่คุณลักษณะที่มีอำนาจจำแนกสูงๆตามฟังก์ชันจุดประสงค์

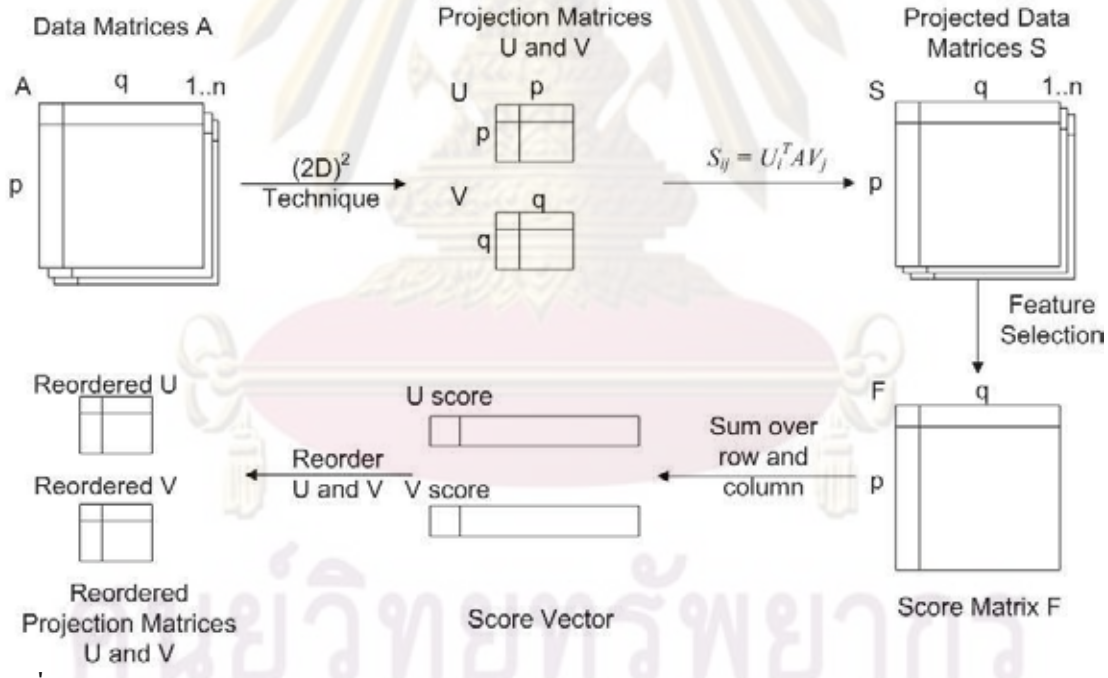
โดยมีแนวคิดจะใช้เทคนิคประเภทคัดเลือกคุณลักษณะประเภทตัวกรอง (Filter) ทำการวัดอำนาจจำแนกของฐานหลักใหม่ตามฟังก์ชันจุดประสงค์ ด้วยเหตุว่าเทคนิคแบบสองมิติสองทิศทางสามารถคำนวณได้เร็วและรองรับข้อมูลจำนวนมากๆได้ดีกว่าเทคนิคแบบหนึ่งมิติ การเพิ่มกระบวนการใหม่เข้าไปอาจใช้เวลามากขึ้นแลกกับความแม่น แต่ฐานหลักที่เป็นผลลัพธ์ของการหา

ค่าเหมาะที่สุดควรได้รับการวัดอำนาจจำแนกใหม่ด้วยการให้คะแนนแต่ละฐานหลักแล้วจึงทำการเลือกฐานหลักที่สร้างคุณลักษณะที่มีอำนาจจำแนกสูงๆเพื่อให้เป็นมิติที่เหลืออยู่หลังการลดมิติ

ซึ่งปัญหาที่เกิดจากการมองข้อมูลเป็นบล็อกนั้น เทคนิคแบบสองมิติก็ประสบปัญหานี้ในทำนองเดียวกัน ดังนั้นการใช้แนวคิดนี้ทดสอบในเทคนิคแบบสองมิติก็ควรที่จะให้ผลลัพธ์ที่ดีขึ้นเช่นกัน โดยอาจทำให้เกิดมุมมองใหม่ๆที่จะเปรียบเทียบเทคนิคแบบสองมิติและสองมิติแบบสองทิศทางว่าเทคนิคใดเหมาะสมกับงานรู้จำมากที่สุด

4.3 การปรับปรุงเทคนิคแบบสองมิติสองทิศทาง

โดยแนวคิดที่จะใช้เทคนิคประเภทคัดเลือกคุณลักษณะร่วมด้วยนั้น ได้ตั้งสมมติฐานว่าควรใช้กับข้อมูลในปริภูมิย่อยที่เกิดจากการลดมิติแล้วเพื่อให้เป็นการวัดผลตัวแบบขั้นสุดท้ายก่อนนำไปใช้งานจริง โดยเน้นวัดในปริภูมิย่อยซึ่งเกิดจากคู่ฐานหลักเพื่อให้เป็นการวัดความสามารถของฐานหลักทั้งสองเมื่อทำการลดมิติร่วมกันดังรูปที่ 16



รูปที่ 16 กระบวนการใช้เทคนิคประเภทคัดเลือกคุณลักษณะเข้าร่วมในเทคนิคแบบสองมิติแบบสองทิศทาง

เริ่มแรกจะใช้เทคนิคแบบสองมิติสองทิศทางกับชุดข้อมูลเพื่อให้ได้เมตริกซ์การฉาย U และ V มาก่อน จากนั้นจึงจับคู่ฐานหลักเพื่อทำการลดมิติของข้อมูลให้มี 1 มิติ แล้วจึงใช้เทคนิคประเภทคัดเลือกคุณลักษณะวัดอำนาจจำแนกโดยถือว่าคุณลักษณะที่เกิดขึ้นเป็นผลงานร่วมของฐานหลักคู่ที่ทำการลดมิติ จากนั้นจะกรองข้อมูลที่มีอำนาจจำแนกต่ำให้มีคะแนนเป็น 0 โดยวัดจากค่ามัชฌิมของ

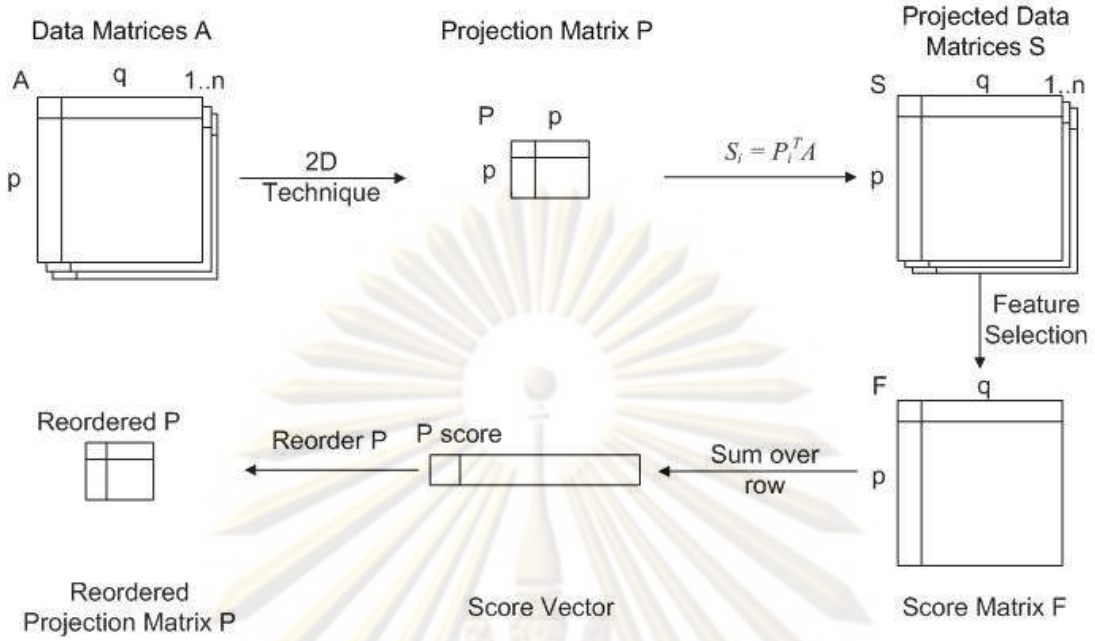
คะแนนทั้งหมดตามจำนวนครั้งในการทำให้เป็นบรรทัดฐาน (Normalize) ที่ต้องการ ทำให้ได้คะแนนที่แบ่งระหว่างคุณลักษณะที่มีอำนาจจำแนกน้อยและมีอำนาจจำแนกมากออกจากกันอย่างชัดเจน จากนั้นจึงรวมคะแนนสำหรับแต่ละฐานหลักที่ใช้ลดมิติตามแนวสดมภ์และแนวแถวตามแนวของมิติที่ทำการลดมิติ เพื่อใช้เป็นคะแนนในการจัดเรียงลำดับเพื่อใช้ในการลดมิติต่อไป โดยขั้นตอนวิธีแสดงในตารางที่ 1

ตารางที่ 1 ขั้นตอนวิธีการเลือกฐานหลักในเทคนิคแบบสองมิติแบบสองทิศทาง

ข้อมูลนำเข้า: เมตริกซ์ข้อมูล A, จำนวนครั้งของการทำให้เป็นบรรทัดฐาน z
ข้อมูลส่งออก: เมตริกซ์การฉายที่จัดเรียงแล้ว U และ V
1. ใช้เทคนิคแบบสองมิติแบบสองทิศทางกับเมตริกซ์ข้อมูล A ทำให้ได้เมตริกซ์การฉาย U และ V
2. นำเวกเตอร์แต่ละสดมภ์ของเมตริกซ์การฉาย U และ V ทำการลดมิติเมตริกซ์ข้อมูล A ให้เหลือ 1×1 มิติตามสมการ (3.57) จนได้เมตริกซ์ข้อมูลที่ถูกฉาย S ที่ S_{ij} แทนข้อมูลที่ถูกลดมิติจากสดมภ์ i และ j ของเมตริกซ์การฉาย U และ V ตามลำดับ
3. ใช้เทคนิคคัดเลือกคุณลักษณะเพื่อทำการจัดลำดับข้อมูลในเมตริกซ์ข้อมูลที่ถูกฉาย S จนได้เมตริกซ์คะแนน F
4. ทำเมตริกซ์คะแนน F ให้เป็นบรรทัดฐานแล้วตัดรายการที่มีค่าน้อยกว่า 0 ให้มีค่าเป็น 0
5. ทำขั้นตอนที่ 4 ซ้ำตามจำนวนครั้งที่ต้องการ z
6. รวบรวมคะแนนตามแนวแถวและแนวสดมภ์เป็นเวกเตอร์คะแนน *Uscore* และ *Vscore* ตามลำดับ
7. จัดเรียงเมตริกซ์การฉาย U และ V ตามคะแนนที่มีนัยสำคัญจากมากไปน้อยตามเวกเตอร์คะแนน *Uscore* และ *Vscore*

4.4 การปรับปรุงเทคนิคแบบสองมิติ

ส่วนเทคนิคแบบสองมิติจะวัดด้วยวิธีที่คล้ายคลึงกันซึ่งสำหรับเทคนิคประเภทสองมิตินั้นจะมีฐานหลักเพียง 1 ชุดคือเมตริกซ์การฉาย P เท่าจำนวนคุณลักษณะตามแนวแถว ดังนั้นจึงใช้เทคนิคประเภทคัดเลือกคุณลักษณะเพื่อคัดกรองคุณลักษณะที่มีอำนาจจำแนกน้อยๆในแต่ละกลุ่มออกไปก่อน ซึ่งแนวคิดสามารถแสดงเป็นภาพได้ดังรูปที่ 17



รูปที่ 17 กระบวนการใช้เทคนิคประเภทคัดเลือกคุณลักษณะเข้าร่วมในเทคนิคแบบสองมิติ

โดยเริ่มจากใช้เทคนิคแบบสองมิติกับชุดข้อมูลเพื่อให้ได้เมตริกซ์การฉาย P มาก่อน จากนั้นจึงใช้แต่ละฐานหลักเพื่อทำการลดมิติของข้อมูล แล้วจึงใช้เทคนิคประเภทคัดเลือกคุณลักษณะวัดอำนาจจำแนกโดยจะกรองข้อมูลที่มีอำนาจจำแนกต่ำให้มีคะแนนเป็น 0 โดยวัดจากค่ามัชฌิมของคะแนนทั้งหมด จากนั้นจึงรวมคะแนนสำหรับแต่ละฐานหลักที่เกี่ยวข้องเพื่อใช้เป็นคะแนนในการจัดเรียงลำดับเพื่อใช้ในการลดมิติต่อไป โดยขั้นตอนวิธีแสดงในตารางที่ 2

ตารางที่ 2 ขั้นตอนวิธีการเลือกฐานหลักในเทคนิคแบบสองมิติ

ข้อมูลนำเข้า: เมตริกซ์ข้อมูล A, จำนวนครั้งของการทำให้เป็นบรรทัดฐาน z
ข้อมูลส่งออก: เมตริกซ์การฉายที่จัดเรียงแล้ว P
1. ใช้เทคนิคแบบสองมิติกับเมตริกซ์ข้อมูล A ทำให้ได้เมตริกซ์การฉาย P
2. นำเวกเตอร์แต่ละสดมภ์ของเมตริกซ์การฉาย P ทำการลดมิติเมตริกซ์ข้อมูล A ให้เหลือ $1\times q$ มิติตามสมการ (3.52) จนได้เมตริกซ์ข้อมูลที่ถูกฉาย S ที่ S_i แทนข้อมูลที่ถูกลดมิติจากสดมภ์ i ของเมตริกซ์การฉาย P
3. ใช้เทคนิคคัดเลือกคุณลักษณะเพื่อทำการจัดลำดับข้อมูลในเมตริกซ์ข้อมูลที่ถูกฉาย S จนได้เมตริกซ์คะแนน F
4. ทำเมตริกซ์คะแนน F ให้เป็นบรรทัดฐานแล้วตัดรายการที่มีค่าน้อยกว่า 0 ให้มีค่าเป็น 0
5. ทำขั้นตอนที่ 4 ซ้ำตามจำนวนครั้งที่ต้องการ z

6. รวบรวมคะแนนตามแนวแถวเป็นเวกเตอร์คะแนน *Pscore*

7. จัดเรียงเมตริกซ์การฉาย *P* ตามคะแนนที่มีนัยสำคัญจากมากไปน้อยตามเวกเตอร์คะแนน *Pscore*

4.5 การเลือกฐานหลักและประมาณมิติเป้าหมาย

ในการประมาณมิติเป้าหมายสำหรับการลดมิตินี้ ได้มีแนวคิด 2 แนวคิดในการเลือกฐานหลักและประมาณมิติเป้าหมาย

1.ใช้เทคนิคคัดเลือกคุณลักษณะประเภทผสม (Hybrid) โดยจะเป็นการใช้เทคนิคประเภทตัวหุ้ม (Wrapper) ร่วมกับคะแนนที่วัดได้ในกระบวนการคัดเลือกคุณลักษณะประเภทตัวกรอง ซึ่งจะใช้คะแนนช่วยจัดเรียงฐานหลักให้เป็นแนวทางในการเลือกชุดของฐานหลักตามอำนาจจำแนกจากมากไปน้อยร่วมกับเทคนิคประเภทตัวหุ้มตามแบบของ John และ Kohavi [59] ที่ใช้การตรวจสอบไขว้ 10 พับ (10-fold Cross Validation) แบ่งเซตฝึกที่เกิดจากการแบ่งชุดฝึกแล้วออกเป็นข้อมูล 10 ส่วนย่อยที่เป็นเซตไม่มีส่วนร่วม (Disjoint Set) จากนั้นจึงใช้ข้อมูล 9 ส่วนย่อยของชุดฝึกในการฝึกแล้วใช้ส่วนย่อยที่เหลือเป็นเซตตรวจสอบ (Validation Set) แล้วตรวจสอบวนไปจนครบทุกส่วนย่อยแล้วนำค่าความแม่น (Accuracy) มาเฉลี่ย โดยใช้การค้นหาแบบสมบูรณ์ (Complete Search) ในการหามิติเป้าหมายโดยอาจเริ่มจากฐานหลักที่มีคะแนนมากที่สุดก่อนจึงค่อยๆเพิ่มฐานหลักที่มีคะแนนรองลงมาจนครบจำนวนมิติแล้ววัดความแม่นจึงค่อยเปลี่ยนไปยังส่วนย่อยถัดไป นอกจากนี้ยังใช้การตัดเล็ม (Prunning) ในการลดปริมาณคำนวณเมื่อชุดฐานหลักใดๆไม่มีความเป็นไปได้ที่จะให้ความแม่นมากกว่าชุดฐานหลักที่ดีที่สุดในขณะนั้น กล่าวคือถึงแม้ว่าในส่วนย่อยอื่นจะได้ความแม่นสูงสุดก็ไม่มากกว่าความแม่นสะสมของชุดฐานหลักที่ดีที่สุดในช่วงเวลาการคำนวณนั้นๆ โดยขั้นตอนวิธีแสดงในตารางที่ 3

วิธีการค้นหาแบบสมบูรณ์นี้มีความเหมาะสมเนื่องจากปริภูมิค่าความแม่นของมิติที่ใกล้เคียงกันมีความแกว่งกวัดมาก (Oscillate) ปริภูมิค่าความแม่นมีลักษณะเหวลึกในบริเวณจำนวนมิติน้อยๆและมีพื้นผิวที่ขรุขระแสดงถึงค่าเหมาะสมที่สุดเฉพาะที่ (Local Optimum) จำนวนมากดังรูปที่ 18 จึงเป็นการยากที่จะใช้อัลกอริทึมเชิงละโมบ (Greedy Algorithm) เช่นการเลือกไปข้างหน้า (Forward Selection) หรือการกำจัดย้อนหลัง (Backward Elimination) ที่จะหยุดเพิ่มหรือลดฐานหลักเมื่อความแม่นในเซตตรวจสอบไม่เพิ่มขึ้น ทำให้มักจะติดในค่าเหมาะสมที่สุดเฉพาะที่ ซึ่งมักเป็นค่าที่ไม่น่าพอใจนัก ดังนั้นหากมีเวลาในการคำนวณมากการใช้เทคนิคประเภทตรวจสอบไขว้ในการประมาณมิติเป้าหมายน่าจะเป็นวิธีการที่เหมาะสม

ตารางที่ 3 ขั้นตอนวิธีในการประมาณมิติเป้าหมายด้วยเทคนิคคัดเลือกคุณลักษณะประเภทผสม

ข้อมูลนำเข้า: เมตริกซ์ข้อมูล A

ข้อมูลส่งออก: เมตริกซ์การฉายที่จัดเรียงและประมาณมิติเป้าหมายแล้ว U และ V

1. แบ่งเมตริกซ์ข้อมูล A เป็น 10 พับ

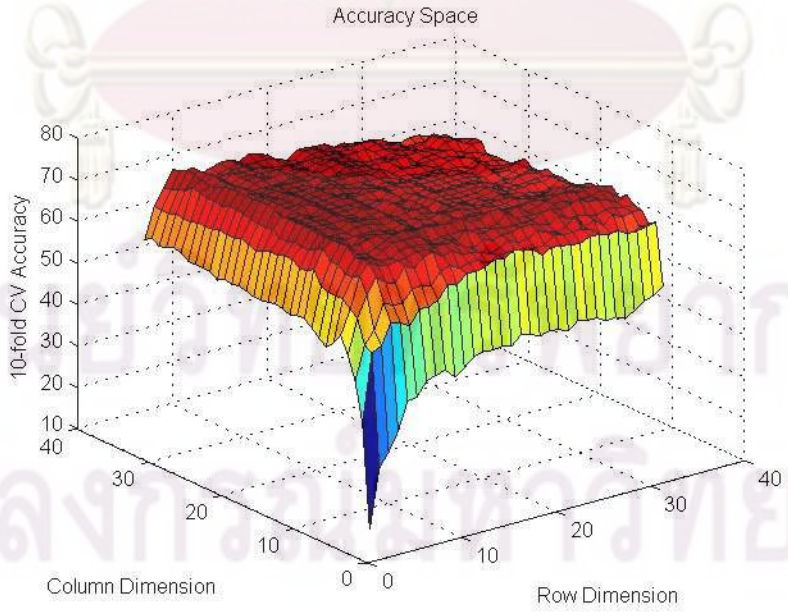
2 ใช้ขั้นตอนวิธีการเลือกฐานหลักในเทคนิคแบบสองมิติแบบสองทิศทางตามตารางที่ 1 กับเมตริกซ์ข้อมูล A โดยจะได้เมตริกซ์การฉายที่จัดเรียงแล้ว U และ V ออกมา

3. สร้างเมตริกซ์ข้อมูล AA จากพับ 9 พับ และเก็บพับสุดท้ายเป็นเซตตรวจสอบ

4. ลดมิติเมตริกซ์ข้อมูล AA ด้วยสมการ (3.57) โดยเริ่มจาก1×1 มิติจนถึงp×q มิติเท่ากับความละเอียดของข้อมูลรูปภาพโดยเลือกตามสดมภ์แรกๆของเมตริกซ์การฉาย U และ V แล้วใช้ฝึกตัวจำแนก Cls ที่ต้องการเพื่อทดสอบด้วยเซตตรวจสอบที่เตรียมไว้ได้ค่าความแม่นเก็บในเมตริกซ์ค่าความแม่นสะสม S ขนาด p×q

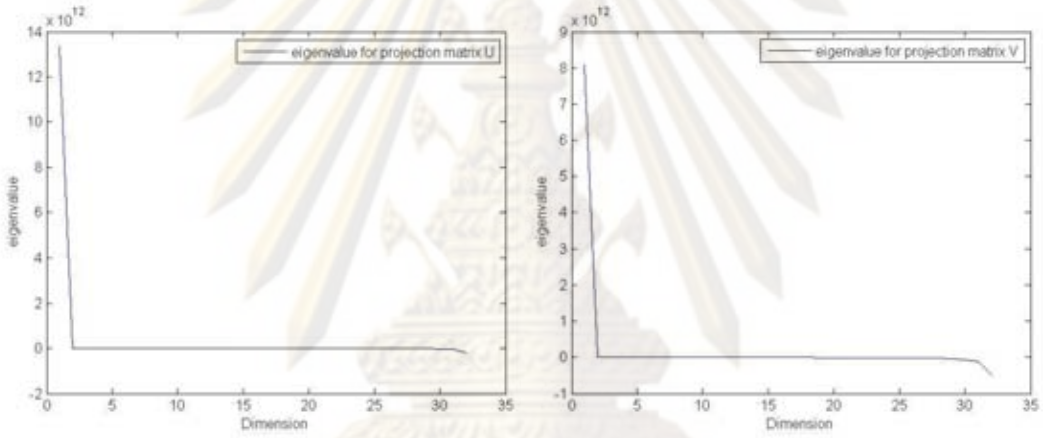
5. ทำซ้ำขั้นตอนที่ 4 จนครบ 10 พับ ระหว่างทำซ้ำให้ข้ามมิติที่ค่าความแม่นสะสมในเมตริกซ์ S ไม่สามารถมากเกินไปกว่าค่าที่มากที่สุดในขณะนั้นแม้ว่าจะได้ค่าความแม่นสูงสุดในพับที่เหลือ

6. ดัชนีของเมตริกซ์ S ที่มีค่าความแม่นสะสมสูงสุดคือมิติเป้าหมายที่ประมาณได้ใช้เลือกสดมภ์แรกๆของเมตริกซ์การฉาย U และ V เท่ากับค่าของเลขดัชนี



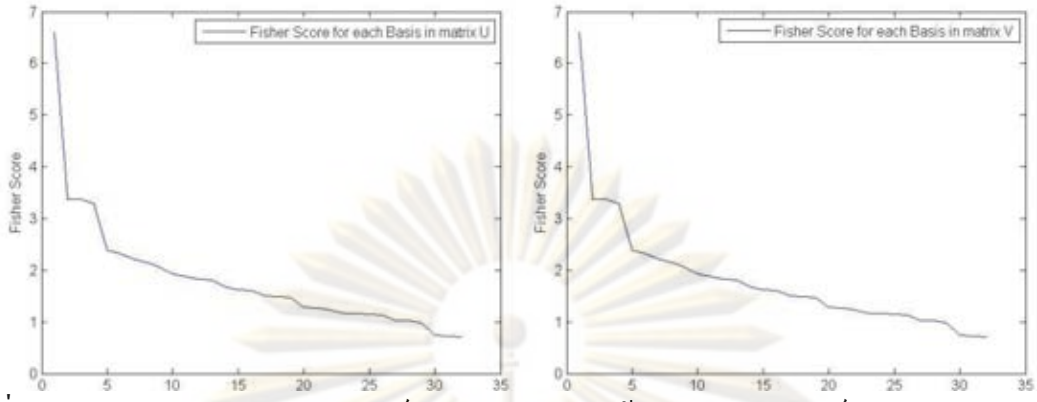
รูปที่ 18 ปริภูมิค่าความแม่นมีความกวัดแกว่งมาก

2. ใช้อัลกอริทึมเชิงละโมบ (Greedy Algorithm) ในการตัดออก (Cutoff) โดยเลือกฐานหลักตามคะแนนส่วนมาก ซึ่งใช้แนวคิดของการเลือกไปข้างหน้า (Forward Selection) และการทดสอบลานหินพัง (Scree Test) ที่จะคัดแต่ฐานหลักที่มีค่าคะแนนมากๆเท่านั้นไปใช้เข้าร่วม ตามทั่วไปแล้วการทดสอบลานหินพังจะใช้ในการวิเคราะห์สเปกตรัมของค่าลักษณะเฉพาะของเทคนิคแบบหนึ่งมิติที่จะเลือกมิติเป้าหมายที่จุดแรกที่ค่าคะแนนลดลงโดยมีความชันมากๆ ซึ่งสามารถวิเคราะห์ได้ด้วยสายตาหรือใช้วิธีทำให้เป็นบรรทัดฐาน (Normalize) แล้วเลือกจุดที่ค่าคะแนนเป็นบวกหรือสะสมผลรวมของคะแนนแล้วใช้วิธีตัดออกเมื่อได้ผลรวมคะแนนสะสมมากกว่าค่าที่กำหนด โดยทั่วไปมักกำหนดให้เป็น 99% โดยสามารถเพิ่มให้เป็น 99.9995% หรือค่าอื่นๆในทำนองนี้ได้ เนื่องจากโดยปกติค่าคะแนนที่มีนัยสำคัญส่วนมากจะมีค่ามากกว่าค่าคะแนนที่ไม่มีนัยสำคัญมากในระดับพันหรือหมื่นเท่าดังรูปที่ 19



รูปที่ 19 ตัวอย่างค่าลักษณะเฉพาะของเมตริกซ์การฉายแนวแถว (ซ้าย) และแนวสดมภ์ (ขวา) จากเทคนิคจากเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง

แต่เทคนิคสองมิติสองทิศทางมีค่าลักษณะเฉพาะที่ทำให้วิเคราะห์ได้ว่าควรลดมิติลงมามากเกินไป ทำให้ติดในหุบของปริภูมิค่าความแม่นในรูปที่ 18 ซึ่งยังคงได้ค่าความแม่นที่ต่ำมากยังไม่ได้ขึ้นมาพบกับบริเวณที่ราบที่มีความกวัดแกว่งมากแต่ค่าความแม่นยังสูงกว่า เมื่อใช้เทคนิคแบบสองมิติสองทิศทางที่นำเสนอพบว่าค่าคะแนนที่ใช้แทนค่าลักษณะเฉพาะจะเปลี่ยนไปโดยทำให้เห็นลานหินพังเป็นชั้นๆหลายๆชั้นเพิ่มขึ้นมาจากค่าลักษณะเฉพาะดังรูปที่ 20 ทำให้ได้ข้อมูลแนวทางที่ทำให้เข้าใจข้อมูลมากกว่าค่าลักษณะเฉพาะ ว่ามีอีกหลายชั้นที่เมื่อลดมิติแล้วจะเป็นไปตามค่าคะแนนจากฟังก์ชันจุดประสงค์มากขึ้นเป็นระดับๆไป

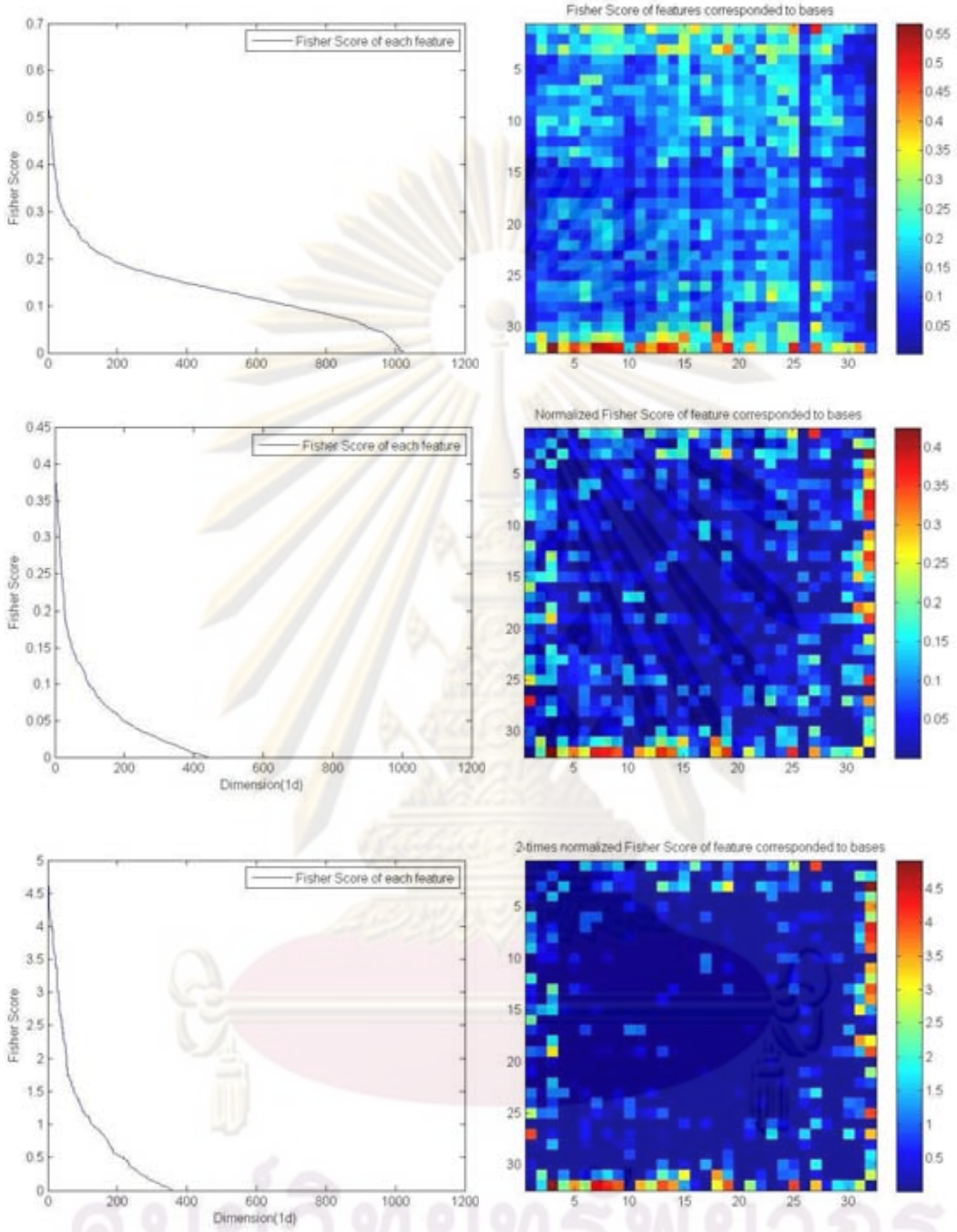


รูปที่ 20 ตัวอย่างค่าคะแนนของเมตริกซ์การฉายแนวแถว (ซ้าย) และแนวสดมภ์ (ขวา) จากเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทางที่นำเสนอ

การเป็นระดับๆไปนี้เราสามารถทำการตัดออกที่แต่ละระดับได้ด้วยการทำให้เป็นบรรทัดฐาน โดยในการทำให้เป็นบรรทัดฐานแต่ละครั้งเราจะสามารถแบ่งแยกคุณลักษณะที่แต่ละฐานหลักมีผลมากและคุณลักษณะที่ฐานหลักมีผลน้อยตามฟังก์ชันจุดประสงค์ได้ตามคะแนนที่ยังคงเหลืออยู่และคะแนนที่ต่ำกว่าค่ามัชฌิมที่ถูกตัดออก ทำให้เป็นประโยชน์ต่อการเลือกฐานหลักที่มีผลมากให้ได้มากๆโดยให้ติดฐานหลักที่มีผลน้อยน้อยที่สุดดังรูปที่ 21

ซึ่งจะเห็นว่าทางสดมภ์ซ้ายจะแสดงค่าคะแนนสำหรับแต่ละคุณลักษณะพบว่าการทำให้เป็นบรรทัดฐานจะตัดตรงส่วนที่มีความชันมากๆและการทำให้เป็นบรรทัดฐานซ้ำจะตัดคุณลักษณะออกน้อยลงแสดงถึงการใกล้มิติเป้าหมาย ส่วนทางสดมภ์ขวาจะเห็นว่ามีคุณลักษณะที่มีคะแนนค่อนไปทางน้อยจำนวนมากโดยสังเกตได้จากสีค่อนไปทางฟ้า เป้าหมายคือการเลือกคุณลักษณะที่มีสีค่อนไปทางแดงให้มากที่สุดและหยุดเมื่อเลือกคุณลักษณะสีค่อนไปทางแดงได้จนเกือบครบแล้ว ทำให้เมื่อทำให้เป็นบรรทัดฐานจะช่วยในการหยุดเลือกได้ง่ายขึ้นเพราะจะตัดคุณลักษณะสีค่อนไปทางฟ้าให้เป็นสีน้ำเงินที่มีค่า 0 ทั้งหมด ทำให้สามารถใช้การตัดออกด้วยคะแนนสะสมได้ง่าย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 21 คะแนนของฟิชเชอร์ที่เปลี่ยนไปในกระบวนการทำให้เป็นบรรทัดฐานได้แก่ คะแนนดั้งเดิม (แถวแรก) หลังทำให้เป็นบรรทัดฐาน 1 ครั้ง (แถวที่สอง) หลังทำให้เป็นบรรทัดฐาน 2 ครั้ง (แถวที่สาม) ของคะแนนจากเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทางที่นำเสนอ

กระบวนการในการเลือกฐานหลักสำหรับแนวคิดนี้เริ่มจากดำเนินกระบวนการของเทคนิคแบบสองมิติสองทิศทางที่นำเสนอจนเสร็จสิ้น จากนั้นจะได้ฐานหลักเป็นเมตริกซ์การฉาย 2 ชุดและ

ค่าคะแนนที่ได้จากเทคนิคคัดเลือกคุณลักษณะประเภทตัวกรองที่มีขนาดเท่าจำนวนคุณลักษณะ จะเริ่มจากการจัดเรียงค่าคะแนนตามอำนาจจำแนกจากมากไปน้อยก่อนกำหนดให้ค่าคะแนนตามอำนาจจำแนกที่มากหมายถึงค่าคะแนนที่ดี จากนั้นจึงเลือกคะแนนที่ดีที่สุดออกมาสะสมไว้พร้อมกับเลือกคู่ฐานหลักที่สร้างคุณลักษณะที่มีคะแนนที่ดีนั้น ต่อมาในการเลือกคะแนนที่ดีที่สุดรองลงมาจะต้องเลือกฐานหลักต่อไปที่มีคะแนนดีที่สุดเมื่อใช้ร่วมกับฐานหลักที่ได้เลือกมาก่อนแล้ว ดังนั้นในการเลือกฐานหลักครั้งต่อๆไปจะเป็นการเลือกฐานหลักที่ดีที่สุดที่เพิ่มคะแนนให้กับชุดฐานหลักที่เลือกมากที่สุดเสมอ จากนั้นจะหยุดกระบวนการเมื่อคะแนนสะสมเกินค่าคงที่ในการตัดออกแล้วจึงได้ชุดฐานหลักเป็นผลลัพธ์และจำนวนฐานหลักเป็นมิติเป้าหมายที่จะลดมิติลงมานั่นเอง โดยขั้นตอนวิธีแสดงในตารางที่ 4

ตารางที่ 4 ขั้นตอนวิธีในการประมาณมิติเป้าหมายด้วยอัลกอริทึมเชิงละโมบและการตัดออก

ข้อมูลนำเข้า: เมตริกซ์ข้อมูล A, ผลรวมสะสมของคะแนน c

ข้อมูลส่งออก: เมตริกซ์การฉายที่จัดเรียงและประมาณมิติเป้าหมายแล้ว Ur และ Vr

1. ใช้ขั้นตอนวิธีการเลือกฐานหลักในเทคนิคแบบสองมิติแบบสองทิศทางตามตารางที่ 1 กับเมตริกซ์ข้อมูล A โดยจะดึงข้อมูลส่งออกเป็นเมตริกซ์การฉายที่ยังไม่ได้จัดเรียง U และ V และเมตริกซ์คะแนน F

2. กำหนดเมตริกซ์ Ur และ Vr เป็นเซตว่างเพื่อใช้แทนเซตของเวกเตอร์ฐานหลักของเมตริกซ์การฉาย U และ V

3. เลือกข้อมูลจากเมตริกซ์ F ที่มีค่ามากที่สุดแล้วดึงเวกเตอร์จากสดมภ์ของเมตริกซ์ U และ V ตามค่าของเลขดัชนีเข้าเพิ่มในเซต Ur และ Vr

4. เลือกเวกเตอร์จากสดมภ์ของเมตริกซ์ U ที่ค่าในเมตริกซ์ F เมื่อค้นหาตามดัชนีเลขสดมภ์ร่วมกับดัชนีจากเซต Vr มีค่าสูงสุด

5. เลือกเวกเตอร์จากสดมภ์ของเมตริกซ์ V ที่ค่าในเมตริกซ์ F เมื่อค้นหาตามดัชนีเลขสดมภ์ร่วมกับดัชนีจากเซต Ur มีค่าสูงสุด

6. เลือกเวกเตอร์จากขั้นตอนที่ 4 หรือ 5 ที่ค่าคะแนนสูงกว่าเพิ่มในเซต Ur หรือ Vr โดยเพิ่มในสดมภ์ถัดไป

7. ทำซ้ำขั้นตอนที่ 4-6 จนกว่าผลรวมของคะแนนในเซต Ur และ Vr คิดเป็นร้อยละเทียบกับผลรวมคะแนนทั้งหมดในเมตริกซ์ F เกินผลรวมสะสมของคะแนน c

บทที่ 5

ผลการทดลอง

การวิจัยได้ดำเนินงานบนโปรแกรม MATLAB และใช้ชุดข้อมูลใบหน้าที่ใช้กันเป็นมาตรฐานได้แก่ ฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (Yale Face Database) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (Extended Yale Face Database B) นอกจากนี้ยังทดสอบเพิ่มเติมกับฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นไอเอสที (MNIST Handwritten Digit Database) ในการทดสอบประสิทธิภาพของเทคนิคและสมมติฐานต่างๆจะใช้วิธีการตรวจสอบไขว้ 10 พับ (10-Fold Cross Validation) ในการทดสอบเปรียบเทียบความแม่นในการจำแนกเป็นสำคัญ

5.1 ชุดข้อมูล

ฐานข้อมูลใบหน้าที่จะใช้ทดสอบแนวคิดประกอบด้วยฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (Yale Face Database) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (Extended Yale Face Database B) โดยฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (Yale Face Database)[1] จะประกอบด้วยข้อมูลใบหน้า 165 ใบหน้าของบุคคล 15 คน ซึ่งจะมีความท้าทายคือใบหน้าจะมีการใส่แว่นและไม่ใส่แว่นเพื่อให้เกิดการบดบังข้อมูลใบหน้าจำนวน 2 รูป แสงเงาเพื่อให้เกิดเงาบดบังใบหน้าโดยแสงจะฉายทางด้านซ้าย ขวา และตรงกลางใบหน้ารวม 3 รูป และที่เหลือ 6 รูปจะเป็นความแตกต่างของสีหน้าตามอารมณ์เช่น ดีใจ ง่วง ตกใจ ขยิบตา เป็นต้น ข้อมูลที่ใช้ในการทดลองจะทำการแปลงความละเอียดของรูปภาพให้มีขนาด 32×32 พิกเซลก่อน

รูปที่ 22 ตัวอย่างใบหน้าจากฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล

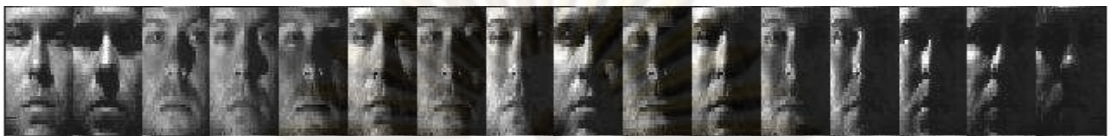
ส่วนฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (Extended Yale Face Database B)[2][60, 61] จะประกอบด้วยใบหน้า 16,128 รูปของบุคคล 28 คน ภายใต้เงื่อนไขท่าทาง 9 ท่าทาง

---

1 http://cvc.yale.edu/projects/yalefaces/yalefaces.html

2 http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html

และแสงเงา 64 แบบ รูปในฐานข้อมูลนี้จะเน้นความท้าทายด้านแสงเงาโดยจะมีภาพที่แสงหรือเงามากจนทำให้ใบหน้าหายไปบางส่วน ข้อมูลที่ใช้ในการทดลองจะเลือกเซตย่อยที่ใบหน้ามีท่าทางค่อนข้างตรงจำนวน 2,414 รูปซึ่งมีจำนวนบุคคลเพิ่มขึ้นเป็น 38 คนแหละทำการแปลงความละเอียดของรูปภาพให้มีขนาด 32×32 พิกเซลก่อน



รูปที่ 23 ตัวอย่างใบหน้าจากฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล

นอกจากนี้ยังทดสอบเพิ่มเติมกับฐานข้อมูลลายมือประกอบด้วยฐานข้อมูลลายมือไบนา-รีอัลฟา (Binary Alphadigits) และฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นไอเอสที (MNIST Handwritten Digit Database) โดยฐานข้อมูลลายมือไบนารีอัลฟา (Binary Alphadigits)[3] จะประกอบด้วยข้อมูลลายมือเขียนตัวเลขอารบิก 0-9 และตัวอักษรภาษาอังกฤษ A-Z ทำให้มีทั้งหมด 36 ประเภทและแต่ละประเภทจะมีข้อมูลตัวอย่างอยู่ 39 ตัวอย่างรวมแล้วมีข้อมูลทั้งหมด 1,404 รูป โดยรูปภาพจะมีความละเอียด 20×16 พิกเซล

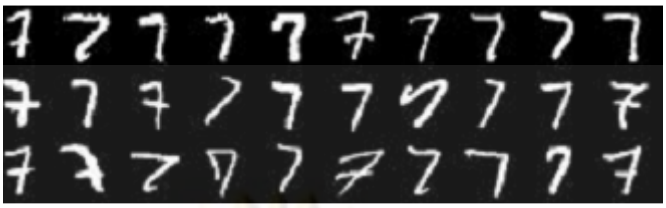
รูปที่ 24 ตัวอย่างลายมือจากฐานข้อมูลไบนารีอัลฟา

ฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นไอเอสที (MNIST Handwritten Digit Database)[4] ประกอบด้วยข้อมูลชุดฝึกเป็นรูปลายมือตัวเลข 60,000 รูปและมีข้อมูลชุดทดสอบเป็นรูปลายมือตัวเลขอีก 10,000 รูปเป็นตัวเลขลายมือเลขอารบิก 0-9 ที่มีความแตกต่างในวิธีการเขียนมาก โดยรูปภาพจะมีความละเอียด 28×28 พิกเซล ทั้งนี้ในการทดลองจะทำการสุ่มเลือกเซตย่อยของข้อมูลจากการนำชุดฝึกและชุดทดสอบมาคละโดยเลือกให้ข้อมูลในแต่ละประเภทมีขนาดใกล้เคียงกันจนได้ชุดข้อมูลใหม่ที่มีข้อมูลอยู่ 5,348 รูป

---

3 http://www.cs.toronto.edu/~roweis/data.html

4 http://yann.lecun.com/exdb/mnist/



รูปที่ 25 ตัวอย่างลายมือตัวเลขจากฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นเอสไอที

5.2 การทดลองและผลการทดลอง

จะทดสอบแนวคิดกับเทคนิคการลดมิติ 4 เทคนิคได้แก่ การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นหรือแอลดีเอ การฉายรักษาสมบัติเฉพาะที่หรือแอลพีพี ตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์หรือดีเอ็นอี และการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่หรือแอลเอสดีเอ โดยจะทดสอบกับเทคนิคแบบสองมิติสองทิศทาง และใช้เทคนิคคัดเลือกคุณลักษณะประเภทตัวกรองได้แก่ คะแนนของฟิชเชอร์เพื่อใช้กับการวิเคราะห์ดิสคริมิแนนต์เชิงเส้น คะแนนลาปลาเชียนเพื่อใช้กับการฉายรักษาสมบัติเฉพาะที่ สำหรับตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์และการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่นั้นจะทำการประดิษฐ์เทคนิคคัดเลือกคุณลักษณะขึ้นมาในทำนองเดียวกันกับคะแนนลาปลาเชียนทำให้ได้เทคนิคคะแนนดีเอ็นอีและคะแนนแอลเอสดีเอตามลำดับ

การทดลองจะแบ่งเป็นการทดสอบค่าความแม่นที่เพิ่มขึ้นโดยยังไม่กำหนดมิติเป้าหมายและการทดสอบค่าความแม่นสำหรับการใช้งานจริงเมื่อกำหนดมิติเป้าหมาย ทั้งสองการทดสอบจะใช้การตรวจสอบไขว้ 10 พับในการเปรียบเทียบความแม่น โดยจะเน้นย้ำนัยสำคัญของค่าความแม่นที่เพิ่มขึ้นด้วยการทดสอบทีชนิดคู่แบบด้านเดียว (One-Tailed Paired t-Test) นอกจากนี้จะใช้การกระจายไบแอส-แวเรียนซ์ในการศึกษาข้อเท็จจริงที่เกิดขึ้นของแนวคิดในฐานข้อมูลใบหน้าอีกด้วย

การทดสอบค่าความแม่นที่เพิ่มขึ้นโดยยังไม่กำหนดมิติเป้าหมายจะทำการฝึกด้วยเทคนิคการลดมิติแบบดั้งเดิมและเทคนิคการลดมิติแบบประยุกต์แนวคิดที่นำเสนอเพื่อหาตัวแบบคือเมตริกซ์การฉายออกมาจากนั้นจึงวัดค่าความแม่นในการจำแนกด้วยอัลกอริทึมเพื่อนบ้านใกล้สุด 1 ตัวเนื่องจากเป็นตัวจำแนกที่อ้างอิงตามระยะทางที่สอดคล้องกับแนวคิดการลดมิติที่จะหาการส่ง (Mapping) จากมิติที่สูงกว่ามายังมิติที่ต่ำกว่าโดยรักษาระยะทางไว้มากที่สุด ทำให้ได้ผลลัพธ์ออกมาเป็นปริภูมิค่าความแม่นตามจำนวนมิติที่เกิดจากฐานหลักแล้วหาค่าที่ดีที่สุดของแต่ละพับมาเฉลี่ยเพื่อแสดงถึงความเป็นไปได้ที่ความแม่นจะเพิ่มขึ้นเมื่อประยุกต์ใช้แนวคิดที่นำเสนอ การทดลองได้ผลลัพธ์ดังตารางที่ 5

ตารางที่ 5 ผลลัพธ์ค่าความแม่นแบบไม่กำหนดมิติเป้าหมายของเทคนิคแบบสองมิติสองทิศทาง

เทคนิค	ฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล	ฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล	ฐานข้อมูลลายมือไบนารีอัลฟา	ฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นเอสไอที
$(2D)^2$LDA	84.33%	90.07%	70.44%	92.74%
$(2D)^2$LDA+FS*	**86.33%**	**95.49%+**	71.67%	95.01%+++
$(2D)^2$LPP	82.33%	88.57%	75.53%	91.75%
$(2D)^2$LPP+LS*	82.67%	93.93%+	**75.93%**	94.87%+++
$(2D)^2$DNE	83.33%	88.95%	75.46%	94.79%
$(2D)^2$DNE+DNES*	85.33%	95.22%++	75.46%	94.87%
$(2D)^2$LSDA	81.67%	89.04%	67.71%	94.95%
$(2D)^2$LSDA+LSDAS*	85.67%	95.22%++	75.79%+++	**95.03%**

* หมายถึงเทคนิคที่นำเสนอ

+ หมายถึงเทคนิคที่นำเสนอมีประสิทธิภาพสูงกว่าเทคนิคดั้งเดิมที่ช่วงความเชื่อมั่น 95%

++ หมายถึงเทคนิคที่นำเสนอมีประสิทธิภาพสูงกว่าเทคนิคดั้งเดิมที่ช่วงความเชื่อมั่น 99%

+++ หมายถึงเทคนิคที่นำเสนอมีประสิทธิภาพสูงกว่าเทคนิคดั้งเดิมที่ช่วงความเชื่อมั่น 99.99%

จากผลการทดลองในตารางที่ 5 จะเห็นว่าการเลือกฐานหลักใหม่ด้วยเทคนิคคัดเลือกคุณลักษณะได้ช่วยเพิ่มค่าความแม่นในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล 2.00% ในการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) 0.33% ในการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) 2.00% ในตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติสองทิศทาง ($(2D)^2$DNE) และ 4.00% ในการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LSDA) โดยเทคนิคลดมิติที่ให้ค่าความแม่นสูงสุดคือเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทางที่ 86.33% จะเห็นว่าในฐานข้อมูลนี้ความแม่นเพิ่มขึ้นไม่มากเพราะจำนวนข้อมูลมีน้อยการลดค่าความผิดพลาดจากพจน์ไบแอสไม่ส่งผลในการเพิ่มความแม่นมากนัก

ส่วนในฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยลค่าความแม่นได้เพิ่มสูงขึ้นมากโดยเพิ่มถึง 5.42% ในการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) 5.37% ในการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) 6.27% ในตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติสองทิศทาง ($(2D)^2$DNE) และ 6.18% ในการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LSDA) โดยเทคนิคลดมิติที่ให้ค่าความแม่นสูงสุดคือเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทางที่ 95.49% ในฐานข้อมูลนี้จำนวนข้อมูลมีสูงขึ้นการลดค่าความผิดพลาดจากพจน์ไบแอสให้ผลในการเพิ่มค่าความแม่นอย่างเห็นได้ชัด

ในฐานข้อมูลลายมือไบนารีอัลฟาได้ค่าความแม่นสูงขึ้นในการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) เพิ่มขึ้น 1.23% และในการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) เพิ่มขึ้น 0.40% แต่ในตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติสองทิศทาง ($(2D)^2$DNE) ค่าความแม่นยังคงเท่าเดิมเนื่องจากไม่พบปัญหาการจัดเรียงเวกเตอร์ผิดลำดับอำนาจจำแนกเพราะตัวเทคนิคได้ออกแบบไปในทางรักษาสมบัติเฉพาะที่มากกว่าสมบัติโดยรวม และในการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LSDA) ค่าความแม่นเพิ่มขึ้นมากถึง 8.08% โดยเทคนิคลดมิติที่ให้ค่าความแม่นสูงสุดคือเทคนิคการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทางที่ 75.93% ในฐานข้อมูลลายมือนี้จะพบว่าความท้าทายได้เปลี่ยนไปโดยเปลี่ยนจากอารมณ์ ท่าทาง และแสงเงาของใบหน้าเป็นความแตกต่างในการเขียนซึ่งประกอบด้วยการเอนเอียง เส้นหนาบาง และส่วนที่เพิ่มขึ้นและหายไปของตัวอักษรและตัวเลขตามบุคคล ทำให้ข้อมูลมีการกระจายตัวสูงกว่าข้อมูลใบหน้าจะสังเกตได้ว่าสมบัติเฉพาะที่มีความสำคัญมากกว่าสมบัติโดยรวมจากเทคนิคที่ดีที่สุดที่เปลี่ยนจากการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นเป็นการฉายรักษาสมบัติเฉพาะที่นั่นเอง

ส่วนในฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นเอสไอทีค่าความแม่นได้เพิ่มสูงขึ้นมากในการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) และในการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) โดยเพิ่มขึ้นถึง 2.37% และ 3.12% ตามลำดับ แต่ในตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติสองทิศทาง ($(2D)^2$DNE) และในการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LSDA) ค่าความแม่นเพิ่มสูงขึ้นไม่มากโดยเพิ่มขึ้นเพียง 0.08% เท่านั้น โดยเทคนิคลดมิติที่ให้ค่าความแม่นสูงสุดคือเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติสองทิศทางที่ 95.03% เนื่องจากเมื่อข้อมูลเพิ่มขึ้นการรักษาระยะขอบเฉพาะที่จะทำได้แม่นยำยิ่งขึ้นและเป็นการโน้มเอียงไปทางสมบัติโดยรวมตามจำนวนข้อมูลที่มีอยู่มาก

ในฐานข้อมูลนี้จำนวนข้อมูลมีสูงมากขึ้นการลดค่าความผิดพลาดจากพจน์ไบแอสและค่าในพจน์แวเรียนซ์ที่เพิ่มขึ้นไม่ได้ช่วยเพิ่มค่าความแม่นให้กับเทคนิคที่ซับซ้อนอย่างตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติสองทิศทาง ($(2D)^2$DNE) และการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LSDA) มากนัก สันนิษฐานว่ามีจำนวนข้อมูลมากเพียงพอและความท้าทายในฐานข้อมูลได้ถูกเอาชนะด้วยตัวเทคนิคเองอยู่แล้ว ในขณะที่เทคนิคพื้นฐานอย่างการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) และการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) การใช้เทคนิคคัดเลือกคุณลักษณะร่วมด้วยยังช่วยในการเอาชนะความท้าทายในฐานข้อมูลอยู่

ตารางที่ 6 ผลลัพธ์ค่าความแม่นแบบไม่กำหนดมิติเป้าหมายของเทคนิคแบบสองมิติ

เทคนิค	ฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล	ฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล	ฐานข้อมูลลายมือไบนารีอัลฟา	ฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นเอสไอที
2DLDA	82.33%	90.40%	69.07%	92.32%
2DLDA+FS*	**83.00%**	91.95%	69.17%	94.60%++
2DLPP	78.33%	91.34%	72.62%	94.05%
2DLPP+LS*	78.67%	93.24%	72.69%	94.64%
2DDNE	79.33%	91.52%	72.18%	94.14%
2DDNE+DNES*	80.67%	93.75%	72.43%	94.38%
2DLSDA	78.33%	91.93%	69.84%	**94.66%**
2DLSDA+LSDAS*	82.00%	**94.48%**	**72.96%+**	**94.66%**

* หมายถึงเทคนิคที่นำเสนอ

+ หมายถึงเทคนิคที่นำเสนอมีประสิทธิภาพสูงกว่าเทคนิคดั้งเดิมที่ช่วงความเชื่อมั่น 95%

++ หมายถึงเทคนิคที่นำเสนอมีประสิทธิภาพสูงกว่าเทคนิคดั้งเดิมที่ช่วงความเชื่อมั่น 99%

เมื่อทดสอบแนวคิดเพิ่มเติมกับเทคนิคแบบสองมิติจะได้ผลการทดลองในตารางที่ 6 จะเห็นว่าการเลือกฐานหลักใหม่ด้วยเทคนิคคัดเลือกคุณลักษณะได้ช่วยเพิ่มค่าความแม่นในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล 0.67% ในการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติ (2DLDA)

0.33% ในการฉายรักษาสมบัติเฉพาะที่แบบสองมิติ (2DLPP) 1.33% ในตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติ (2DDNE) และ 3.67% ในการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติ (2DLSDA) โดยเทคนิคลดมิติที่ให้ค่าความแม่นสูงสุดคือเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติที่ 83.00% จะเห็นว่าในฐานข้อมูลนี้เทคนิคแบบสองมิติดั้งเดิมไม่ได้มีค่าความแม่นดีกว่าเทคนิคแบบสองมิติสองทิศทางดั้งเดิม แต่การจัดเรียงฐานหลักใหม่ก็เพิ่มความแม่นให้เล็กน้อยเช่นกันแต่ไม่เพิ่มมากเท่ากับในเทคนิคแบบสองมิติสองทิศทาง

ส่วนในฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยลค่าความแม่นได้เพิ่มสูงขึ้นมากกว่าโดยเพิ่มถึง 1.55% ในการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติ (2DLDA) 1.90% ในการฉายรักษาสมบัติเฉพาะที่แบบสองมิติ (2DLPP) 2.23% ในตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติ (2DDNE) และ 2.55% ในการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติ (2DLSDA) โดยเทคนิคลดมิติที่ให้ค่าความแม่นสูงสุดคือเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติที่ 94.48% ในฐานข้อมูลนี้เทคนิคแบบสองมิติดั้งเดิมมีค่าความแม่นสูงกว่าเทคนิคแบบสองมิติสองทิศทางดั้งเดิม แต่เมื่อจัดเรียงฐานหลักใหม่พบว่าแม้ค่าความแม่นจะสูงขึ้นมากแต่ไม่สูงขึ้นมากเท่ากับเทคนิคแบบสองมิติสองทิศทาง

ส่วนในฐานข้อมูลลายมือไบนารีอัลฟาค่าความแม่นได้เพิ่มสูงขึ้น 0.10% ในการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติ (2DLDA) 0.07% ในการฉายรักษาสมบัติเฉพาะที่แบบสองมิติ (2DLPP) 0.25% ในตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติ (2DDNE) และ 3.12% ในการฉายรักษาเฉพาะที่แบบสองมิติ (2DLPP) โดยเทคนิคลดมิติที่ให้ค่าความแม่นสูงสุดคือเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติที่ 72.69% ในฐานข้อมูลนี้เทคนิคแบบสองมิติดั้งเดิมมีค่าความแม่นต่ำกว่าเทคนิคแบบสองมิติสองทิศทางดั้งเดิม แต่เมื่อจัดเรียงฐานหลักใหม่พบว่าจะให้ค่าความแม่นสูงขึ้นเล็กน้อยเท่านั้น

ส่วนในฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นเอสไอทีค่าความแม่นได้เพิ่มสูงขึ้นโดยเพิ่มขึ้น 2.32% ในการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติ (2DLDA) 0.59% ในการฉายรักษาสมบัติเฉพาะที่แบบสองมิติ (2DLPP) 0.24% ในตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติ (2DDNE) และ ไม่เพิ่มเลยในการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติ (2DLSDA) โดยเทคนิคลดมิติที่ให้ค่าความแม่นสูงสุดคือเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติที่ 94.66% ในฐานข้อมูลนี้เทคนิคการฉายรักษาสมบัติเฉพาะที่และการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติดั้งเดิมมีค่าความแม่นสูงกว่าเทคนิคแบบสองมิติสองทิศทางดั้งเดิม แต่เมื่อจัดเรียงฐานหลักใหม่พบว่าแม้ค่าความแม่นจะสูงขึ้นมากแต่ไม่สูงขึ้นมากเท่ากับเทคนิคแบบสองมิติสองทิศทาง ส่วนการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นและตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบ

สองมิติแบบดั้งเดิมค่าความแม่นต่ำกว่าเทคนิคแบบสองมิติสองทิศทางดั้งเดิม แต่เมื่อจัดเรียงฐานหลักใหม่พบว่าแม้ค่าความแม่นจะสูงขึ้นมากแต่ไม่สูงขึ้นมากเท่ากับเทคนิคแบบสองมิติสองทิศทางเช่นกัน

ตารางที่ 7 ผลลัพธ์ค่าความแม่นเมื่อกำหนดมิติเป้าหมายของเทคนิคแบบสองมิติสองทิศทาง

เทคนิค	ฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล	ฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล	ฐานข้อมูลลายมือไบนารีอัลฟา	ฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นเอสไอที
$(2D)^2$LDA+wrapper	75.00%	-	-	-
$(2D)^2$LDA+FS+wrapper*	77.67%	-	-	-
$(2D)^2$LDA+FS+cutoff*	**79.00%**	86.45%	70.00%	93.92%
$(2D)^2$LPP+wrapper	71.67%	-	-	-
$(2D)^2$LPP+LS+wrapper	72.33%	-	-	-
$(2D)^2$LPP+LS+cutoff*	70.67%	**92.65%**	**73.66%**	**94.66%**
$(2D)^2$DNE+wrapper	77.33%	-	-	-
$(2D)^2$DNE+DNES+wrapper*	78.67%	-	-	-
$(2D)^2$DNE+DNES+cutoff*	75.33%	87.76%	72.06%	92.00%
$(2D)^2$LSDA+wrapper	73.00%	-	-	-
$(2D)^2$LSDA+LSDAS+wrapper*	74.00%	-	-	-
$(2D)^2$LSDA+LSDAS+cutoff*	75.33%	92.54%	71.30%	92.48%

* หมายถึงเทคนิคที่นำเสนอ
- หมายถึงไม่สะดวกในการคำนวณ

การทดสอบค่าความแม่นสำหรับการใช้งานจริงเมื่อกำหนดมิติเป้าหมายจะทำการเลือกฐานหลักเปรียบเทียบความแม่นในการจำแนกด้วยอัลกอริทึมเพื่อนบ้านใกล้สุด 1 ตัวระหว่างเทคนิค

แบบดั้งเดิมโดยจะเลือกให้มีความผิดพลาดแบบทั่วไป (Generalized Error) ต่ำสุดด้วยวิธีตัวหุ้มที่ใช้การตรวจสอบไขว้ และแนวคิดในการเลือกฐานหลักใหม่ที่นำเสนอทั้งแบบใช้เทคนิคการคัดเลือกคุณลักษณะแบบผสมและแบบใช้อัลกอริทึมเชิงละโมบ โดยในทุกๆพับจะประมาณค่ามิติเป้าหมายมาค่าหนึ่งและใช้ในการลดมิติในพับนั้นๆและจะประมาณใหม่เมื่อเปลี่ยนพับ การทดลองได้ผลลัพธ์ดังตารางที่ 7

จากผลการทดลองในตารางที่ 7 พบว่าในการทดสอบค่าความแม่นเมื่อกำหนดมิติเป้าหมายในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยลพบว่าจะได้ค่าความแม่นที่เพิ่มขึ้นเมื่อใช้เทคนิคที่นำเสนอมากกว่าเทคนิคต้นฉบับเมื่อใช้การตรวจสอบไขว้ในการประมาณมิติเป้าหมาย โดยในเทคนิคประเภทดิสคริมิแนนต์ได้แก่การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) และการวิเคราะห์ดิสคริมิแนนต์เฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LSDA) แนวคิดอัลกอริทึมเชิงละโมบในการตัดออกให้ผลดีที่สุด ส่วนเทคนิคประเภทเน้นสมบัติเฉพาะที่และข้อมูลเพื่อนบ้านอย่างการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) และตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติสองทิศทาง ($(2D)^2$DNE) การใช้เทคนิคคัดเลือกคุณลักษณะแบบผสมจะได้ผลลัพธ์ที่ดีที่สุด ทั้งนี้ค่าความแม่นที่สูงที่สุดได้จากการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) และการตัดออกที่ 79.00%

ส่วนในฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล ฐานข้อมูลลายมือไบนารีอัลฟาและฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นเอสไอทีพบว่าการใช้เทคนิคคัดเลือกคุณลักษณะแบบผสมไม่สามารถทำงานได้เนื่องจากปริมาณการทำงานขึ้นกับจำนวนข้อมูล ทำให้ต้องการเวลามากเกินไปไม่เหมาะสมต่อการใช้งานจริง ในขณะที่แนวคิดอัลกอริทึมเชิงละโมบในการตัดออกสามารถทำงานได้อย่างรวดเร็วเพราะขึ้นกับจำนวนคุณลักษณะไม่ขึ้นกับจำนวนข้อมูล ทั้งนี้ค่าความแม่นที่สูงที่สุดได้จากการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) และการตัดออกที่ 92.65% 73.66% และ 94.66% ตามลำดับ ถึงแม้จะมีความแม่นเมื่อไม่กำหนดมิติเป้าหมายจะต่ำในบางฐานข้อมูลก็ตาม เนื่องจากการฉายรักษาสมบัติเฉพาะที่มีสมบัติจากทฤษฎีกราฟเชิงเสป็กตรัมที่สามารถตรวจจับจำนวนกลุ่มในการรวมกลุ่มได้ดีด้วยการตัดออก ทำให้สามารถนำมาใช้ในการตัดออกกับแนวคิดที่นำเสนอได้เป็นอย่างดี

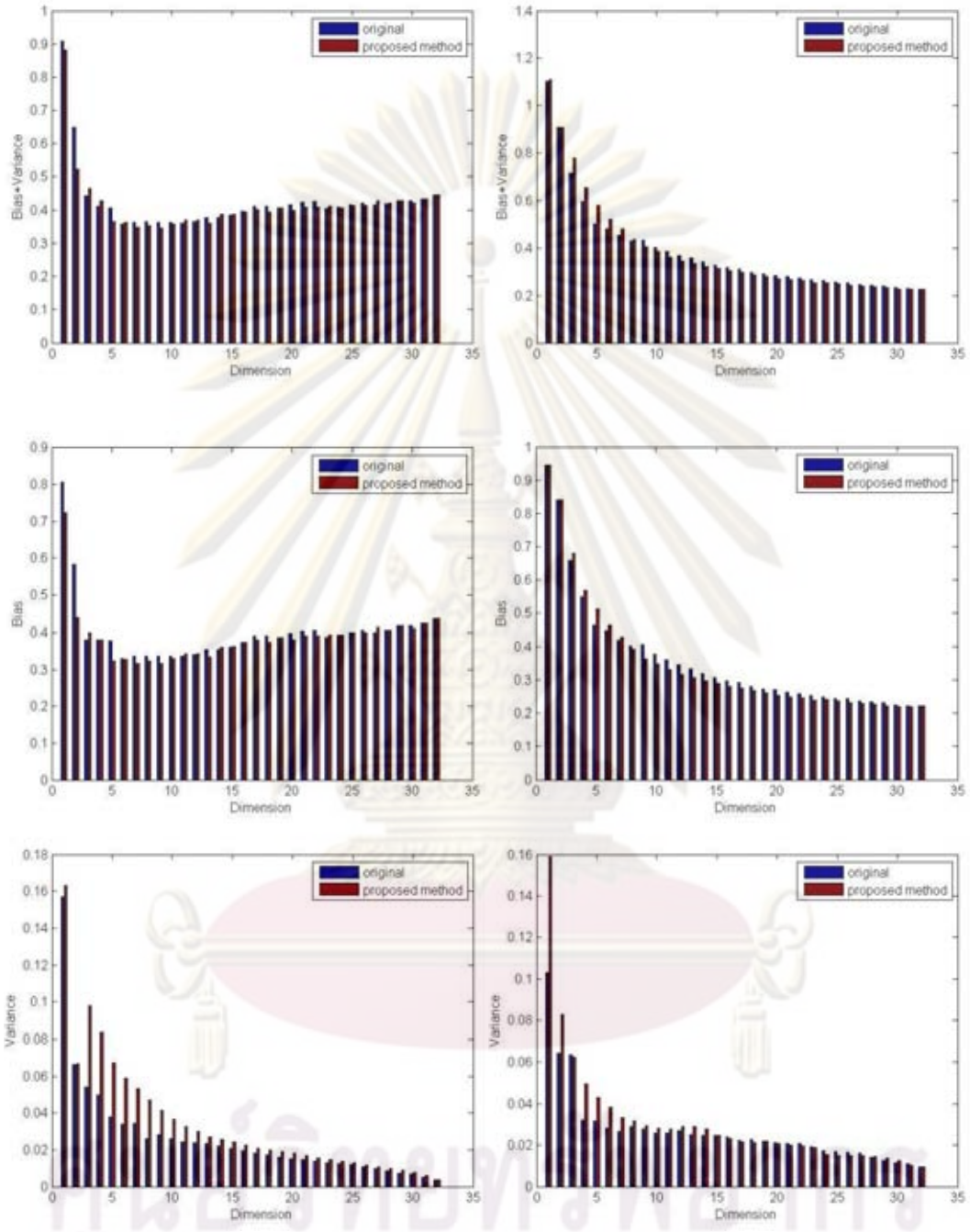
นอกจากนี้จะใช้การกระจายไบแอส-แวเรียนซ์ในการศึกษาความเปลี่ยนแปลงของค่าความผิดพลาดแบบทั่วไปในกรอบของไบแอส-แวเรียนซ์สำหรับวิเคราะห์ค่าความแม่นที่เปลี่ยนไปตามมิติเป้าหมายแบบ $n \times n$ เมื่อ n มีค่าตั้งแต่ 1 ถึง 32 ที่ประมาณด้วยวิธีของ Bauer และ Kohavi ได้ผลลัพธ์ดังรูปที่ 25 ถึง 28 ทั้งนี้พจน์ไบแอสที่ประมาณหมายถึงผลรวมของพจน์ไบแอสกำลังสองในสมการ (3.93) และสัญญาณรบกวน เนื่องจากเมื่อใช้ข้อมูลประเภทแทนสมมติฐานตัวแบบที่เหมาะ

ที่สุดที่ไม่สามารถหาได้ในทางทฤษฎีและปฏิบัติทำให้ไม่สามารถแยกพจน์ไบแอสและสัญญาณรบกวนได้

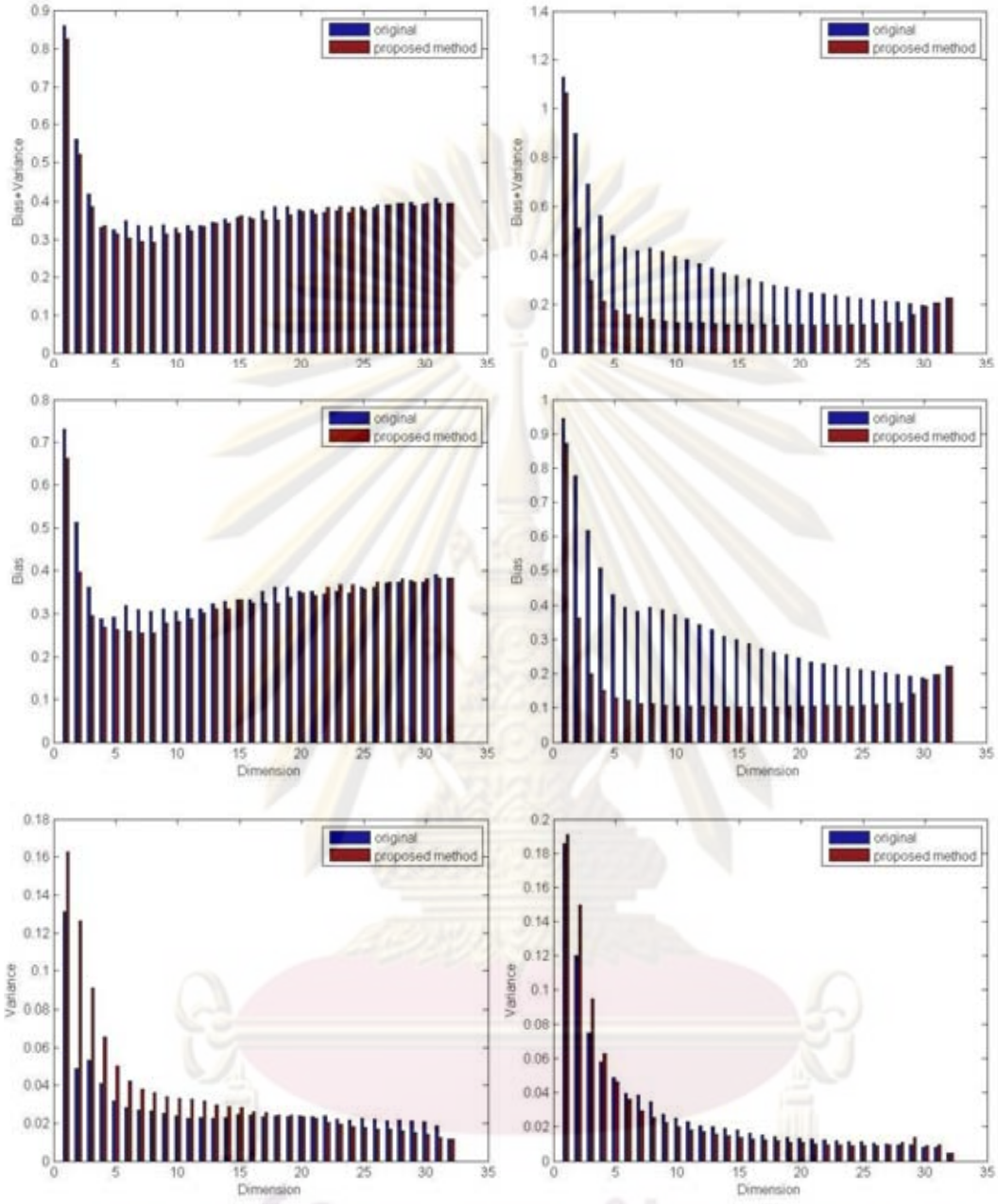
จะเห็นว่าในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยลค่าความผิดพลาดจะน้อยสุดเมื่อลดมิติลงมามากสังเกตได้จากเส้นกราฟที่มีลักษณะเป็นอุ้งลงมาแสดงว่าการลดมิติสามารถช่วยเพิ่มค่าความแม่นได้อย่างดีจากความเชื่อที่ว่าการลดมิติที่ดีที่สุดคือจุดที่พจน์ไบแอสและแวเรียนซ์สมดุลกันที่สุด ส่วนในฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยลพบว่าเทคนิคดั้งเดิมมีค่าความผิดพลาดน้อยสุดเมื่อลดมิติมาเพียงเล็กน้อยเท่านั้นแสดงว่าการลดมิติอาจไม่สามารถช่วยลดปริมาณการคำนวณและเพิ่มความแม่นในฐานข้อมูลนี้ได้มากนัก

แต่เมื่อใช้แนวคิดที่นำเสนอพบว่าเมื่อเรียงลำดับฐานหลักใหม่ด้วยเทคนิคคัดเลือกคุณลักษณะทำให้ลดพจน์ไบแอสได้ดีแลกด้วยการเพิ่มพจน์แวเรียนซ์ที่มีค่าน้อยทำให้ค่าความผิดพลาดต่ำลง ทั้งนี้จะเห็นได้ชัดในการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) ตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติสองทิศทาง ($(2D)^2$DNE) และการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LSDA) ที่ช่วยในการลดมิติมากขึ้นสังเกตได้จากเส้นกราฟที่มีลักษณะเป็นอุ้ง แต่การวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) ถึงแม้จะให้ค่าความแม่นที่ดีที่สุดในการตรวจสอบไขว้ 10 พับเมื่อไม่กำหนดมิติเป้าหมายแต่กลับลดมิติได้น้อยทำให้ยังคงสิ้นเปลืองปริมาณการคำนวณอยู่

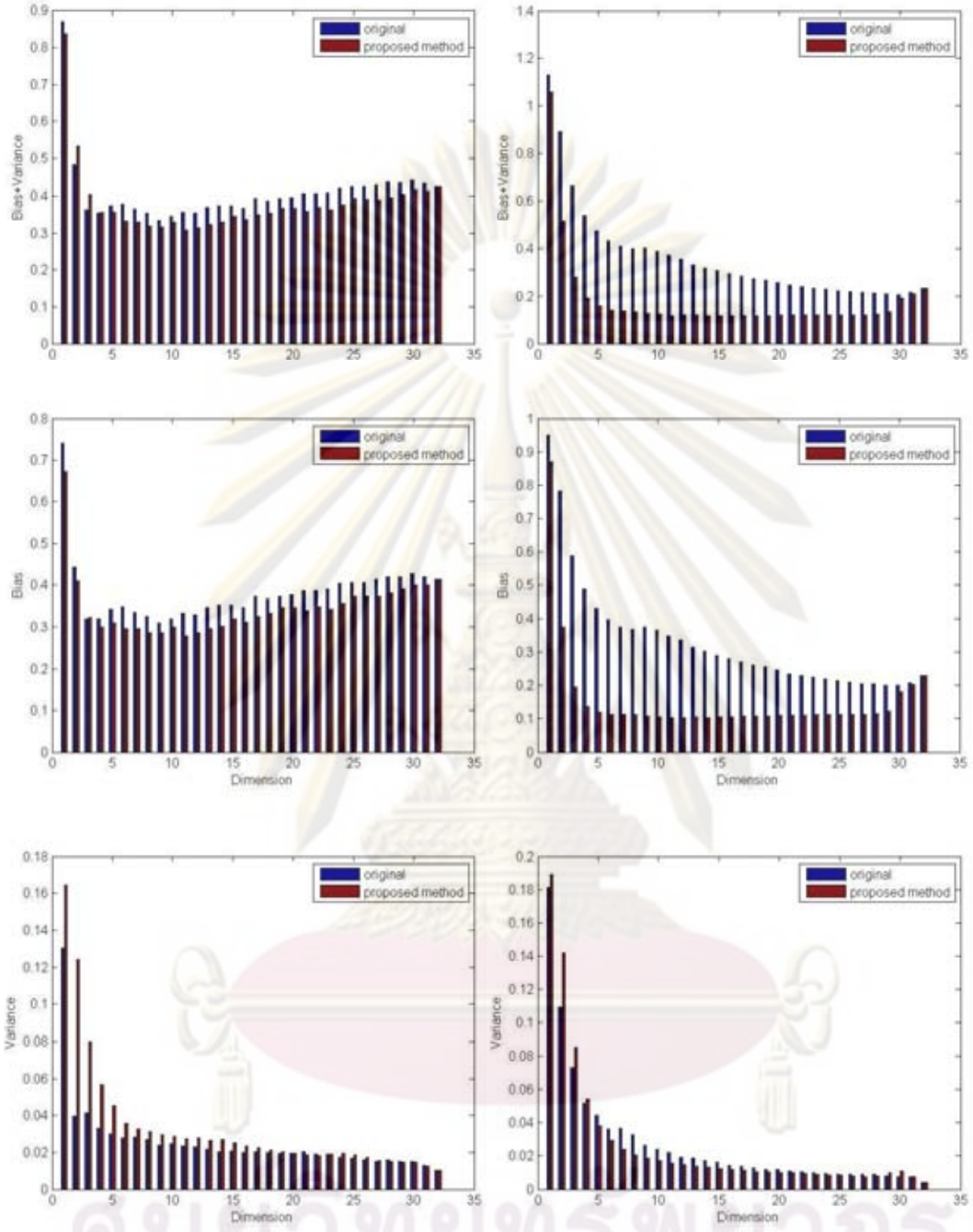
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



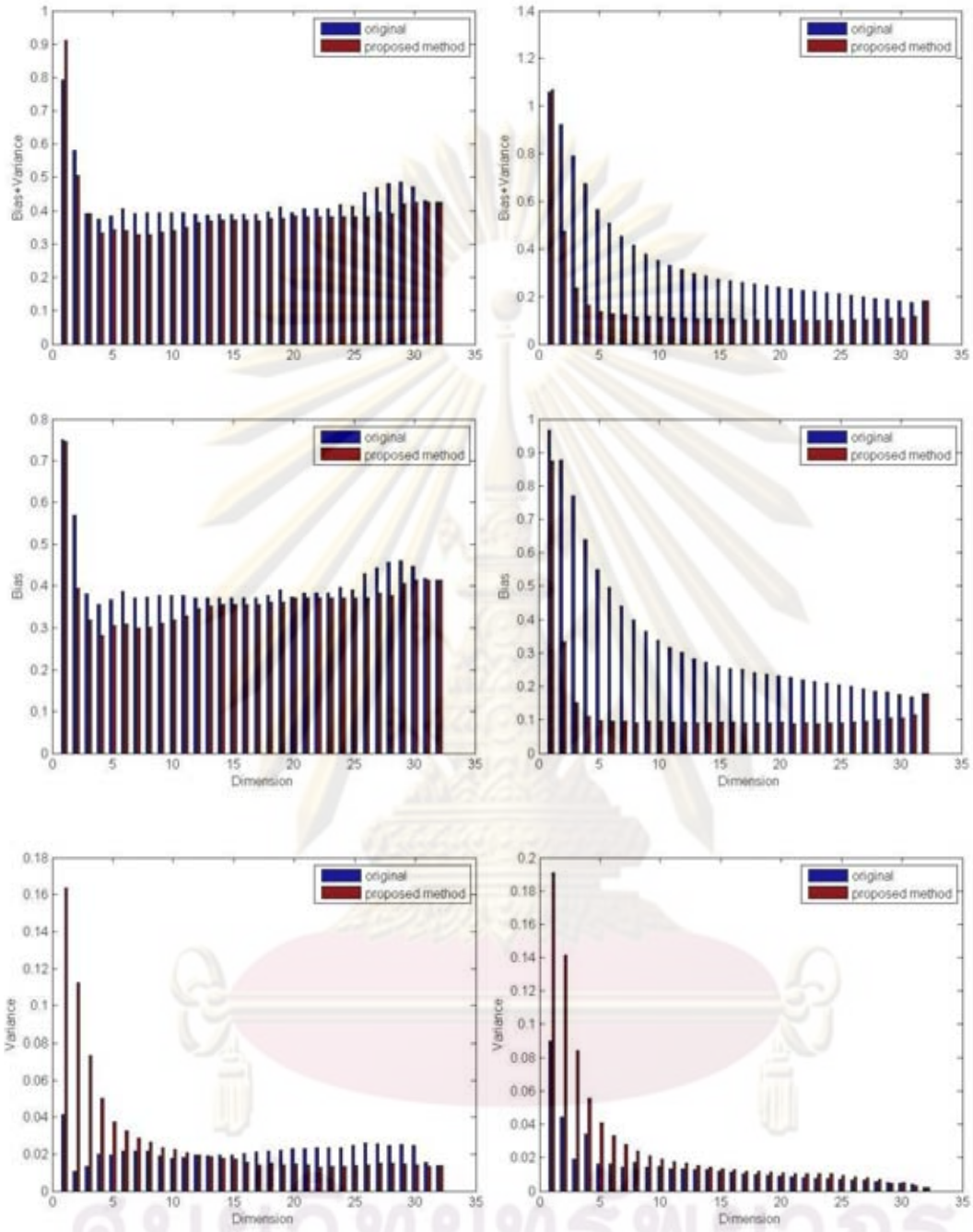
รูปที่ 26 ค่าความผิดพลาดทั่วไป (แถวบน) ไบแอส (แถวกลาง) แวเรียนซ์ (แถวล่าง) ที่ประมาณด้วยวิธีของ Bauer และ Kohavi สำหรับเทคนิคการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นแบบสองมิติสองทิศทาง ($(2D)^2$LDA) ในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (สดมภ์ซ้าย) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (สดมภ์ขวา)



รูปที่ 27 ค่าความผิดพลาดทั่วไป (แถวบน) ไบแอส (แถวกลาง) แวเรียนซ์ (แถวล่าง) ที่ประมาณด้วยวิธีของ Bauer และ Kohavi สำหรับเทคนิคการฉายรักษาสมบัติเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LPP) ในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (สดมภ์ซ้าย) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (สดมภ์ขวา)



รูปที่ 28 ค่าความผิดพลาดทั่วไป (แถวบน) ไบแอส (แถวกลาง) แวเรียนซ์ (แถวล่าง) ที่ประมาณด้วยวิธีของ Bauer และ Kohavi สำหรับเทคนิคตัวฝึงตัวเพื่อนบ้านแบบดิสคริมิแนนต์แบบสองมิติสองทิศทาง ($(2D)^2DNE$) ในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (สดมภ์ซ้าย) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (สดมภ์ขวา)



รูปที่ 29 ค่าความผิดพลาดทั่วไป (แถวบน) ไบแอส (แถวกลาง) แวเรียนซ์ (แถวล่าง) ที่ประมาณด้วยวิธีของ Bauer และ Kohavi สำหรับเทคนิคการวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่แบบสองมิติสองทิศทาง ($(2D)^2$LSDA) ในฐานข้อมูลใบหน้าของมหาวิทยาลัยเยล (สดมภ์ซ้าย) และฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล (สดมภ์ขวา)

บทที่ 6

สรุปผลการวิจัยและข้อเสนอแนะ

6.1 สรุปผลการวิจัย

ในวิทยานิพนธ์นี้ได้กล่าวถึงปัญหาการจัดเรียงเวกเตอร์ลักษณะเฉพาะตามอำนาจจำแนกสำหรับการลดมิติในเทคนิคการลดมิติแบบสองมิติและสองมิติสองทิศทางที่คำนวณโดยมองข้อมูลเป็นบล็อกทำให้การคำนวณหยาบและการหาค่าเหมาะที่สุดไม่เหมาะสม โดยนำเสนอแนวคิดของการเลือกฐานหลักใหม่ด้วยเทคนิคคัดเลือกคุณลักษณะประเภทตัวกรองเพื่อแบ่งคุณลักษณะที่มีอำนาจจำแนกมากและน้อยออกจากกันโดยการทำให้เป็นบรรทัดฐาน ทั้งนี้จำนวนครั้งของการทำให้เป็นบรรทัดฐานสามารถกำหนดด้วยการวิเคราะห์จากแผนภาพหรือจากการทดสอบการกระจายไบแอส-แวเรียนซ์ ซึ่งให้สมมติฐานว่าค่าความแม่นควรจะสูงขึ้น

นอกจากนี้ยังกล่าวถึงปัญหาในการประมาณมิติเป้าหมายในเทคนิคแบบสองมิติสองทิศทางที่มีความไม่เหมาะสมจากการใช้ค่าลักษณะเฉพาะเป็นเกณฑ์และนำเสนอวิธีประมาณมิติเป้าหมายเพื่อการใช้งานจริงให้เป็นไปตามฟังก์ชันจุดประสงค์มากขึ้น 2 วิธีได้แก่ วิธีใช้เทคนิคคัดเลือกคุณลักษณะแบบผสมเพื่อแก้ปัญหาปริภูมิความแม่นที่กวัดแกว่งมากโดยมีความเหมาะสมในฐานข้อมูลที่มีขนาดเล็ก และวิธีใช้อัลกอริทึมเชิงละโมบในการตัดออกเพื่อใช้แทนค่าลักษณะเฉพาะโดยมีความเหมาะสมในฐานข้อมูลที่มีขนาดใหญ่

ผลการทดลองในงานรู้จำใบหน้าได้ทดสอบบนฐานข้อมูลใบหน้าของมหาวิทยาลัยเยลและฐานข้อมูลใบหน้าเพิ่มเติมบีของมหาวิทยาลัยเยล ส่วนผลการทดลองในงานรู้จำลายมือได้ทดสอบบนฐานข้อมูลลายมือไบนารีอัลฟาและฐานข้อมูลลายมือตัวเลขของเอ็มเอ็นเอสไอที และได้ทดสอบแนวคิดกับเทคนิคสกัดคุณลักษณะ 4 เทคนิคได้แก่ การวิเคราะห์ดิสคริมิแนนต์เชิงเส้น การฉายรักษาสมบัติเฉพาะที่ ตัวฝังตัวเพื่อนบ้านแบบดิสคริมิแนนต์ การวิเคราะห์ดิสคริมิแนนต์ไวเฉพาะที่ ผลการทดลองแสดงให้เห็นว่าเทคนิคแบบสองมิตินั้นได้รับผลกระทบจากปัญหาการจัดเรียงเวกเตอร์ลักษณะเฉพาะและเทคนิคแบบสองมิติสองทิศทางจะได้รับผลกระทบจากปัญหานี้มากสังเกตได้จากค่าความแม่นที่สูงขึ้นเมื่อแก้ไขเทคนิคด้วยแนวคิดที่นำเสนอทั้งในการทดลองที่ไม่กำหนดมิติเป้าหมายและการทดลองเมื่อกำหนดมิติเป้าหมายเพื่อใช้งานจริง

โดยเมื่อวิเคราะห์ด้วยการกระจายไบแอส-แวเรียนซ์พบว่าแนวคิดที่นำเสนอช่วยลดพจน์ไบแอสได้มากแลกด้วยการเพิ่มพจน์แวเรียนซ์เล็กน้อยทำให้ค่าความผิดพลาดทั่วไปลดลง มีผลให้ความแม่นสูงขึ้นโดยเฉพาะในฐานข้อมูลขนาดใหญ่ที่ค่าไบแอสที่ลดลงยังคงมีผลอยู่ นอกจากนี้ยังทำให้เห็นจำนวนมิติที่แต่ละเทคนิคสามารถลดเพื่อขจัดความห่างมากในการลดปริมาณการคำนวณได้อีกด้วย

ทั้งนี้ในฐานข้อมูลลายมือจะเห็นถึงความสำคัญและการแลกเปลี่ยนของสมบัติเฉพาะที่และสมบัติโดยรวมเมื่อข้อมูลเพิ่มขึ้น โดยเมื่อข้อมูลเพิ่มขึ้นสมบัติโยรวมจะมีความสำคัญมากขึ้นแต่อย่างไรก็ตามเทคนิคที่ดีที่สุดคือเทคนิคที่มีการแลกเปลี่ยนของสมบัติเฉพาะที่และสมบัติโดยรวมเหมาะสมกับความท้าทายในชุดข้อมูลมากที่สุด

ในการใช้งานจริงพบว่าการฉายรักษาสมบัติเฉพาะที่เมื่อนำมาประยุกต์การตัดออกกับแนวคิดที่นำเสนอพบว่าให้ผลลัพธ์ที่น่าพอใจเนื่องจากยังคงรักษาค่าความแม่นได้มาก แสดงถึงการประมาณมิติเป้าหมายที่แม่นยำกว่าเทคนิคอื่นๆ สำหรับการใช้เทคนิคคัดเลือกคุณลักษณะในการเลือกฐานหลัก

นอกจากนี้ในวิทยานิพนธ์นี้ยังได้สนับสนุนแนวคิดการใช้เทคนิคคัดเลือกคุณลักษณะต่อจากเทคนิคสกัดคุณลักษณะในการประมาณมิติเป้าหมายและหาตัวแบบสำหรับการลดมิติ ทำให้เพิ่มความยืดหยุ่นในการใช้งานจริงและการประดิษฐ์เทคนิคใหม่ๆต่อไป

6.2 ข้อเสนอแนะ

อย่างไรก็ตามในประเด็นของการหาค่าเหมาะที่สุดที่เหมาะสมในเทคนิคแบบสองมิติสองทิศทางและเทคนิคคัดเลือกคุณลักษณะอื่นๆที่สามารถประยุกต์ใช้ในการหาตัวแบบรวมถึงผลลัพธ์ในการเรียนรู้ของเครื่องและงานรู้จำอื่นๆเช่น การจำแนกข้อความ และผลลัพธ์ในตัวจำแนกอื่นๆที่ไม่ใช่อัลกอริทึมเพื่อนบ้านใกล้สุดยังคงเป็นประเด็นที่ต้องมีการศึกษาวิจัยต่อไป

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

[1] H. Zheng and Y. Zhang. Feature selection for high-dimensional data in astronomy. Advances in Space Research. vol. 41. (2008): 1960-1963.

[2] Y. Saeys, I. Inza and P. Larranaga. A review of feature selection techniques in bioinformatics. Bioinformatics. vol. 23. no. 19. (2007): 2507-2517. DOI 10.1093/bioinformatics/btm344.

[3] Y. Chen, Y. Li, X. Cheng and L. Guo. Survey and taxonomy of feature selection algorithms in intrusion detection system. In Proc. Inscrypt. LNCS. (2006): pp. 153-167.

[4] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. Journal of Machine Learning Research. vol. 3. (2003): 1157-1171.

[5] K. Kira and L.A. Rendell. A practical approach to feature selection. In Proc. International Conference on Machine Learning (1992): pp. 249-256.

[6] I. Kononenko. Estimating attributes: Analysis and extensions of RELIEF. In Proc. European Conference on Machine Learning. (1994).

[7] H. Almuallim and T.G. Dietterich. Learning with many irrelevant features. In Proc. National Conference on Artificial Intelligence. MIT Press. (1992): pp. 547-552.

[8] X. He, D. Cai and P. Niyogi. Laplacian score for feature selection. In Proc. Advances in Neural Information Processing Systems. (2005).

[9] D. Koller and M. Sahami. Toward optimal feature selection. In Proc. International Conference on Machine Learning. (1996).

[10] H. Peng, F. Long and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. On Pattern Analysis and Machine Intelligence. vol. 27. no. 8. (2005): 1226-1238.

[11] S. Das. Filters, wrappers and a boosting-based hybrid for feature selection. In Proc. International Conference on Machine Learning. (2001).

[12] H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. IEEE Transaction on Knowledge and Data Engineering. vol. 17. no. 4. (2005): 491-502.

[13] L.v.d. Maaten, E. Postma and J.v.d. Henrik. Dimensionality reduction: A comparative review. Journal of Machine Learning Research. vol. 10. no. 1-14. (2009).

[14] K. Pearson. On lines and planes of closest fit to systems of points in space. Philosophical Magazine. vol. 2. (1901): 559-572.

[15] W.S. Torgerson. Multidimensional scaling I: Theory and method. Psychometrika. vol. 17. (1952): 401-419.

[16] J.B. Tenenbaum, V.d. Silva and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. Science. vol. 290. no. 5500. (2000): 2319-2323.

[17] B. Scholkopf, A.J. Smola and K.R. Muller. Non-linear component analysis as a kernel eigenvalue problem. Neural Computation. vol. 10. no. 5. (1998): 1299-1319.

[18] S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by Locally Linear Embedding. Science. vol. 290. no. 5500. (2000): 2323-2326.

[19] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In Proc. Advances in Neural Information Processing Systems. (2002).

[20] G.E. Hinton and R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. Science. vol. 313. no. 5789. (2006): 504-507.

[21] Y.W. Teh and S.T. Roweis. Automatic alignment of hidden representations. In Proc. Advances in Neural Information Processing Systems. (2002).

[22] R.A. Fisher. The use of multiple measurements in taxonomic problems. Ann of Eugenics. (1936): 179-185.

[23] K. Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press, 1990.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

[24] L. Yang and R. Lin, Distance metric learning: A comprehensive survey, Michigan State University, 2006.

[25] E.P. Xing, A.Y. Ng, M.I. Jordan and S. Russell. Distance metric learning, with application to clustering with side-information. In Proc. Neural Information Processing Systems. (2003).

[26] S. Mika, G. Ratsch, J. Weston, B. Scholkopf and K.R. Mullers. Fisher discriminant analysis with kernels. In Proc. Neural Networks for Signal Processing Workshop. (1999): pp. 41-48.

[27] S. Yan, D. Xu, B. Zhang and H. Zhang. Graph embedding: A general framework for dimensionality reduction. In Proc. IEEE International Conference on Computer Vision and Pattern Recognition. (2005).

[28] W. Zhang, X. Xue, Z. Sun, Y.-F. Guo and H. Lu. Optimal dimensionality of metric space for classification. In Proc. International Conference on Machine Learning. (2007).

[29] J. Goldberger, S. Roweis, G. Hinton and R. Salakhutdinov. Neighbourhood component analysis. In Proc. Advance in Neural Information Processing Systems. (2005).

[30] D. Cai, X. He, K. Zhou, J. Han and H. Bao. Locality Sensitive Discriminant Analysis. In Proc. International Joint Cenferences on Artificial Intelligence. (2007).

[31] X. He and P. Niyogi. Locality preserving projections. In Proc. Advances in Neural Information Processing Systems. (2003).

[32] X. He, D. Cai, S. Yan and H. Zhang. Neighborhood preserving embedding. In Proc. IEEE International Conference on Computer Vision. (2005).

[33] A.N. Gorban. Principal manifolds for data visualization and dimension reduction. Springer, 2008.

[34] C.M. Bishop. Pattern recognition and machine learning. Springer, 2006.

[35] I.T. Jolliffe. Principal component analysis. Springer, 2002.

[36] M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In Proc. IEEE Conf. Computer Vision and Pattern Recognition. (1991): pp. 586-591.

[37] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. vol. 19. no. 7. (1997): 711-720. DOI http://dx.doi.org/10.1109/34.598228.

[38] X. He, S. Yan, Y. Hu, P. Niyogi and H.-J. Zhang. Face recognition using laplacianfaces. IEEE Trans. Pattern Anal. Mach. Intell. vol. 27. no. 3. (2005): 328-340. DOI http://doi.ieeecomputersociety.org/10.1109/TPAMI.2005.55.

[39] X. He, D. Cai and W. Min. Statistical and computational analysis of locality preserving projection. In Proc. International Conference on Machine Learning. (2005).

[40] W. Zhang, X. Xue, H. Lu and Y.F. Gao. Discriminant neighborhood embedding for classification. Pattern Recognition. vol. 39. (2006): 2240-2243.

[41] J. Yang, D. Zhang, A.F. Frangi and J.Y. Yang. Two-dimensional PCA: A new approach to appearance-based face representation and recognition. . IEEE Trans. On Pattern Analysis and Machine Intelligence. vol. 26. no. 1. (2004): 131-137.

[42] D. Zhang and Z.-H. Zhou. (2D)2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition Neurocomputing. vol. 69. (2005): 224-239.

[43] J. Ye. Generalized low rank approximations of matrices. In Proc. International Comference on Machine Learning. (2004).

[44] J. Ye. Generalized low rank approximations of matrices. Machine Learning. vol. 61. (2005): 161-191.

[45] L. Ming and B. Yuan. 2D-LDA: a statistical linear discriminant analysis for image matrix. Pattern Recognition. vol. 26. no. 5. (2005): 527-532.

[46] S. Noushath, G.H. Kumar and P. Shivakumara. (2D)2LDA: An efficient approach for face recognition. Pattern Recognition. vol. 39. (2006): 1396-1410.

[47] J. Ye, R. Jarnardan and Q. Li. Two-dimensional linear discriminant analysis. In Proc. Neural Information Processing Systems. (2004).

[48] S. Chen, H. Zhao, M. Kong and B. Luo. 2D-LPP: A two-dimensional extension of locality preserving projections. Neurocomputing. vol. 70. (2007): 912-921.

[49] X. He, D. Cai and P. Niyogi. Tensor subspace analysis. In Proc. Neural Information Processing Systems. (2006).

[50] W. Zhang, X. Xue, Z. Sun, Y.F. Guo, M. Chi and H. Lu. Efficient feature extraction for image classification. In Proc. International Conference on Computer Vision. (2007).

[51] X. Liu, Z. Wang, J. Liu and Z. Feng. Face recognition with locality sensitive discriminant analysis based on matrix representation. In Proc. International Joint Conference on Neural Networks. (2008).

[52] R.O. Duda, P.E. Hart and D.G. Stork. Pattern classification. Wiley-Interscience, 2000.

[53] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. Machine Learning. vol. 36. no. 1-2. (1999): 105-139.

[54] Z. Wang, S. Chen, J. Liu and D. Zhang. Pattern representation in feature extraction and classifier design: Matrix versus vector. IEEE Trans. On Neural Networks. vol. 19. no. 5. (2008): 758-769.

[55] L. Wang, X. Wang and J. Feng. On image matrix based feature extraction algorithms. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS. vol. 36. no. 1. (2006): 194-197.

[56] Z. Liang, Y. Li and P. Shi. A note on two-dimensional linear discriminant analysis. Pattern Recognition Letters. vol. 29. (2008): 2122-2128.

[57] W.-S. Zheng, J.H. Lai and S.Z. Li. 1D-LDA vs. 2D-LDA: When is vector-based linear discriminant analysis. Pattern Recognition. vol. 41. (2008): 2156-2172.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

[58] N. Nyugen, W. Liu and S. Venkatesh. Random subspace two-dimensional PCA for face recognition. In Proc. Advances in Multimedia Information Processing. Springer. (2007): pp. 655-664.

[59] G.H. John and R. Kohavi. Irrelevant features and the subset selection problem. In Proc. International Conference on Machine Learning. (1994): pp. 121-129.

[60] A. Georghiades, P. Belhumeur and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intell. vol. 23. no. 6. (2001): 643-660.

[61] K.C. Lee, J. Ho and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. vol. 27. no. 5. (2005): 684-698.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นายพีรธรรม วิริยธรรมภูมิสำเร็จการศึกษาวิศวกรรมศาสตรบัณฑิต (เกียรตินิยมอันดับ 2) สาขาวิศวกรรมคอมพิวเตอร์ จากภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2551 (เมษายน 2552) และได้เข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2552 (พฤษภาคม 2552)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย