ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

การจัดกลุ่มข้อมูลโดยเซ็ลฟออร์แกไนซิงแม็ปและการจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล

SELF-ORGANIZING MAP AND PARTITIONING METHOD IN CLUSTER

H006296

โดย

ณัฐพงษ์ ฉัตรแก้วโพธิ์ทอง

NATTAPONG CHATKAEWPOTONG

อาจารย์ที่ปรึกษา

รศ.ดร. อาริต ธรรมโน

QW. ณ389ก 2552

ฉบหมู่.....
เลขทะเบียน 06296
วัน,เดือน,ปี 17 ก.พ. 2554

b.12303033
i.....

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ภาคเรียนที่ 1 ปีการศึกษา 2552

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

SELF-ORGANIZING MAP AND PATTITIONING METHOD IN CLUSTER

NATTAPONG CHATKAEWPOTONG

A REPORT SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS OF THE COURSE

SYSTEM DEVELOPMENT PROJECT

MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY

FACULTY OF INFORMATION TECNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

1/ 2009

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2009

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	การเปรียบเทียบอัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ปสำหรับการจัดกลุ่มข้อมูล
นักศึกษา	นายณัฐพงษ์ ฉัตรแก้วโพธิ์ทอง
รหัสนักศึกษา	49066439
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2552
อาจารย์ที่ปรึกษา	รศ.ดร.อาริต ธรรมโน

บทคัดย่อ

การจัดกลุ่มข้อมูลเป็นกระบวนการหนึ่งในการทำ Data Mining เพื่อจัดกลุ่มของข้อมูลโดยที่เราไม่ทราบกลุ่มของข้อมูลไว้ การเลือกใช้วิธีการที่เหมาะสมจะช่วยให้การจัดกลุ่มมีประสิทธิภาพมากยิ่งขึ้น โครงการพัฒนาระบบงานนี้จึงพัฒนาโปรแกรมขึ้นเพื่อเปรียบเทียบความสามารถในการจัดกลุ่มข้อมูล โดยใช้การจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล (Partitioning Method) ได้แก่ K- means , PAM และ โครงข่ายประสาทเทียมแบบไม่มีผู้สอน SOM(Self-organizing map) โดยศึกษาจากประสิทธิภาพด้านผลลัพธ์ที่ได้ เพื่อให้ได้อัลกอริทึมที่เหมาะสมที่สุดในการจัดกลุ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Self- Organizing Map and Partitioning Method in Cluster
Student	Mr Nattapong Chatkaewpothong
Student ID.	49066439
Degree	Master of Science
Program	Information Technology
Major	Information Science
Academic Year	2009
Advisor	Associate. Professor Arit Thummano, Ph.D

ABSTRACT

Clustering is once process in many process of data mining for classifies data into supervised class. Using a proper system will make us do this method effectively. So, An Application in this project is developed to compare between clustering with Self- organizing map (SOM) and Partitioning Method (such as K-means and PAM) by studying from results of three algorithms to get an algorithm which is appropriate for clustering.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

โครงการศึกษาพัฒนาระบบการเปรียบเทียบอัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ปสำหรับการจัดกลุ่มข้อมูลจะสำเร็จลุล่วงไปไม่ได้เลย ถ้าไม่ได้รับการช่วยเหลือและแรงสนับสนุนจากบุคคลสำคัญหลายท่าน ดังต่อไปนี้

ข้าพเจ้าขอกราบขอบพระคุณ รศ.ดร.อาริต ธรรมโน ซึ่งเป็นอาจารย์ที่ปรึกษาโครงการนี้ ที่ให้ความกรุณาให้คำแนะนำ และปรึกษา ข้าพเจ้ารู้สึกซาบซึ้งและขอขอบพระคุณเป็นอย่างสูง

ขอกราบพระคุณคณาจารย์ของสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุก ๆ ท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้า

ขอขอบคุณ ภญ. นฤมล กาญจนชัยรัตน์ ที่คอยให้กำลังใจเสมอมา

ขอขอบคุณเพื่อน ๆ พี่ ๆ น้อง ๆ ทุกคนที่ให้คำแนะนำ

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้าที่เป็นกำลังใจ และให้การสนับสนุนในทุกเรื่องๆ ทำให้ข้าพเจ้าสามารถทำโครงการพัฒนาระบบงานฉบับนี้สำเร็จลุล่วงด้วยดี

คุณค่าและประโยชน์จากโครงการพัฒนาระบบงานฉบับนี้ ข้าพเจ้าขอมอบแด่ผู้มีพระคุณทุกท่าน

ณัฐพงษ์ ฉัตรแก้วโพธิ์ทอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VII

สารบัญรูป VIII

บทที่ 1 บทนำ 1

1.1 ความเป็นมา 1

1.2 วัตถุประสงค์ 1

1.3 ขอบเขตการดำเนินงาน 2

1.4 ขั้นตอนและวิธีดำเนินงาน 2

1.5 ประโยชน์ที่คาดว่าจะได้รับ 2

1.6 เครื่องมือที่ใช้ในการพัฒนาระบบ 3

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง 4

2.1 การจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล(Partitioning Method) 4

2.1.1 K-means 4

2.1.2 การทำงานของ k-means 5

2.1.3 ตัวอย่างการจัดกลุ่มของ k-means 6

2.1.4 K-medoids 9

2.1.5 กรณีการย้ายกลุ่มของ Medoids แบบต่างๆ 9

2.1.6 การทำงานของPAM 10

2.1.7 ตัวอย่างการจัดกลุ่มของ PAM 11

2.2 ทฤษฎีโครงข่ายประสาทเทียม 15

2.2.1 โครงสร้างของโครงข่ายประสาทเทียม 15

2.2.2 การทำงานของโครงข่ายประสาทเทียม 17

2.2.3 ลักษณะการเรียนรู้ของโครงข่ายประสาทเทียม 18

2.3 อัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ป 18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

2.3.1 ลักษณะของอัลกอริทึมเซ็ลฟ์ออร์แกไนซิงแม็ป 18
2.3.2 โครงสร้างของเซ็ลฟ์ออร์แกไนซิงแม็ป 19
2.3.3 ขั้นตอนการทำงานของเซ็ลฟ์ออร์แกไนซิงแม็ป 19
2.3.4 อัลกอริทึมในการพัฒนาระบบการแบ่งกลุ่มข้อมูล
ด้วยเซ็ลฟ์ออร์แกไนซิงแม็ป 20

บทที่ 3 วิธีการดำเนินงาน 21
3.1 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม 21
3.2 ลักษณะอินพุทและเอาท์พุทของโปรแกรม 21
3.2.1 ลักษณะอินพุทของโปรแกรม 21
3.2. ลักษณะเอาท์พุทของโปรแกรม 22
3.3 ขั้นตอนการทำงานของโปรแกรม 24
3.4 ส่วนติดต่อผู้ใช้ของโปรแกรม 27
3.4.1 หน้าต่าง Menu 27
3.4.2 หน้าต่างหลักของโปรแกรมจัดกลุ่มโดยใช้ K-means 27
3.4.3 หน้าต่างหลักของโปรแกรมจัดกลุ่มโดยใช้ PAM 29
3.4.4 หน้าต่างหลักของโปรแกรมจัดกลุ่มโดยใช้ SOM 31
3.5 การใช้งานโปรแกรมจัดกลุ่มข้อมูล 33
3.5.1 การใช้งานโปรแกรมจัดกลุ่มโดยใช้ k-meansม 33
3.5.2 การใช้งานโปรแกรมจัดกลุ่มโดยใช้ PAM 37
3.5.3 การใช้งานโปรแกรมจัดกลุ่มโดยใช้ SOM 40

บทที่ 4 ผลการทดลอง 44
4.1 ข้อมูลที่ใช้กับโปรแกรม 44
4.1.1 Iris Data 44
4.1.2 Wine Data 45
4.1.3 Diabetes Data 45
4.1.4 Breast Data 46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.2 ทดสอบค่าที่ใช้ในการจัดกลุ่มของ SOM 46
4.2.1 ทดสอบค่าในข้อมูล Iris Data 46
4.2.2 ทดสอบค่าในข้อมูล Diabetes Data 48
4.2.3 ทดสอบค่าในข้อมูล Wine Data 49
4.2.4 ทดสอบค่าในข้อมูล Breast Data 50
4.3 ผลการทดลอง 51
4.3.1 Iris Data 51
4.3.2 Diabetes Data 54
4.3.3 Wine Data 56
4.3.4 Breast Data 59
4.4 วิเคราะห์ผลการทดลอง 61
4.4.1 วิเคราะห์ผลการทดลอง 61

บทที่ 5 สรุปผลการทดลองและข้อเสนอแนะ 63
5.1 สรุปผลการศึกษา 63
5.2 ประโยชน์ที่ได้รับจากการศึกษาและพัฒนาระบบ 64
5.3 ข้อเสนอแนะ 64

บรรณานุกรม 65

ประวัติผู้เขียน 66

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

4.1 ผลการทดสอบค่า Iris Data 46

4.2 ผลการทดสอบค่า Diabetes Data 47

4.3 ผลการทดสอบค่า Wine Data 48

4.4 ผลการทดสอบค่า Breast Data 49

4.5 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 1 50

4.6 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 2 51

4.7 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 3 51

4.8 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 4 51

4.9 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 5 52

4.10 ผลการทดลองจัดกลุ่มข้อมูล Iris โดยเฉลี่ย 52

4.11 ผลการทดลองจัดกลุ่มข้อมูล Diabetes Data ครั้งที่ 1 53

4.12 ผลการทดลองจัดกลุ่มข้อมูล Diabetes Data ครั้งที่ 2 53

4.13 ผลการทดลองจัดกลุ่มข้อมูล Diabetes Data ครั้งที่ 3 54

4.14 ผลการทดลองจัดกลุ่มข้อมูล Diabetes Data ครั้งที่ 4 54

4.15 ผลการทดลองจัดกลุ่มข้อมูล Diabetes Data ครั้งที่ 5 54

4.16 ผลการทดลองจัดกลุ่มข้อมูล Diabetes Data โดยเฉลี่ย 55

4.17 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 1 55

4.18 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 2 56

4.19 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 3 56

4.20 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 4 56

4.21 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 5 57

4.22 ผลการทดลองจัดกลุ่มข้อมูล wine data โดยเฉลี่ย 57

4.23 ผลการทดลองจัดกลุ่มข้อมูล Breast data ครั้งที่ 1 58

4.24 ผลการทดลองจัดกลุ่มข้อมูล Breast data ครั้งที่ 2 58

4.25 ผลการทดลองจัดกลุ่มข้อมูล Breast data ครั้งที่ 3 59

4.26 ผลการทดลองจัดกลุ่มข้อมูล Breast data ครั้งที่ 4 59

4.27 ผลการทดลองจัดกลุ่มข้อมูล Breast data ครั้งที่ 5 59

4.28 ผลการทดลองจัดกลุ่มข้อมูล Breast data โดยเฉลี่ย 60

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้า

2.1 ขั้นตอนการทำงานของ k-means 5
2.2 อัลกอริทึมการทำงานของ k-means 6
2.3 แสดงการย้ายกลุ่มของ Medoids กรณีแรก 9
2.4 แสดงการย้ายกลุ่มของ Medoids กรณีที่สอง 9
2.5 แสดงการย้ายกลุ่มของ Medoids กรณีที่สาม 10
2.6 แสดงการย้ายกลุ่มของ Medoids กรณีที่สี่ 10
2.7 แสดงการทำงานของ PAM 11
2.8 โครงสร้างของเซลล์ประสาท 16
2.9 สถาปัตยกรรมของโครงข่ายประสาทเทียม 17
2.10 ขั้นตอนการทำงานของโครงข่ายประสาทเทียม 17
2.11 Two self-organized feature maps 18
2.12 อัลกอริทึมการทำงานของ SOM 20
3.1 ลักษณะของอินพุทโปรแกรม 21
3.2 ลักษณะของเอาท์พุทโปรแกรม 22
3.3 ลักษณะของเอาท์พุทโปรแกรม 23
3.4 แสดง use case diagram 24
3.5 แสดง sequence diagram 25
3.6 แสดง sequence diagram ส่วนของ k-means 26
3.7 แสดง sequence diagram ส่วนของ PAM 27
3.8 แสดง sequence diagram ส่วนของ SOM 28
3.9 หน้าต่างMenu 28 28
3.10 หน้าต่างของการจัดกลุ่มโดย K-means 28
3.11 ข้อความแจ้งเตือนการกำหนดค่ากลุ่ม 29
3.12 ตัวอย่างการแสดงข้อมูลการจัดกลุ่ม 29
3.13 ตัวอย่างการแสดงข้อมูลศูนย์กลางการจัดกลุ่ม 30
3.14 หน้าต่างของการจัดกลุ่มโดย PAM 30
3.15 ข้อความแจ้งเตือนการกำหนดค่ากลุ่ม 31

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

3.16 ตัวอย่างการแสดงข้อมูลการจัดกลุ่ม 31
3.17 ตัวอย่างการแสดงข้อมูลศูนย์กลางการจัดกลุ่ม..... 31
3.18 หน้าต่างของการจัดกลุ่มโดย SOM..... 32
3.19 ข้อความแจ้งเตือนการกำหนดค่ากลุ่ม 32
3.20 ตัวอย่างการแสดงข้อมูลการจัดกลุ่ม 33
3.21 ตัวอย่างการแสดงข้อมูลศูนย์กลางการจัดกลุ่ม..... 33
3.22 แสดงการนำข้อมูลเข้าสู่โปรแกรม 33
3.23 แสดงการเลือกข้อมูลเข้าสู่โปรแกรม..... 33
3.24 แสดงการใช้งานของโปรแกรม 33
3.25 แสดงข้อมูลที่ได้ในโปรแกรม excel..... 35
3.26 การแสดงผลและข้อมูลเปรียบเทียบ 36
3.27 การแสดงผลในโปรแกรม excel..... 36
3.28 แสดงการนำข้อมูลเข้าสู่โปรแกรม..... 37
3.29 แสดงการเลือกข้อมูลเข้าสู่โปรแกรม..... 37
3.30 แสดงการใช้งานของโปรแกรม 38
3.31 แสดงข้อมูลที่ได้ในโปรแกรม excel..... 38
3.32 การแสดงผลและข้อมูลเปรียบเทียบ 39
3.33 การแสดงผลในโปรแกรม excel..... 39
3.34 แสดงการนำข้อมูลเข้าสู่โปรแกรม..... 40
3.35 แสดงการเลือกข้อมูลเข้าสู่โปรแกรม..... 40
3.36 แสดงการใช้งานของโปรแกรม 41
3.37 แสดงข้อมูลที่ได้ในโปรแกรม excel..... 41
3.38 การแสดงผลและข้อมูลเปรียบเทียบ 42
3.39 การแสดงผลในโปรแกรม excel..... 42
4.1 ตัวอย่างข้อมูล Iris Data 43
4.2 ข้อมูล Wine Data..... 44
4.3 ข้อมูล Diabetes Data..... 45
4.4 ข้อมูล Breast Cancer Data..... 46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมา

ในปัจจุบันความก้าวหน้าทางเทคโนโลยีด้านสารสนเทศได้รับการพัฒนาขึ้นอย่างมาก ทำให้มีการเก็บข้อมูลในรูปแบบต่างๆ อย่างต่อเนื่อง เพื่อที่จะนำข้อมูลเหล่านั้นมาใช้ประโยชน์อย่างเต็มที่ โดยกระบวนการในการวิเคราะห์ข้อมูลในการทำการแบ่งกลุ่มข้อมูล ได้แก่ การจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล (Partitioning Method) และ การทำงานโครงข่ายประสาทเทียม (Neural Network) ที่สามารถนำผลการการแบ่งกลุ่มข้อมูลไปใช้ประโยชน์ในการแบ่งกลุ่มข้อมูลได้ทั้งสองรูปแบบ จึงเกิดความคิดว่าข้อมูลชุดเดียวกัน อัลกอริทึมใดสามารถทำการแบ่งกลุ่มข้อมูลได้มีประสิทธิภาพได้มากกว่ากัน โดยการพัฒนาโปรแกรมการเปรียบเทียบประสิทธิภาพสำหรับการแบ่งกลุ่มข้อมูล โดยการใช้ข้อมูลชุดเดียวกันในการแบ่งกลุ่มข้อมูลผ่านกระบวนการของทั้ง แบบการจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล (Partitioning Method) และ การทำงานโครงข่ายประสาทเทียม (Neural Network)

การพัฒนาโปรแกรมเปรียบเทียบการแบ่งกลุ่มข้อมูลด้วยการจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล(Partitioning Method) ที่มีอัลกอริทึมในการทำงานคือ K-means อัลกอริทึม และ PAM (Partitioning around Medoids) เป็นอัลกอริทึมหนึ่งใน K- medoids และ การจัดกลุ่มแบบโครงข่ายประสาทเทียม (Neural Network) โดยวิธี SOM (Self-organizing map)ในการแบ่งกลุ่มข้อมูลเพื่อนำมาเปรียบเทียบอัลกอริทึมในการทำงานในการแบ่งกลุ่มข้อมูล

1.2 วัตถุประสงค์

โครงการพัฒนาระบบงานเรื่องการศึกษาเปรียบเทียบอัลกอริทึมในการแบ่งกลุ่มข้อมูลมีวัตถุประสงค์ดังนี้

1. เพื่อศึกษาการจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล(Partitioning Method) มี K-means และ PAM(Partitioning around Medoids)
2. เพื่อศึกษาระบบโครงข่ายประสาทเทียม (Neural Network) แบบที่มีการทำงานโดยใช้วิธีของ SOM (Self-organizing map)
3. เพื่อพัฒนาโปรแกรมแบบจำลองด้วยรูปแบบ การจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล (Partitioning Method) K-means , PAM และระบบโครงข่ายประสาทเทียม (Neural Network) SOM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. เพื่อเปรียบเทียบประสิทธิภาพทางด้าน การจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล (Partitioning Method) มี K-means, PAM และการทำงานโครงข่ายประสาทเทียม (Neural Network) SOM

1.3 ขอบเขตการดำเนินงาน

การศึกษาโครงงานนี้ได้กำหนดขอบเขตในการศึกษาเป็นการพัฒนาระบบการจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล(Partitioning Method) โดยวิธี K-means , PAM และ การจัดกลุ่มแบบโครงข่ายประสาทเทียม (Neural Network) โดยวิธี SOM ในการแบ่งกลุ่มข้อมูล เพื่อเปรียบเทียบประสิทธิภาพผลลัพธ์ในการแบ่งกลุ่มข้อมูล ผลจากการทำการทดลอง โดยใช้ข้อมูลชุดเดียวกันเพื่อนำมาเปรียบเทียบผลลัพธ์ที่ได้จากการแบ่งกลุ่มข้อมูลของแต่ละอัลกอริทึมนำมาวิเคราะห์ว่าแบบใดให้ผลลัพธ์ได้ถูกต้องกว่า ซึ่งมีขอบเขตการศึกษาหลักๆ ดังต่อไปนี้

- การจัดกลุ่มแบบแบ่งส่วนข้อมูล(Partitioning Method) มี K-means , PAM
- การจัดกลุ่มแบบโครงข่ายประสาทเทียม (Neural Network) SOM

1.4 ขั้นตอนและวิธีการดำเนินงาน

เพื่อให้การศึกษาเป็นไปตามวัตถุประสงค์ และขอบเขตที่กำหนด จึงได้กำหนดขั้นตอนในการดำเนินงานไว้ ดังนี้

1. ศึกษาทฤษฎีที่เกี่ยวข้องกับการจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล(Partitioning Method)
2. ศึกษาทฤษฎีที่เกี่ยวข้องกับการจัดกลุ่มข้อมูลแบบโครงข่ายประสาทเทียม (Neural Network)
3. กำหนดวัตถุประสงค์ในการพัฒนาโปรแกรม
4. พัฒนาโปรแกรมเตรียมข้อมูลเพื่อการจัดกลุ่มกลุ่มข้อมูล
5. ทดสอบการใช้งานโปรแกรม
6. สรุปผลการศึกษาและข้อเสนอแนะ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

ทำให้ทราบวิธีการพัฒนาระบบที่มีการนำกระบวนการการแบ่งกลุ่มข้อมูล(Clustering) โดยการจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล(Partitioning Method) และการจัดกลุ่มข้อมูลแบบโครงข่ายประสาทเทียม (Neural Network) มาประยุกต์ใช้กับการทำนายข้อมูลใน การแบ่งกลุ่มข้อมูลว่าระบบได้มีประสิทธิภาพในการจำแนกกลุ่มที่แม่นยำและมีความเหมาะสมที่จะนำรูปแบบใดรูปแบบหนึ่งไปใช้เป็นเครื่องมือในการจำแนกกลุ่มข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.6 เครื่องมือที่ใช้ในการพัฒนาระบบ

Visual Basic.NET หรือ VB.NET เป็นเครื่องมือที่ใช้พัฒนาโปรแกรม Visual Programming บนระบบปฏิบัติการ Windows ซึ่งได้รับการพัฒนามาจากภาษา BASIC (Beginners All Purpose Symbolic Instruction Code) ซึ่งเป็นภาษาโปรแกรมที่ได้รับความนิยมอย่างแพร่หลายสำหรับผู้เริ่มต้นหัดเขียนโปรแกรมคอมพิวเตอร์ เนื่องจากภาษา BASIC เป็นโปรแกรมที่สามารถทำความเข้าใจได้ง่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

ในหัวข้อนี้จะกล่าวถึงทฤษฎีพื้นฐานต่างๆ ที่เกี่ยวข้องในการพัฒนาระบบ เนื้อหาในบทนี้จะกล่าวถึง การจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล (Partitioning Method) ได้แก่ K- means ,PAM (Partitioning around Medoids) และ ทฤษฎีโครงข่ายประสาทเทียม(Neural Network) ได้แก่ SOM (Self-organizing map) เป็นอัลกอริทึมที่ใช้แบ่งกลุ่มข้อมูล โดยมีรายละเอียดที่เกี่ยวข้องกับการพัฒนาระบบงาน ดังนี้

2.1 การจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล(Partitioning Method)

การจัดกลุ่มประเภทนี้จะทำการสร้าง k พาร์ทิชั่น โดยแต่ละพาร์ทิชั่นจะแสดงถึงข้อมูลที่ถูกแบ่งออกเป็นกลุ่ม (โดยที่ $k \leq n$) ในแต่ละกลุ่มจะประกอบไปด้วยข้อมูลอย่างน้อยที่สุด 1 ชุดข้อมูล และแต่ละชุดข้อมูลจะต้องถูกจัดให้อยู่ในกลุ่มข้อมูลใดเพียงกลุ่มเดียวเท่านั้น การจัดกลุ่มประเภทนี้ต้องระบุค่า k หรือจำนวนกลุ่มที่ต้องการ โดยกระบวนการจัดกลุ่มเริ่มจากการสร้างกลุ่มตั้งต้น จากนั้นอัลกอริทึมจะทำการวนซ้ำเพื่อปรับข้อมูลให้เหมาะสมโดยการย้ายข้อมูลจากกลุ่มหนึ่งไปยังอีกกลุ่มหนึ่งที่มีความเหมาะสมกว่า โดยที่กลุ่มที่ดีจะต้องสามารถแบ่งให้ข้อมูลที่มีความใกล้ชิดกันหรือมีความสัมพันธ์กันอยู่ในกลุ่มเดียวกัน ส่วนข้อมูลที่มีความแตกต่างกันอยู่ในกลุ่มที่ไกลออกไป ซึ่งแต่ละอัลกอริทึมจะใช้เทคนิคในการพิจารณาความคล้ายคลึงกันของข้อมูลของแต่ละกลุ่มแตกต่างกันออกไป โดยอัลกอริทึมที่จะแสดงได้แก่ K-means , K-medoids

2.1.1 K- means

K-means อัลกอริทึมเป็นอัลกอริทึมหนึ่งใน Partitioning Methods เป็นการจัดกลุ่มข้อมูลที่มีความคล้ายคลึงกันอยู่กลุ่มเดียวกันโดยการใช้ค่าเฉลี่ยของข้อมูลที่ถูกจัดให้อยู่กลุ่มเดียวกันเป็นตัวแทนของกลุ่มนั้น การจัดกลุ่มข้อมูลจะทำการแบ่งข้อมูลออกเป็น k กลุ่ม และใช้ค่าเฉลี่ยของข้อมูลสร้างจุดศูนย์กลางใหม่ โดยจะพยายามปรับปรุงการแบ่งกลุ่มข้อมูล จากการเคลื่อนย้ายข้อมูลจากกลุ่มจนกระทั่งได้ข้อมูลที่มีความคล้ายคลึงกันอยู่ในกลุ่มเดียวกันและข้อมูลที่มีความแตกต่างกันอยู่คนละ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.2 การทำงานของ K- means

1. กำหนดจำนวน k กลุ่มเพื่อเริ่มจุดศูนย์กลางของข้อมูล
2. สุ่มเลือกข้อมูลมา k ข้อมูลตามจำนวนกลุ่มเพื่อเป็นจุดศูนย์กลางของข้อมูล
3. กำหนดสมาชิกของข้อมูลให้อยู่ในกลุ่มข้อมูล โดยการวัดระยะห่างจากสูตร (Euclidean distance) ของสมาชิกกับศูนย์กลางของกลุ่ม หากสมาชิกอยู่ใกล้จุดศูนย์กลางของกลุ่มใดมากสุดก็นำไปรวมกับกลุ่มนั้น

   สมการ2.1 ฟังก์ชั่นวัดระยะห่างจากสูตร (Euclidean distance)

$$D_{(i,j)} = \sqrt{(X_{in} - C_{jn})^2} \qquad (2.1)$$

4. ในแต่ล่ะกลุ่มคำนวณค่าเฉลี่ยใหม่เพื่อสร้างจุดศูนย์กลางของกลุ่มใหม่
5. ทำขั้นตอน 3-4 ซ้ำจนกระทั่งศูนย์กลางของกลุ่มไม่เปลี่ยนแปลง

(a) (b) (c)

รูปที่ 2.1 ขั้นตอนการทำงานของ K- means

จากรูป อธิบายการทำงานของ K- means ได้ดังนี้ ภาพ (a) เราเลือกข้อมูลมา 3 ตัวเพื่อเป็นจุดศูนย์กลางของกลุ่มข้อมูลที่เราเลือกเป็นจุดศูนย์กลางของกลุ่มข้อมูลถูกแทนที่ด้วยเครื่องหมาย "+" หลังจากนั้นวัดระยะห่างจากแต่ละข้อมูลกับจุดศูนย์กลางของกลุ่มนั้น ข้อมูลใดอยู่ใกล้ศูนย์กลางไหนมากที่สุดก็จะไปอยู่ในกลุ่มนั้น การกระจายสมาชิกข้อมูลแทนที่ด้วย จุดสีดำ ภาพ (b) เมื่อทำการรวมกลุ่มเรียบร้อยแล้ว จะคำนวณค่าเฉลี่ยของข้อมูลแต่ละกลุ่มใหม่เพื่อที่จะปรับปรุงศูนย์กลางของแต่ละกลุ่ม ถ้าข้อมูลใกล้เคียงกับกลุ่มอื่นก็จะทำการย้ายกลุ่ม ภาพ(c) กระบวนการนี้จะทำการวนซ้ำจนกระทั่งศูนย์กลางกลุ่มไม่เปลี่ยนแปลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Algorithm: *k*-means. The *k*-means algorithm for partitioning based on the mean value of the object in the cluster.
Input: The number of cluster *k* and a database containing *n* objects.
Output: A set of *k* clusters that minimizes the squared-error criterion.
Method:
(1) arbitrarily choose *k* objects as the initial cluster center;
(2) repeat
(3) (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
(4) update the cluster means, i.e., calculate the mean value of the objects for each cluster;
(5) until no change;

รูปที่ 2.2 อัลกอริทึมการทำงานของ K- means

2.1.3 ตัวอย่างการจัดกลุ่มของ K- means

ชุดข้อมูลมีทั้งหมด 10 ชุด จำนวนกลุ่ม 2 กลุ่ม

จำนวนกลุ่มเท่ากับ k = 2

ชุดข้อมูล

X1 [2 6]

X2 [3 4]

X3 [3 8]

X4 [4 7]

X5 [6 2]

X6 [6 4]

X7 [7 3]

X8 [7 4]

X9 [8 5]

X10[7 6]

ขั้นตอนที่ 1 สุ่มชุดข้อมูลเท่ากับจำนวนกลุ่ม k=2

จุดศูนย์กลางที่สุ่มได้

C1[3 4]

C2[7 4]

ขั้นตอนที่2 หาระยะห่างระหว่างข้อมูลแต่ละชุดกับจุดศูนย์กลาง

$D(x1,c1) = \text{sqrt}((2-3)^2+(6-4)^2=2.23$

$D(x2,c1) = \text{sqrt}(3-3)^2+(4-4)^2=0$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

D(x3,c1) = sqrt(3-3)²+(8-4)²=4

D(x4,c1) = sqrt(4-3)²+(7-4)²=3.16

D(x5,c1) = sqrt(6-3)²+(2-4)²=3.60

D(x6,c1) = sqrt(6-3)²+(4-4)²=3

D(x7,c1) = sqrt(7-3)²+(3-4)²=4.12

D(x8,c1) = sqrt(7-3)²+(4-4)²=4

D(x9,c1) = sqrt(8-3)²+(5-4)²=5.09

D(x10,c1) = sqrt(7-3)²+(6-4)²=4.47

D(x1,c2) =sqrt((2-7)²+(6-4)²=5.38

D(x2,c2) = sqrt(3-7)²+(4-4)²=4

D(x3,c2) = sqrt(3-7)²+(8-4)²=5.65

D(x4,c2) = sqrt(4-7)²+(7-4)²=4.24

D(x5,c2) = sqrt(6-7)²+(2-4)²=2.23

D(x6,c2) = sqrt(6-7)²+(4-4)²=1

D(x7,c2) = sqrt(7-7)²+(3-4)²=1

D(x8,c2) = sqrt(7-7)²+(4-4)²=0

D(x9,c2) = sqrt(8-7)²+(5-4)²=1.41

D(x10,c2) = sqrt(7-7)²+(6-4)²=2

ขั้นตอนที่ 3 คำนวณหาค่าเฉลี่ยใหม่เพื่อสร้างจุดศูนย์กลางของกลุ่มใหม่

กลุ่ม C1

X1 [2 6]

X2 [3 4]

X3 [3 8]

X4 [4 7]

C1=[(2+3+3+4)/4 ,(6+4+8+7)/4]

C2=[(6+6+7+7+8+7)/6 , (2+4+3+4+5+6)/6]

จุดศูนย์กลางใหม่ที่คำนวณได้

C1=[3 6.25]

C2=[6.83 4]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำขั้นตอนที่สองและสามซ้ำ จนกระทั่งกลุ่มไม่เปลี่ยนแปลง ระยะห่างระหว่างข้อมูลแต่ละชุดกับจุดศูนย์กลาง

D(x1,c1) =sqrt((2-3)²+(6-6.25)²=1.03

D(x2,c1) = sqrt(3-3)²+(4-6.25)²=2.25

D(x3,c1) = sqrt(3-3)²+(8-6.25)²=1.75

D(x4,c1) = sqrt(4-3)²+(7-6.25)²=1.25

D(x5,c1) = sqrt(6-3)²+(2-6.25)²=5.20

D(x6,c1) = sqrt(6-3)²+(4-6.25)²=3.75

D(x7,c1) = sqrt(7-3)²+(3-6.25)²=5.15

D(x8,c1) = sqrt(7-3)²+(4-6.25)²=4.58

D(x9,c1) = sqrt(8-3)²+(5-6.25)²=5.15

D(x10,c1) = sqrt(7-3)²+(6-6.25)²=4.00

D(x1,c2) =sqrt((2-6.83)²+(6-4)²=5.22

D(x2,c2) = sqrt(3-6.83)²+(4-4)²=3.83

D(x3,c2) = sqrt(3-6.83)²+(8-4)²=5.53

D(x4,c2) = sqrt(4-6.83)²+(7-4)²=4.12

D(x5,c2) = sqrt(6-6.83)²+(2-4)²=2.16

D(x6,c2) = sqrt(6-6.83)²+(4-4)²=0.68

D(x7,c2) = sqrt(7-6.83)²+(3-4)²=1.01

D(x8,c2) = sqrt(7-6.83)²+(4-4)²=0.17

D(x9,c2) = sqrt(8-6.83)²+(5-4)²=1.53

D(x10,c2) = sqrt(7-6.83)²+(6-4)²=2.00

หลังจากพิจารณากลุ่มไม่มีการเปลี่ยนแปลงกลุ่มจะ ได้การจัดกลุ่มได้ดังนี้

X1 [2 6] กลุ่มที่ 1

X2 [3 4] กลุ่มที่ 1

X3 [3 8] กลุ่มที่ 1

X4 [4 7] กลุ่มที่ 1

X5 [6 2] กลุ่มที่ 2

X6 [6 4] กลุ่มที่ 2

X7 [7 3] กลุ่มที่ 2

X8 [7 4] กลุ่มที่ 2

X9 [8 5] กลุ่มที่ 2

X10[7 6] กลุ่มที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.1.4 K- medoids

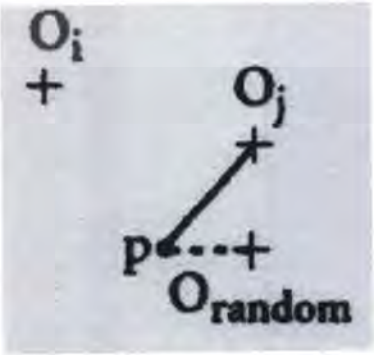
เป็นเทคนิคที่พัฒนาขึ้นต่อจาก K- means เนื่องจาก K- means มีความอ่อนไหวต่อข้อมูลรบกวน ดังนั้นอัลกอริทึม K-medoids จึงถูกออกแบบมาเพื่อขจัดข้อด้อยดังกล่าว โดยการใช้อ๊อบเจกต์ที่อยู่ใกล้กับจุดกึ่งกลางของคลัสเตอร์มากที่สุดเป็นตัวแทนของทุกอ๊อบเจกต์ในคลัสเตอร์นั้นแทนการใช้ค่าเฉลี่ยของคลัสเตอร์ โดยที่กระบวนการคล้ายกับ K- means คือการจัดกลุ่มข้อมูลโดยพิจารณาความคล้ายคลึงกันของอ๊อบเจกต์กับตัวแทนของคลัสเตอร์แต่ต่างกันเพียงวิธีหาตัวแทนคลัสเตอร์เท่านั้น PAM(Partitioning around Medoids)เป็นอัลกอริทึมหนึ่งใน K- medoids การจัดกลุ่มข้อมูลจะทำการแบ่งข้อมูลออกเป็น k กลุ่มและใช้อ๊อบเจกต์ที่ใกล้จุดศูนย์กลางมากที่สุดเป็นศูนย์กลางและเลือกอ๊อบเจกต์ที่ไม่ใช่ศูนย์กลางของกลุ่มหาระยะห่างเปรียบเทียบค่าที่ได้กับศูนย์กลางถ้าค่าที่ได้น้อยกว่า จะเปลี่ยนจุดศูนย์กลาง จนกระทั่งได้ข้อมูลที่มีความคล้ายคลึงกันอยู่ในกลุ่มเดียวกันศูนย์กลางไม่เปลี่ยนแปลง

2.1.5 กรณีการย้ายกลุ่มของ Medoids แบบต่างๆ

กรณีแรก ถ้า P อยู่ในกลุ่ม Oj Oj เป็นจุดศูนย์กลาง ถ้าจุดศูนย์กลาง Oj ถูกแทนที่ด้วยจุดศูนย์กลาง Orandom , $i \neq j$ P อยู่ใกล้กับจุดศูนย์กลาง ดังนั้น P จะไปอยู่กลุ่ม Oi

รูปที่ 2.3 แสดงการย้ายกลุ่มของ Medoids กรณีแรก

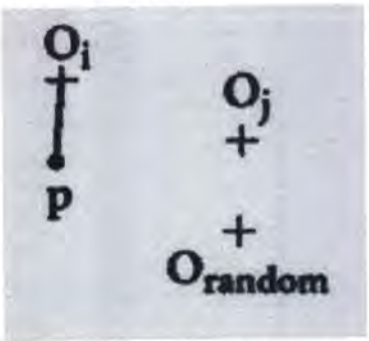
กรณีที่สอง ถ้า P อยู่ในกลุ่ม Oj Oj เป็นจุดศูนย์กลาง ถ้าจุดศูนย์กลาง Oj ถูกแทนที่ด้วยจุดศูนย์กลาง Orandom และ P อยู่ใกล้กับจุดศูนย์กลาง Orandom ดังนั้น P จะไปอยู่กลุ่ม Orandom



รูปที่ 2.4 แสดงการย้ายกลุ่มของ Medoids กรณีที่สอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กรณีที่สาม ถ้า P อยู่ในกลุ่ม Oi Oi เป็นจุดศูนย์กลาง ถ้าจุดศูนย์กลาง Oj ถูกแทนที่ด้วยจุดศูนย์กลาง Orandom , $i \neq j$ ดังนั้น P จะไปอยู่กลุ่มเดิม



รูปที่ 2.5 แสดงการย้ายกลุ่มของ Medoids กรณีที่สาม

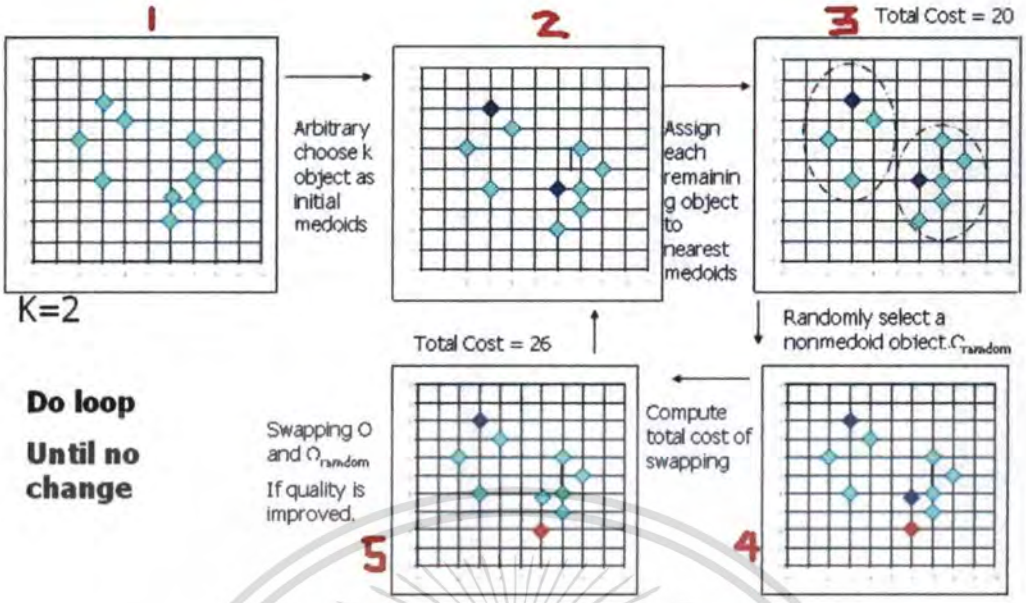
กรณีที่สี่ ถ้า P อยู่ในกลุ่ม Oi Oi เป็นจุดศูนย์กลาง , $i \neq j$ แล้ว P อยู่ใกล้กับจุดศูนย์กลาง Orandom ถ้าจุดศูนย์กลาง Oj ถูกแทนที่ด้วยจุดศูนย์กลาง Orandom ดังนั้น P จะไปอยู่กลุ่ม Orandom

รูปที่ 2.6 แสดงการย้ายกลุ่มของ Medoids กรณีที่สี่

2.1.6 การทำงานของPAM

1. กำหนดจำนวน k กลุ่มเพื่อเริ่มจุดศูนย์กลาง (medoids) ของข้อมูล
2. สุ่มเลือกข้อมูลมา k ข้อมูลตามจำนวนกลุ่มเพื่อเป็นจุดศูนย์กลางของข้อมูล
3. กำหนดสมาชิกของข้อมูลให้อยู่ในกลุ่มข้อมูล โดยการวัดระยะห่างจากสูตร (Euclidean distance) ของสมาชิกกับศูนย์กลางของกลุ่ม โดยค่าที่ได้กำหนดเป็น TCcurrent
4. เลือกข้อมูลทั้งหมดที่ไม่ใช้จุดศูนย์กลาง (non-medoids) ทำการวัดระยะห่างกับข้อมูล ค่าที่ได้กำหนดเป็น TC เลือกค่า TC น้อยที่สุดมาเปรียบเทียบกับ TCcurrent ถ้า TC น้อยกว่า TCcurrent เปลี่ยนจุดศูนย์กลางข้อมูลจาก TC เป็น TCcurrent
5. ทำซ้ำจนกระทั่งไม่มีการเปลี่ยนแปลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.7 แสดงการทำงานของ PAM

จากรูปแสดงการจัดกลุ่มโดยกลุ่มที่ต้องการจัดกลุ่ม 2 กลุ่ม k=2 รูปที่ 2 จะทำการสุ่มเลือกข้อมูลที่จะเป็นจุดศูนย์กลาง รูปที่ 3 คำนวณหาระยะห่างจุดศูนย์กลาง ได้ค่า TCcurrent รูปที่ 4 ทำการวัดระยะห่างกับข้อมูลทั้งหมดที่ไม่ใช้จุดศูนย์กลาง (non-medoids) หาค่า TC รูปที่ 5 เปรียบเทียบค่า TCcurrent กับ TC ถ้า TCcurrent น้อยกว่า TC ไม่เปลี่ยนจุดศูนย์กลางทำการทดลองจนกว่าจะครบข้อมูลทุกตัว

2.1.7 ตัวอย่างการจัดกลุ่มของ PAM

ชุดข้อมูลมีทั้งหมด 10 ชุด จำนวนกลุ่ม 2 กลุ่ม

จำนวนกลุ่มเท่ากับ k = 2

ชุดข้อมูล

X1 [2 6]

X2 [3 4]

X3 [3 8]

X4 [4 7]

X5 [6 2]

X6 [6 4]

X7 [7 3]

X8 [7 4]

X9 [8 5]

X10[7 6]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่1 สุ่มชุดข้อมูลเท่ากับจำนวนกลุ่ม k=2

จุดศูนย์กลางที่สุ่มได้

C1[3 4]

C2[7 4]

ขั้นตอนที่2 หาระยะห่างระหว่างข้อมูลแต่ละชุดกับจุดศูนย์กลาง

D(x1,c1) =sqrt((2-3)²+(6-4)²=2.23

D(x2,c1) = sqrt(3-3)²+(4-4)²=0

D(x3,c1) = sqrt(3-3)²+(8-4)²=4

D(x4,c1) = sqrt(4-3)²+(7-4)²=3.16

D(x5,c1) = sqrt(6-3)²+(2-4)²=3.60

D(x6,c1) = sqrt(6-3)²+(4-4)²=3

D(x7,c1) = sqrt(7-3)²+(3-4)²=4.12

D(x8,c1) = sqrt(7-3)²+(4-4)²=4

D(x9,c1) = sqrt(8-3)²+(5-4)²=5.09

D(x10,c1) = sqrt(7-3)²+(6-4)²=4.47

D(x1,c2) =sqrt((2-7)²+(6-4)²=5.38

D(x2,c2) = sqrt(3-7)²+(4-4)²=4

D(x3,c2) = sqrt(3-7)²+(8-4)²=5.65

D(x4,c2) = sqrt(4-7)²+(7-4)²=4.24

D(x5,c2) = sqrt(6-7)²+(2-4)²=2.23

D(x6,c2) = sqrt(6-7)²+(4-4)²=1

D(x7,c2) = sqrt(7-7)²+(3-4)²=1

D(x8,c2) = sqrt(7-7)²+(4-4)²=0

D(x9,c2) = sqrt(8-7)²+(5-4)²=1.41

D(x10,c2) = sqrt(7-7)²+(6-4)²=2

ขั้นตอนที่3 นำค่าระยะห่างระหว่างข้อมูลแต่ละชุดกับจุดศูนย์กลางที่มีค่าน้อยมาบวกกันเพื่อได้ค่า TOTAL COST current เพื่อใช้ในการเปรียบเทียบกับจุดศูนย์กลางอื่นๆ

ค่า TCcurrent = 2.23+0+4+3.16+2.23+1+1+0+1.41+2=17.03

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่4 เปลี่ยนจุดศูนย์กลาง C1,C2 กับข้อมูลทุกชุด เลือกค่า cost น้อยที่สุดเป็นตัวแทนของจุดศูนย์กลางนั้น ดังตัวอย่าง

เปลี่ยนจุดศูนย์กลาง C1 กับข้อมูลทุกชุด

C1 [2 6] C2 [7 4]

ค่า TC =13.74

C1 [3 4] C2 [7 4]

ค่า TC =17.03

C1 [3 8] C2 [7 4]

ค่า TC =14.69

C1 [4 7] C2 [7 4]

ค่า TC =13.85

C1 [6 2] C2 [7 4]

ค่า TC =20.86

C1 [6 4] C2 [7 4]

ค่า TC =21.3

C1 [7 3] C2 [7 4]

ค่า TC =25.5

C1 [7 4] C2 [7 4]

ค่า TC =26.34

C1 [8 5] C2 [7 4]

ค่า TC =24.93

C1 [7 6] C2 [7 4]

ค่า TC =22.27

ค่า TC ของ C1 [2 6] ที่น้อยกว่า TC current ของ C1[3 4] เท่ากับ 13.74

จุดศูนย์กลางของ C1 เท่ากับ [2 6]

เปลี่ยนจุดศูนย์กลาง C2 กับข้อมูลทุกชุด

C1 [2 6] C2[2 6]

ค่า TC =28.39

C1 [2 6] C2 [3 4]

ค่า TC =35.27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

C1 [2 6] C2 [3 8]

ค่า TC =35.27

C1 [2 6] C2 [4 7]

ค่า TC =29.49

C1 [2 6] C2 [6 2]

ค่า TC =19.46

C1 [2 6] C2 [6 4]

ค่า TC =15.56

C1 [2 6] C2 [7 3]

ค่า TC =15.74

C1 [2 6] C2 [7 4]

ค่า TC =17.57

C1 [2 6] C2 [8 5]

ค่า TC =19.45

C1 [2 6] C2 [7 6]

ค่า TC =13.74

ค่า TC ของ [7 4] ที่น้อยกว่า TC current ของ C2 [7 4] เท่ากับ 17.57

จุดศูนย์กลางของ C2 เท่ากับ [7 4]

ขั้นตอนที่5 นำกลุ่มที่ได้ไปหาระยะห่างระหว่างข้อมูลแต่ละชุดกับจุดศูนย์กลาง

D(x1,c1) =sqrt((2-2)²+(6-6)²=0

D(x2,c1) = sqrt(3-2)²+(4-6)²=2.23

D(x3,c1) = sqrt(3-2)²+(8-6)²=2.23

D(x4,c1) = sqrt(4-2)²+(7-6)²=2.23

D(x5,c1) = sqrt(6-2)²+(2-6)²=4.47

D(x6,c1) = sqrt(6-2)²+(4-6)²=4.47

D(x7,c1) = sqrt(7-2)²+(3-6)²=5.83

D(x8,c1) = sqrt(7-2)²+(4-6)²=5.38

D(x9,c1) = sqrt(8-2)²+(5-6)²=6.08

D(x10,c1) = sqrt(7-2)²+(6-6)²=5.00

D(x1,c2) =sqrt((2-2)²+(6-4)²=2

D(x2,c2) = sqrt(3-7)²+(4-4)²=4

D(x3,c2) = sqrt(3-7)²+(8-4)²=5.65

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$D(x4,c2) = \sqrt{(4-7)^2+(7-4)^2}=4.24$

$D(x5,c2) = \sqrt{(6-7)^2+(2-4)^2}=2.23$

$D(x6,c2) = \sqrt{(6-7)^2+(4-4)^2}=1$

$D(x7,c2) = \sqrt{(7-7)^2+(3-4)^2}=1$

$D(x8,c2) = \sqrt{(7-7)^2+(4-4)^2}=0$

$D(x9,c2) = \sqrt{(8-7)^2+(5-4)^2}=1.41$

$D(x10,c2) = \sqrt{(7-7)^2+(6-4)^2}=1.41$

หลังจากพิจารณาการจัดกลุ่มได้ดังนี้

X1 [2 6] กลุ่มที่ 1

X2 [3 4] กลุ่มที่ 1

X3 [3 8] กลุ่มที่ 1

X4 [4 7] กลุ่มที่ 1

X5 [6 2] กลุ่มที่ 2

X6 [6 4] กลุ่มที่ 2

X7 [7 3] กลุ่มที่ 2

X8 [7 4] กลุ่มที่ 2

X9 [8 5] กลุ่มที่ 2

X10[7 6] กลุ่มที่ 2

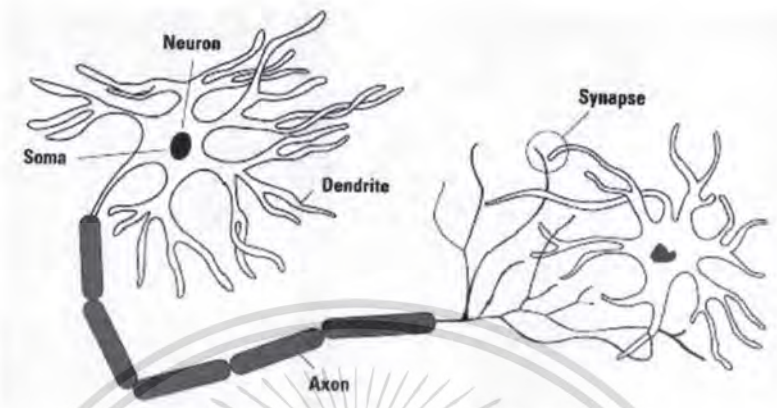
2.2 ทฤษฎีโครงข่ายประสาทเทียม

2.2.1 โครงสร้างของโครงข่ายประสาทเทียม

โครงสร้างของโครงข่ายประสาทเทียม มีลักษณะคล้ายคลึงกับสมองของมนุษย์ซึ่งประกอบด้วยเซลล์ประสาทที่เรียกว่านิวรอล (Neuron) จำนวนมากเชื่อมต่อกัน โดยแต่ละนิวรอลจะมีเส้นเชื่อมต่อกันเพื่อทำการรับส่งข้อมูล ซึ่งแต่ละนิวรอลสามารถมีอินพุทได้หลายอินพุต แต่จะมีเอาท์พุตเพียงเอาท์พุตเดียว เอาท์พุตจะส่งต่อไปยังอินพุตของนิวรอลอื่นๆ เป็นการส่งสัญญาณในระยะเวลาที่รวดเร็วมาก เนื่องจากมีรูปแบบการส่งสัญญาณในลักษณะขนาน (Parallel) กัน ทำให้สามารถทำงานพร้อมๆ กันได้หลายงาน และสำหรับวิธีการแก้ปัญหาของมนุษย์จะใช้ประสบการณ์ที่เกิดขึ้นจากการเรียนรู้ข้อมูลต่างๆ ในอดีตนำมาวิเคราะห์ แต่ในขณะที่สถาปัตยกรรมของคอมพิวเตอร์นั้น จะประกอบด้วยหน่วยประมวลผลเพียงอย่างเดียวเท่านั้น รวมถึงในการทำงานนั้นจะใช้ชุดคำสั่งที่เป็นลำดับ ทำให้การแก้ปัญหาต้องทำตามลำดับที่แน่นอน จึงอาจมีปัญหาที่ไม่สามารถแก้ปัญหาโดยใช้การทำงานในลักษณะนี้ได้ เช่น การจดจำหน้าตาของคน ซึ่งสมองของมนุษย์นั้นสามารถแก้ปัญหาลักษณะเช่นนี้ได้ ถึงแม้เวลาจะผ่านมาสักระยะหนึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แล้ว ซึ่งต่างจากคอมพิวเตอร์ ดังนั้นจึงได้มีการพัฒนาคอมพิวเตอร์ให้สามารถประมวลผลได้ในลักษณะที่คล้ายคลึงกับสมองของมนุษย์ หรือที่รียกว่า โครงข่ายประสาทเทียม (Artificial Neural Network)



รูปที่ 2.8 โครงสร้างของเซลล์ประสาท

โครงข่ายประสาทเทียม ประกอบด้วยกลุ่มของข้อมูลอินพุต และข้อมูลเอาท์พุต ที่มีการเชื่อมต่อกันนั้น จะมีการให้น้ำหนัก (Weights) เพื่อให้ได้ผลลัพธ์ที่ถูกต้องมากที่สุด ลักษณะการเชื่อมต่อระหว่างหน่วยที่อยู่ต่างชั้น (Layer) เพื่อทำการส่งสัญญาณมี 2 ลักษณะ คือ

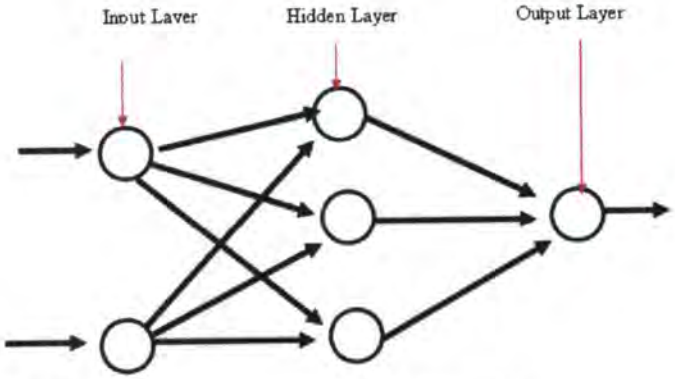
- Feedforward : ลักษณะการไหลของข้อมูลของการเชื่อมต่อลักษณะนี้คือ ข้อมูลจะไหลไปตามการเชื่อมต่อในทิศทางเดียว จะไม่มีการไหลย้อนกลับ
- Feedback : ลักษณะการไหลของข้อมูลของการเชื่อมต่อลักษณะนี้คือ ข้อมูลจะไหลย้อนกลับได้ หรืออาจจะมีการไหลของข้อมูลในลักษณะเรียกซ้ำได้

โดยทั่วไปโครงข่ายประสาทเทียมจะแบ่งออกเป็น 3 ชั้น (Layer) ด้วยกันคือ

1. ชั้นอินพุต (Input Layer) ทำหน้าที่รับสัญญาณจากภายนอกผ่านเข้ามาในโครงข่าย แล้วกระจายไปในชั้นถัดไป โดยไม่มีการคำนวณใดๆ
2. ชั้นซ่อน (Hidden Layer) เป็นชั้นที่อยู่ระหว่างกลางระหว่างชั้นอินพุตและชั้นเอาท์พุต ซึ่งในชั้นนี้สามารถมีได้มากกว่า 1 ชั้น
3. ชั้นเอาท์พุต (Output Layer) ทำหน้าที่ส่งผลลัพธ์ที่คำนวณได้ออกจากโครงข่าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ห้องสมุดคณะเทคโนโลยีสารสนเทศ พระจอมเกล้าลาดกระบัง

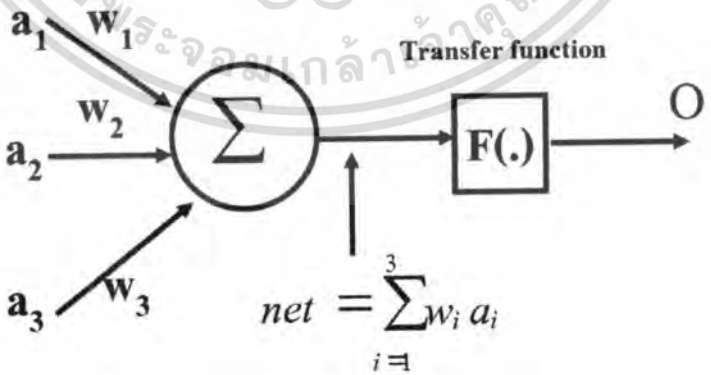


รูปที่ 2.9 สถาปัตยกรรมของโครงข่ายประสาทเทียม

2.2.2 การทำงานของโครงข่ายประสาทเทียม

การทำงานของโครงข่ายประสาทเทียมจะขึ้นอยู่กับค่าน้ำหนัก และฟังก์ชันโยกย้าย (Transfer Function) โดยมีการกำหนดค่าน้ำหนักเริ่มต้นของโครงข่ายประสาทเทียมอยู่ระหว่าง -0.5 ถึง 0.5 และสำหรับฟังก์ชันโยกย้ายหรือฟังก์ชันกระตุ้นนั้น จะถือเป็นฟังก์ชันที่ทำหน้าที่กระตุ้นการทำงานให้เกิดการย้ายค่าอินพุต ไปยังเอาท์พุตซึ่งจะมีค่าอยู่ระหว่าง 0 ถึง 1 โดยทั่วไปแล้วฟังก์ชันโยกย้ายที่นิยมใช้ คือ Sigmoid Function ซึ่งมีขั้นตอนการทำงานดังนี้

- ขั้นตอนแรก เมื่อได้อินพุตเข้ามาในโครงข่ายแล้ว ก็จะคูณค่าอินพุตนั้นกับค่าน้ำหนักในแต่ละลิ้งค์
- ขั้นที่สอง ทำการบวกผลลัพธ์ที่ได้จากขั้นตอนแรกเข้าด้วยกัน
- ขั้นที่สาม จะนำผลรวมในขั้นตอนที่สองมาผ่านฟังก์ชั่นโยกย้าย แล้วจะได้ผลลัพธ์เป็นค่าเอาท์พุตของโครงข่าย



รูปที่ 2.10 ขั้นตอนการทำงานของโครงข่ายประสาทเทียม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
06296

2.2.3 ลักษณะการเรียนรู้ของโครงข่ายประสาทเทียม

ลักษณะการเรียนรู้ของโครงข่ายประสาทเทียม สามารถแบ่งได้เป็น 2 ประเภท คือ

1. การเรียนรู้แบบมีผู้สอน (Supervised Learning)

การเรียนรู้แบบมีผู้สอน คือ จะมีการกำหนดเซตของการฝึกหัดให้กับโครงข่ายประสาทเทียม ซึ่งเซตนี้จะประกอบด้วยอินพุต และเอาท์พุตที่ต้องการ โดยทำการจับคู่สอน (Train Pair) เมื่อทำการป้อนอินพุตให้กับโครงข่ายแล้ว โครงข่ายจะทำการประมวลผลจนได้คำตอบ และค่าถ่วงน้ำหนัก (Weight) ออกมาชุดหนึ่ง สำหรับคำตอบที่ได้จริง (Actual Output) เพื่อนำมาคำนวณค่าผิดพลาด โดยสามารถหาได้ จากการนำคำตอบที่เราต้องการ (Target Output) ลบกับคำตอบที่ได้จริงจากการเรียนรู้ของโครงข่าย ถ้าค่าความผิดพลาดที่ได้จากการลบกันระหว่างค่าที่ได้จริงกับค่าที่เราต้องการแล้วยังมีค่าสูง หรือ ในบางกรณีอาจจะต้องปรับให้มีค่าเป็นศูนย์ จึงต้องการทำปรับค่าถ่วงน้ำหนักไปจนกว่าจะได้ค่าที่ยอมรับได้จึงหยุดการสอนโครงข่าย

2. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning)

การเรียนรู้แบบไม่มีผู้สอน คือ การเรียนรู้โดยที่ไม่ต้องมีการจับคู่สอน เนื่องจากเมื่อเราใส่ค่าอินพุตเซตเข้าสู่โครงข่ายแล้วมันจะพยายามจัดกลุ่มอินพุตเซตที่มีลักษณะเดียวกันให้เอาท์พุตออกมาจากโครงข่ายที่เดียวกัน คือ มีการปรับตัวเองภายใน (Self-Organization) การเรียนรู้โดยวิธีนี้จะป้อนอินพุตเข้าสู่โครงข่ายภายใน โครงข่ายจะมีเอาท์พุตโหนด อยู่หลายโหนด โดยแต่ละโหนด แทนกลุ่มของข้อมูลที่มีคุณสมบัติเหมือนกัน เมื่อป้อนอินพุตเข้าสู่โครงข่าย โครงข่ายจะทำการคำนวณค่าความสัมพันธ์ที่มีภายในเซตของอินพุตของโครงข่าย การเรียนรู้ คือการใช้ค่าน้ำหนักเป็นตัวแยกความแตกต่างของอินพุต ทำให้การเรียนรู้ในลักษณะนี้ผู้ใช้งานจะต้องทำการกำหนดเอาท์พุตเอง ส่งผลให้ค่าที่ได้อาจจะไม่ถูกต้อง

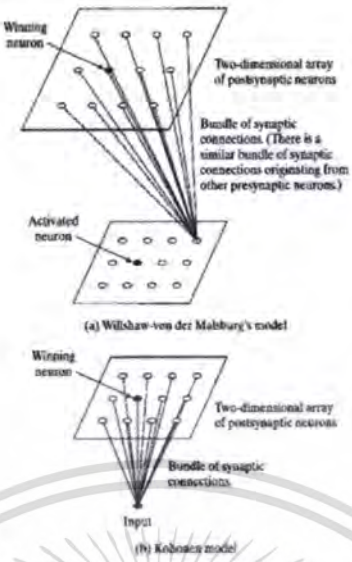
2.3 อัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ป

2.3.1 ลักษณะของอัลกอริทึมเซ็ลฟออร์แกไนซิงแม็ป

เซ็ลฟออร์แกไนซิงแม็ปพัฒนาโดย Kohonen แบบจำลอง SOM เป็นแบบจำลองระบบโครงข่ายแบบไม่มีผู้สอน (Unsupervised Learning) คือ การเรียนรู้ที่ไม่ต้องมีการจับคู่สอน เมื่อใส่ค่าอินพุตเซตเข้าสู่โครงข่ายแล้วมันจะพยายามจัดกลุ่มอินพุตเซตที่มีลักษณะเดียวกันให้เอาท์พุตออกมาจากโครงข่ายที่เดียวกัน คือ การปรับตัวภายใน (Self –Organizing) การเรียนรู้โดยวิธีนี้จะป้อนอินพุตเข้าสู่โครงข่ายภายใน โครงข่ายจะมีเอาท์พุตโหนดอยู่หลายโหนด โดยแต่ละโหนดแทนกลุ่มของข้อมูลที่มีคุณสมบัติเหมือนกัน เมื่อป้อนอินพุตเข้าสู่โครงข่าย โครงข่ายจะคำนวณหาค่าความสัมพันธ์ที่มีภายในเซตของอินพุตของโครงข่าย การเรียนรู้ คือ การใช้ค่าน้ำหนักเป็นตัวแยกความแตกต่างของอินพุต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2 โครงสร้างของเซ็ลฟออร์แกไนซิงแม็ป



รูปที่ 2.11 Two self-organized feature maps

เซ็ลฟออร์แกไนซิงแม็ป ประกอบด้วยชั้นของเซลล์ประสาทที่แท้จริงเพียงชั้นเดียวคือ ชั้นฮิด-เดนเลเยอร์ เพราะเป็นชั้นที่นำค่าอินพุตเข้ามาคิดคำนวณ แล้วนำค่าที่ได้ส่งไปชั้นเอาต์พุตต่อไป ลักษณะของ Self-Organizing เป็นการปรับตัวภายใน การเรียนรู้โดยวิธีนี้จะป้อนอินพุตเข้าสู่เครือข่าย ภายในเครือข่ายจะมีเอาท์พุตอยู่หลายโหนด โดยแต่ละโหนดแทนกลุ่มของข้อมูลที่มีคุณสมบัติเหมือนกัน เมื่อป้อนอินพุตเข้าสู่เครือข่าย เครือข่ายจะทำการคำนวณหาค่าความสัมพันธ์ที่มีภายในเซตของอินพุต จุดหนึ่งที่ไม่เหมือนกับการทำ Clustering ในแบบอื่นๆ ของ SOM ไม่ได้ใช้การเปรียบเทียบระหว่างข้อมูลด้วยกันเองแต่เป็นการเปรียบเทียบข้อมูลกับ neural output

2.3.3 ขั้นตอนการทำงานของเซ็ลฟออร์แกไนซิงแม็ป

วิธีการคิดของ Kohonen's Self-Organizing Map คือการทำซ้ำของในแต่ละข้อมูลเพื่อที่จะได้ค่าน้ำหนักของข้อมูลที่มีอยู่ทั้งหมดตามจำนวนกลุ่มที่ต้องการแบ่ง มีวิธีการคิดดังนี้

การทำงานของ SOM (Self-organizing map)

1. กำหนดค่าโหนด Neural และกำหนดอัตราการเรียนรู้ (α) ระยะห่างโหนดข้างเคียง (λ) และ ค่าซิกม่า(σ)
2. เริ่มต้นให้กับ weight vector วิธีการหนึ่งที่ใช้ในการกำหนดค่า weight vector คือการสุ่มจากข้อมูล
3. ทำการคำนวณข้อมูลกับค่า weight vector เพื่อหาค่า winner neural โดยใช้ minimum-distance Euclidean เป็นดังสมการ2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\| x - m_c \| = \min \| x - m_i \| \qquad (2.2.1)$$

4. Winner neural คำนวณหาน้ำหนักใหม่ โดยใช้สมการ 2.2

$$m_i(t+1) = m_i(t) + \alpha(t)(x(t) - m_i(t)) \qquad (2.2.2)$$

โดย Neighborhood Function มักใช้สมการของ Gaussian ซึ่งเป็นดังสมการ 2.3

$$m_i(t+1) = m_i(t) + h_{ci}[x(t) - m_i(t)]$$

$$h_{ci} = \alpha(t) \exp\left(-\frac{\| r_i - r_c \|^2}{\sigma^2(t)} \right) \qquad (2.2.3)$$

ซึ่ง $0 < \alpha(t) < 1$ เป็นค่าแฟกเตอร์ หรือค่า Learning rate

5. เริ่มทำซ้ำตั้งแต่ข้อสองจนกระทั่งอัตราการเรียนรู้เท่ากับศูนย์หรือครบรอบการทำงานในแต่ละกลุ่ม

2.3.4 อัลกอริทึมในการพัฒนาระบบการแบ่งกลุ่มข้อมูลด้วยเซ็ลฟ์ออร์แกไนซิงแม็ป

1. Initialize random assign each a unique position in low-dimension grid (map)
2. Initialize α_0 and σ_0 to some values
3. **While** the number of iterations and user-specified limit **do**
4. Take random point x from the data set
5. Find the winner
   $\forall i, \| x(t) - m_c(t) \| \leq \| x(t) - m_i(t) \|$
6. Update the node in map
   $m_i(t+1) = m_i(t) + h_{c(x),i}(x(t) - m_i(t))$
7. Reduce the neighborhood radius and learning rate
8. **End While**

รูปที่ 2.12 อัลกอริทึมการทำงานของ SOM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

วิธีการดำเนินการศึกษา

ในบทนี้จะกล่าวถึงวิธีการพัฒนาโครงการ โดยที่จะแบ่งเป็นเครื่องมือที่ใช้ใน การพัฒนาโปรแกรม การทำงานของโปรแกรม ลักษณะอินพุทและเอาท์พุทของโปรแกรม ขั้นตอนการทำงานของโปรแกรมรูปแบบการทำงานและส่วนติดต่อผู้ใช้ของโปรแกรม

3.1 เครื่องมือที่ใช้ในการพัฒนาโปรแกรม

โครงการนี้เป็นการพัฒนาโปรแกรมประยุกต์ขึ้น เพื่อใช้ในการจัดกลุ่มข้อมูลโดยการเปรียบเทียบอัลกอริทึมที่ใช้ในการจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล (Partitioning Method) ได้แก่ K-means ,PAM (Partitioning around Medoids) และ ทฤษฎีโครงข่ายประสาทเทียม(Neural Network) ได้แก่ SOM (Self-organizing map) ภาษาคอมพิวเตอร์ที่ใช้การพัฒนาในโครงการนี้ใช้ Microsoft Visual Studio 2005

3.2 ลักษณะอินพุทและเอาท์พุทของโปรแกรม

3.2.1 ลักษณะอินพุทของโปรแกรม

อินพุทของโปรแกรมคือ ข้อมูลจำนวนหลายรูปแบบ โดยที่แต่ละรูปแบบต้องเป็นข้อมูลประเภทเดียวกันและมีลักษณะของข้อมูลเป็นตัวเลข ส่วนเอาท์พุทของโปรแกรมคือจำนวนกลุ่มของข้อมูลอินพุท แสดงกลุ่มของข้อมูลอินพุท ตัวอย่างลักษณะอินพุทของโปรแกรมแสดงดังรูป

รูปที่3.1 ลักษณะของอินพุทโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปสามารถอธิบายลักษณะของอินพุทได้คือ

- ในแต่ละบรรทัดคือ อินพุท 1 รูปแบบ
- อินพุทแต่ละรูปแบบจะมีคุณสมบัติต่างกันโดยการเว้นช่องว่างเพื่อแสดงคุณสมบัตินั้น
- ข้อมูลในแต่ละคุณสมบัติจะมีค่าเป็นตัวเลขเท่านั้น

3.2.2 ลักษณะเอาท์พุทของโปรแกรม

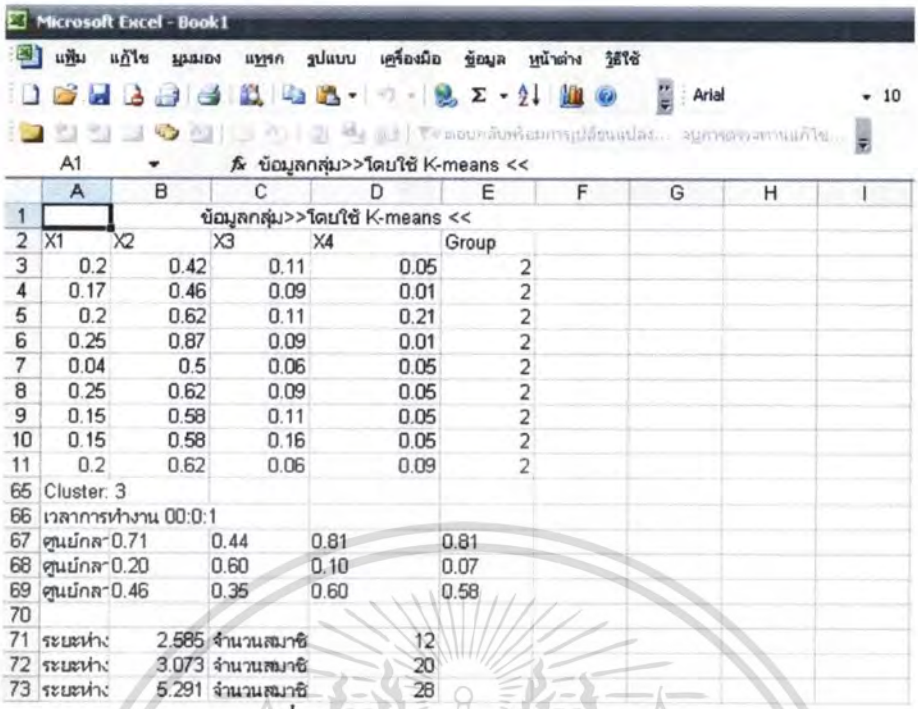
เอาท์พุทของโปรแกรมคือจำนวนกลุ่มของข้อมูลอินพุทที่ผู้ใช้ต้องการ แสดงกลุ่มของข้อมูลอินพุทข้อมูลเวลาการทำงานของอัลกอริทึมที่ใช้ในการจัดกลุ่ม ระยะห่างระหว่างข้อมูลกับจุดศูนย์กลาง การแสดงเอาท์พุทของโปรแกรมมี 2 ส่วนส่วนแรกแสดงในโปรแกรมและส่วนที่สองแสดงในโปรแกรม Excel เพื่อให้ผู้ใช้สามารถนำข้อมูลที่ได้ไปดำเนินการต่อได้ ตัวอย่างลักษณะเอาท์พุทของโปรแกรมแสดงดังรูป

รูปที่3.2 ลักษณะของเอาท์พุทโปรแกรม

จากรูปสามารถอธิบายลักษณะของเอาท์พุทได้คือ

- ข้อมูลในตาราง Results จะมีข้อมูลที่ใช้ในการจัดกลุ่ม และข้อมูลกลุ่มที่ได้
- ข้อมูลในตาราง Group จะมีข้อมูลจุดศูนย์กลางของกลุ่ม จำนวนสมาชิกในกลุ่มนั้นและข้อมูลระยะห่างของศูนย์กลางกับข้อมูลรวมของกลุ่มนั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



Microsoft Excel - Book1

แฟ้ม แก้ไข มุมมอง แทรก รูปแบบ เครื่องมือ ข้อมูล หน้าต่าง วิธีใช้

Arial 10

A1 ข้อมูลกลุ่ม>>โดยใช้ K-means <<

	A	B	C	D	E	F	G	H	I
1	ข้อมูลกลุ่ม>>โดยใช้ K-means <<								
2	X1	X2	X3	X4	Group				
3	0.2	0.42	0.11	0.05	2				
4	0.17	0.46	0.09	0.01	2				
5	0.2	0.62	0.11	0.21	2				
6	0.25	0.87	0.09	0.01	2				
7	0.04	0.5	0.06	0.05	2				
8	0.25	0.62	0.09	0.05	2				
9	0.15	0.58	0.11	0.05	2				
10	0.15	0.58	0.16	0.05	2				
11	0.2	0.62	0.06	0.09	2				
65	Cluster: 3								
66	เวลาการทำงาน 00:0:1								
67	ศูนย์กล	0.71	0.44	0.81	0.81				
68	ศูนย์กล	0.20	0.60	0.10	0.07				
69	ศูนย์กล	0.46	0.35	0.60	0.58				
70									
71	ระยะห่า	2.585	จำนวนสมาชิ	12					
72	ระยะห่า	3.073	จำนวนสมาชิ	20					
73	ระยะห่า	5.291	จำนวนสมาชิ	28					

รูปที่3.3 ลักษณะของเอาท์พุทโปรแกรม

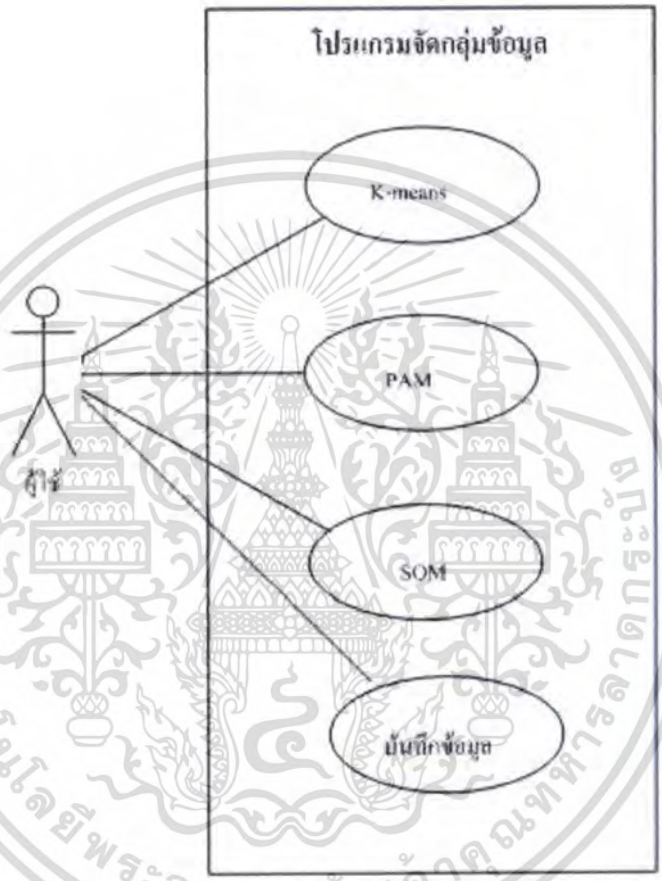
จากรูปสามารถอธิบายลักษณะของเอาท์พุทจากโปรแกรม excel ได้คือ

- วิธีที่ใช้ในการจัดกลุ่ม และข้อมูลกลุ่มที่ได้
- เวลาการทำงาน
- ข้อมูลจุดศูนย์กลางของกลุ่ม จำนวนสมาชิกในกลุ่มนั้นและข้อมูลระยะห่างของศูนย์กลางกับข้อมูลรวมของกลุ่มนั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 ขั้นตอนการทำงานของโปรแกรม

ขั้นตอนการทำงานของโปรแกรมจะมีการทำงานออกเป็น 4 ส่วนตามวิธีการจัดกลุ่มที่ผู้ใช้ทำการเลือกวิธีการจัดกลุ่มได้แก่ K- means , PAM ,SOM ในส่วนสุดท้ายเป็นการแสดงข้อมูลที่จัดกลุ่มได้ซึ่งจะแสดงการทำงานของโปรแกรมโดยรูปที่ 3.4 จะแสดง usecase diagram ของโปรแกรมดังต่อไปนี้

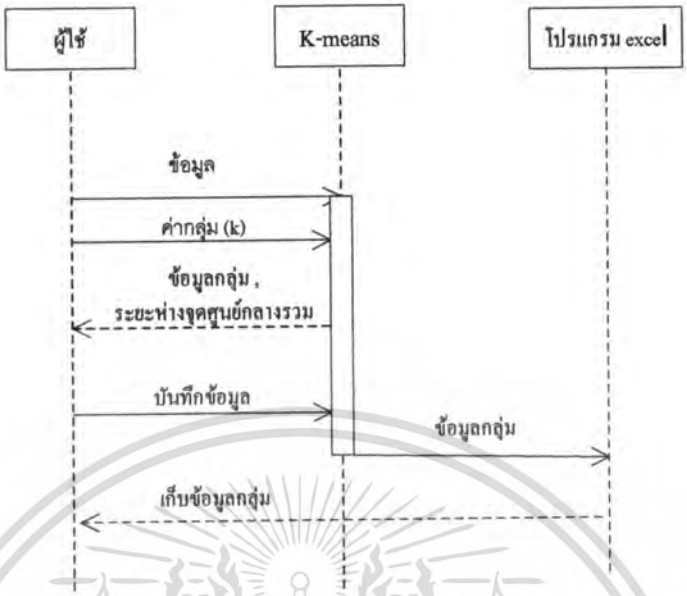


รูปที่3.4 แสดง use case diagram ของโปรแกรม

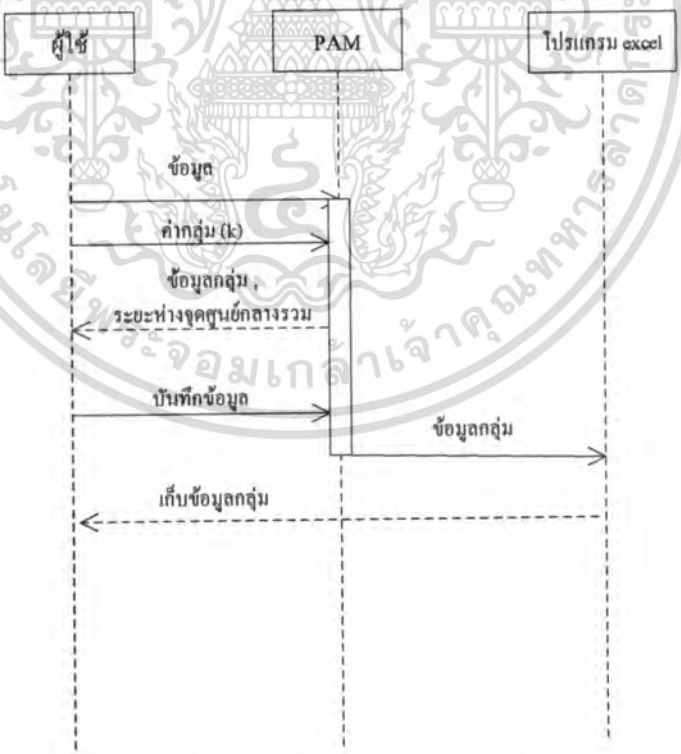
รูปที่ 3.5 ถึง 3.8 จะแสดง sequence diagram ทำให้เห็นการทำงานในขั้นตอนต่างๆของการทำงานของโปรแกรมโดยที่ รูปที่ 3.5 จะแสดงขั้นตอนทั้งหมดของโปรแกรม รูปที่ 3.6 แสดง sequence diagram ส่วนของ k-means รูปที่ 3.7 แสดง sequence diagram ส่วนของ PAM รูปที่ 3.8 แสดง sequence diagram ส่วนของ SOM

รูปที่3.5 แสดง sequence diagram ของโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

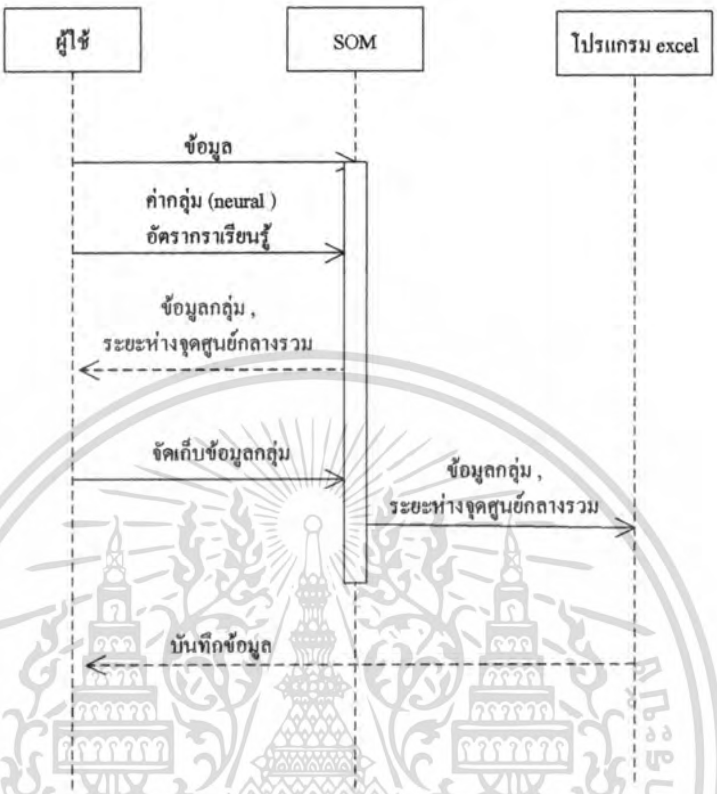


รูปที่3.6 แสดง sequence diagram ส่วนของ K- means



รูปที่3.7 แสดง sequence diagram ส่วนของ PAM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



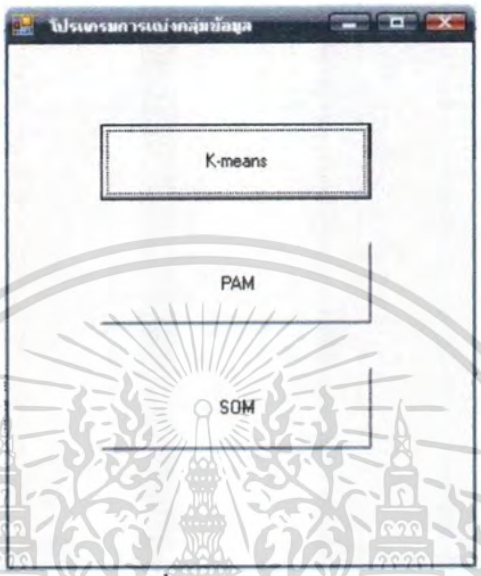
รูปที่3.8 แสดง sequence diagram ส่วนของ SOM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 ส่วนติดต่อผู้ใช้ของโปรแกรม

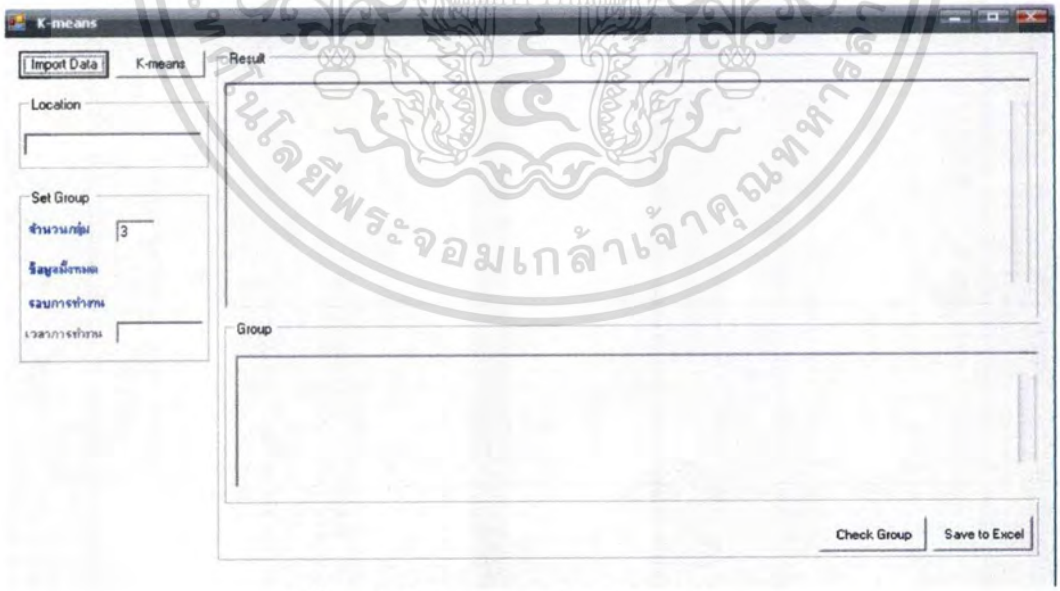
3.4.1 หน้าต่าง Menu

เพื่อเลือกว่าจะในการแบ่งกลุ่มแบบในในการทำงานโดยจะมีในเลือก 3 แบบคือ K- means, PAM, SOM



รูปที่3.9 หน้าต่างMenu

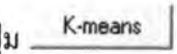
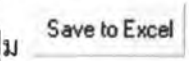
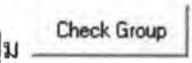
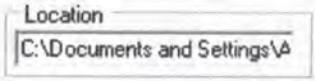
3.4.2 หน้าต่างหลักของโปรแกรมจัดกลุ่มโดยใช้ K-means



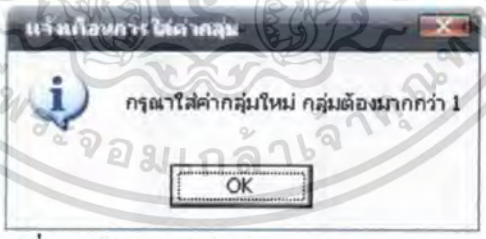
รูปที่3.10 หน้าต่างของการจัดกลุ่มโดย K- means

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในส่วนนี้จะเป็นการใช้ K-means ในการแบ่งกลุ่มข้อมูลโดยมีการทำงานดังนี้

- ปุ่ม จะเป็นปุ่มสำหรับการนำข้อมูลที่ต้องการแบ่งกลุ่มเข้ามาในโปรแกรม
- ปุ่ม  จะเป็นปุ่มสำหรับการแบ่งกลุ่มโดย K- means
- ปุ่ม  จะเป็นปุ่มสำหรับนำผลที่ได้ไปแสดงใน Microsoft office excel
- ปุ่ม  จะเป็นปุ่มสำหรับนำผลที่ได้และข้อมูลกลุ่มไปแสดงใน Microsoft office excel
-  จะเป็นข้อความสำหรับการนำข้อมูลที่ใช้ในการแบ่งกลุ่ม

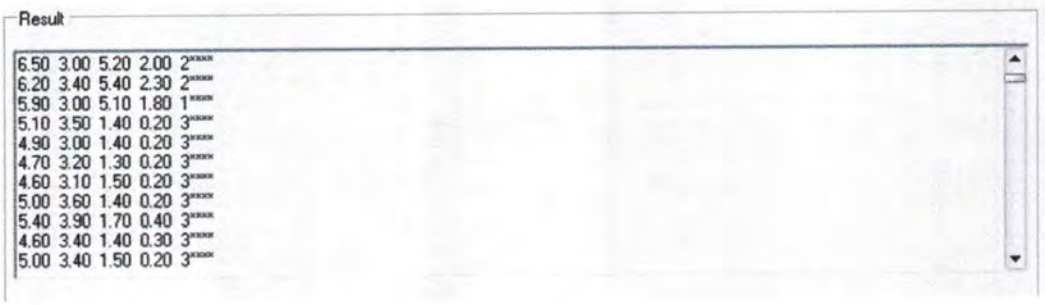
- หน้าต่างในการใส่ค่าการแบ่งกลุ่มโดยผู้ใช้งานจะใส่กลุ่มที่ต้องการแบ่งในช่อง insert cluster โดย ค่าที่ใส่นั้นต้องมากกว่า 1 กลุ่มขึ้นไป จำนวนข้อมูลทั้งหมดที่ใช้ในการคำนวณ และจำนวนรอบการทำงาน แสดงเวลาในการทำงาน



รูปที่3.11 ข้อความแจ้งเตือนการกำหนดค่ากลุ่ม

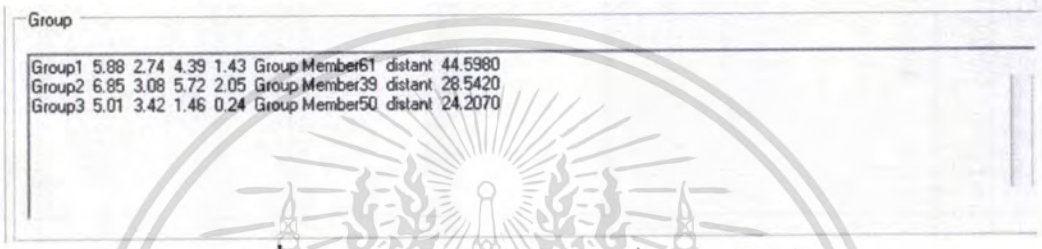
- ผู้ใช้งานจะใส่กลุ่มที่ต้องการแบ่งในช่อง insert cluster โดย ค่าที่ใส่นั้นต้องมากกว่า 1 กลุ่มขึ้นไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่3.12 ตัวอย่างการแสดงข้อมูลการจัดกลุ่ม

- หน้าต่างแสดงข้อมูลที่ต้องการแบ่งกลุ่ม และข้อมูลกลุ่มที่โปรแกรมคำนวณได้



รูปที่3.13 ตัวอย่างการแสดงข้อมูลศูนย์กลางการจัดกลุ่ม

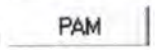
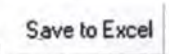
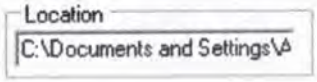
- หน้าต่างแสดงข้อมูลจุดศูนย์กลางของกลุ่มต่าง จำนวนสมาชิก และค่าระยะห่างระหว่างข้อมูลกับจุดศูนย์กลาง

3.4.3 หน้าต่างหลักของโปรแกรมจัดกลุ่มโดยใช้ PAM

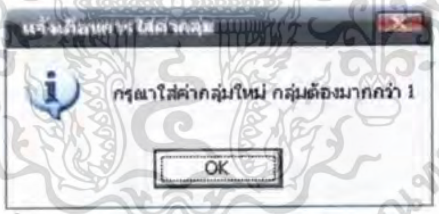
รูปที่3.14 หน้าต่างของการจัดกลุ่มโดย PAM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในส่วนนี้จะเป็นการใช้ PAM ในการแบ่งกลุ่มข้อมูลโดยมีการทำงานดังนี้

- ปุ่ม จะเป็นปุ่มสำหรับการนำข้อมูลที่ต้องการแบ่งกลุ่มเข้ามาในโปรแกรม
- ปุ่ม  จะเป็นปุ่มสำหรับการแบ่งกลุ่มโดย PAM
- ปุ่ม  จะเป็นปุ่มสำหรับนำผลที่ได้ไปแสดงใน Microsoft office excel
- ปุ่ม จะเป็นปุ่มสำหรับนำผลที่ได้และข้อมูลกลุ่มไปแสดงใน Microsoft office excel
-  จะเป็นข้อความสำหรับการนำข้อมูลที่ใช้ในการแบ่งกลุ่ม

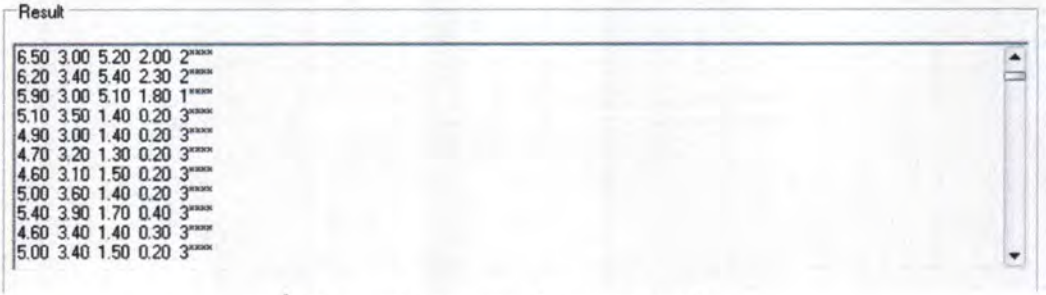
- หน้าต่างในการใส่ค่าการแบ่งกลุ่มโดยผู้ใช้งานจะใส่กลุ่มที่ต้องการแบ่งในช่อง insert cluster โดย ค่าที่ใส่นั้นต้องมากกว่า 1 กลุ่มขึ้นไป และจำนวนข้อมูลทั้งหมดที่ใช้ในการคำนวณ เวลาการทำงาน



รูปที่3.15 ข้อความแจ้งเตือนการกำหนดค่ากลุ่ม

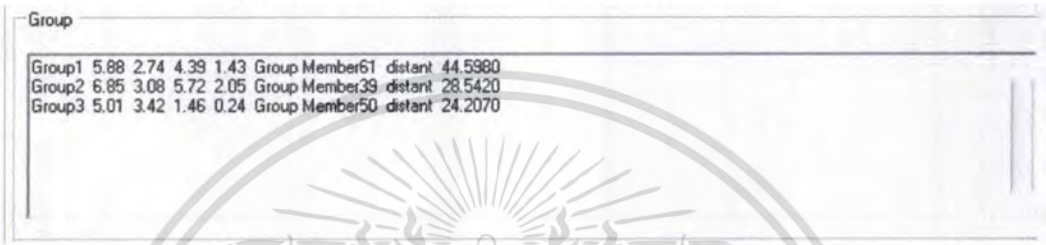
- ผู้ใช้งานจะใส่กลุ่มที่ต้องการแบ่งในช่อง insert cluster โดย ค่าที่ใส่นั้นต้องมากกว่า 1 กลุ่มขึ้นไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่3.16 ตัวอย่างการแสดงข้อมูลการจัดกลุ่ม

- หน้าต่างแสดงข้อมูลที่ต้องการแบ่งกลุ่ม และข้อมูลกลุ่มที่โปรแกรมคำนวณได้



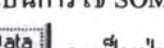
รูปที่3.17 ตัวอย่างการแสดงข้อมูลศูนย์กลางการจัดกลุ่ม

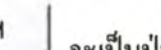
- หน้าต่างแสดงข้อมูลจุดศูนย์กลางของกลุ่มต่าง จำนวนสมาชิก และค่าระยะห่างระหว่างข้อมูลกับจุดศูนย์กลาง

3.4.4 หน้าต่างหลักของโปรแกรมจัดกลุ่มโดยใช้ SOM

รูปที่3.18 หน้าต่างของการจัดกลุ่มโดย SOM

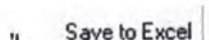
ในส่วนนี้จะเป็นการใช้ SOM ในการแบ่งกลุ่มข้อมูลโดยมีการทำงานดังนี้

- ปุ่ม 

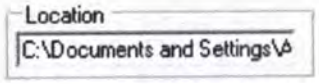
จะเป็นปุ่มสำหรับการนำข้อมูลที่ต้องการแบ่งกลุ่มเข้ามาในโปรแกรม
- ปุ่ม 

จะเป็นปุ่มสำหรับการแบ่งกลุ่มโดย SOM

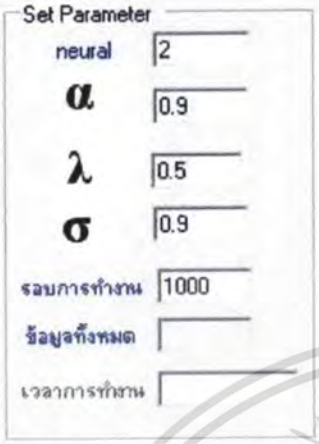
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ปุ่ม 

จะเป็นปุ่มสำหรับนำผลที่ได้ไปแสดงใน Microsoft office excel

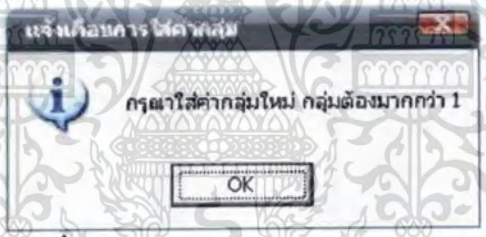
- 

จะเป็นข้อความสำหรับการนำข้อมูลที่ใช้ในการแบ่งกลุ่ม



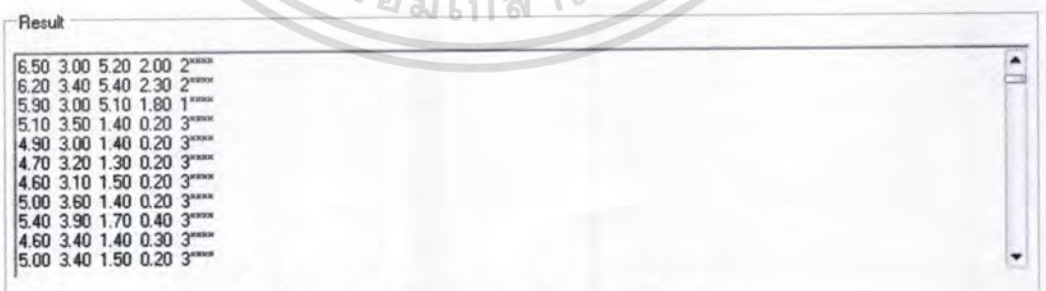
- หน้าต่างในการใส่ค่าพารามิเตอร์ที่ใช้ในการแบ่งกลุ่มผู้ใช้สามารถกำหนด neural อัตราการเรียนรู้(α) ระยะห่างในการปรับน้ำหนักการเรียนรู้ (σ) อัตราการเรียนรู้โหนดข้างเคียง (λ) และจำนวนรอบการทำงาน

-



รูปที่3.19 ข้อความแจ้งเตือนการกำหนดค่ากลุ่ม

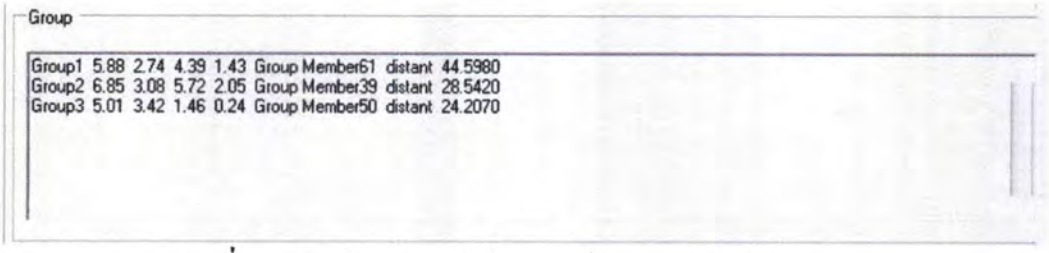
- ผู้ใช้งานจะใส่กลุ่มที่ต้องการแบ่งในช่อง insert cluster โดย ค่าที่ใส่นั้นต้องมากกว่า 1 กลุ่มขึ้นไป



รูปที่3.20 ตัวอย่างการแสดงข้อมูลการจัดกลุ่ม

- หน้าต่างแสดงข้อมูลที่ต้องการแบ่งกลุ่ม และข้อมูลกลุ่มที่โปรแกรมคำนวณได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

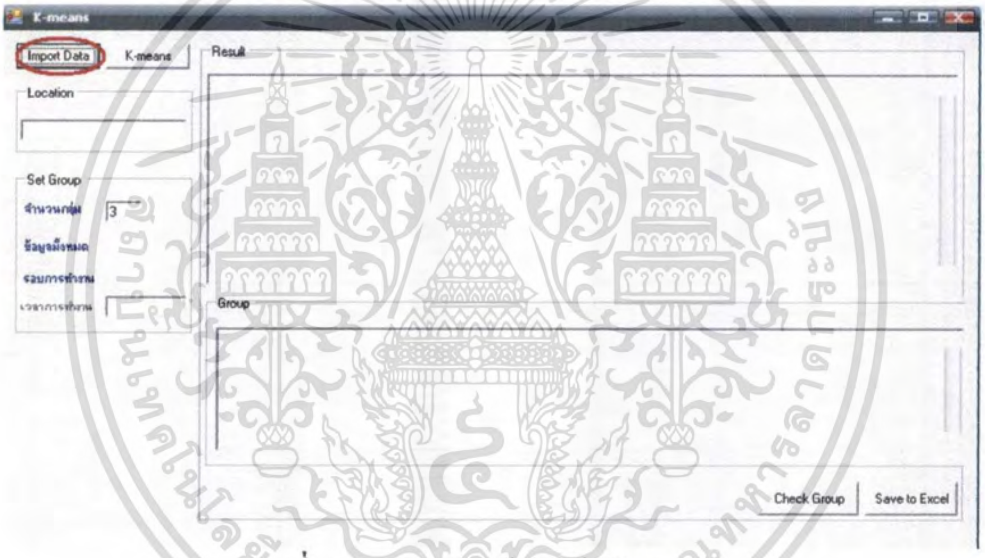


รูปที่3.21 ตัวอย่างการแสดงข้อมูลศูนย์กลางการจัดกลุ่ม

- หน้าต่างแสดงข้อมูลจุดศูนย์กลางของกลุ่มต่าง จำนวนสมาชิก และค่าระยะห่างระหว่างข้อมูลกับจุดศูนย์กลาง

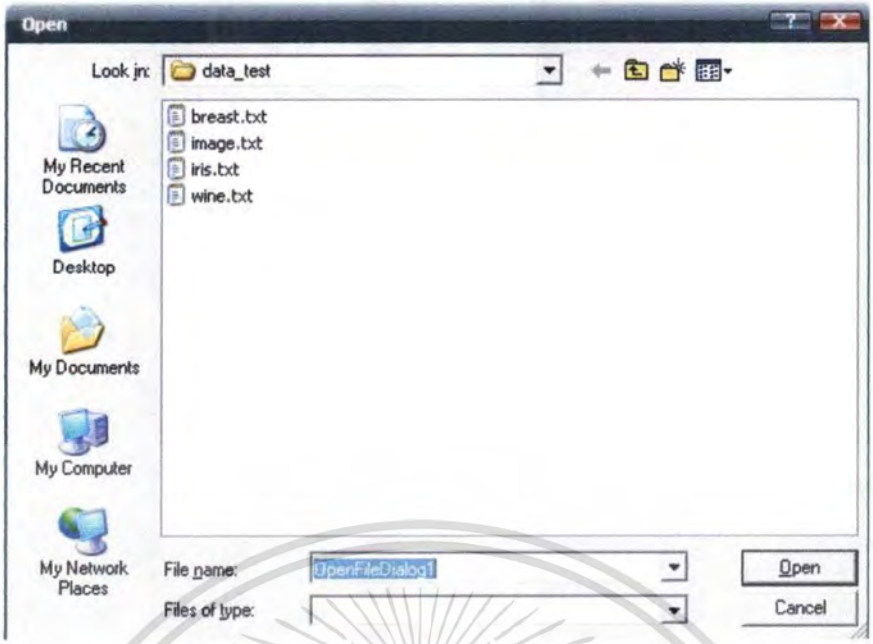
3.5 การใช้งานโปรแกรมจัดกลุ่มข้อมูล

3.5.1 การใช้งานโปรแกรมจัดกลุ่มโดยใช้ K- means



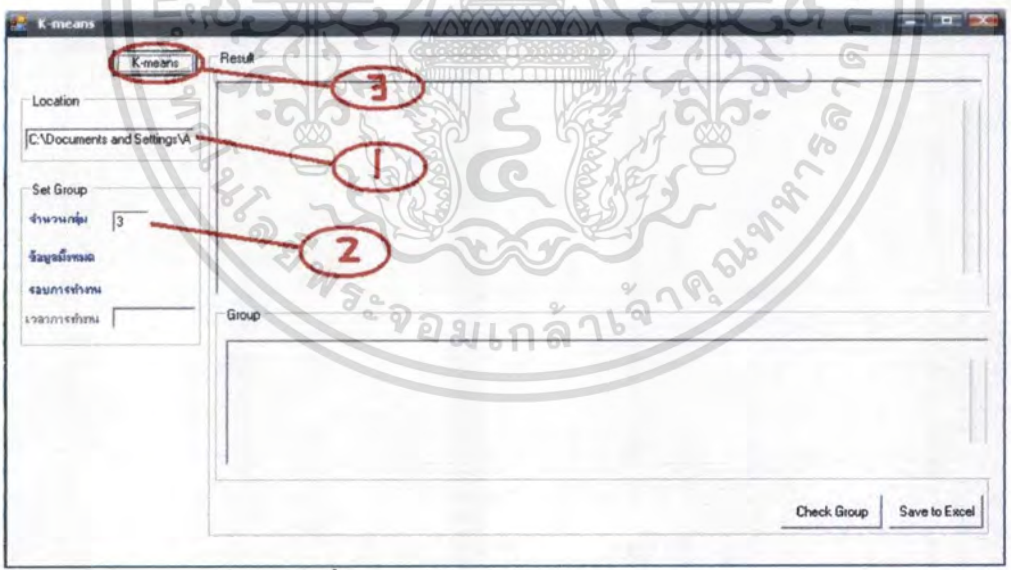
รูปที่3.22 แสดงการนำข้อมูลเข้าสู่โปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่3.23 แสดงการเลือกข้อมูลเข้าสู่โปรแกรม

การใช้งานโปรแกรมการจัดกลุ่มโดยใช้ K- means ในส่วนแรกเริ่มจากการนำข้อมูลที่ต้องการจัดกลุ่มเข้าสู่โปรแกรมโดยการกดปุ่ม import data จากนั้นเลือกข้อมูลที่ต้องการจัดกลุ่มดังรูป

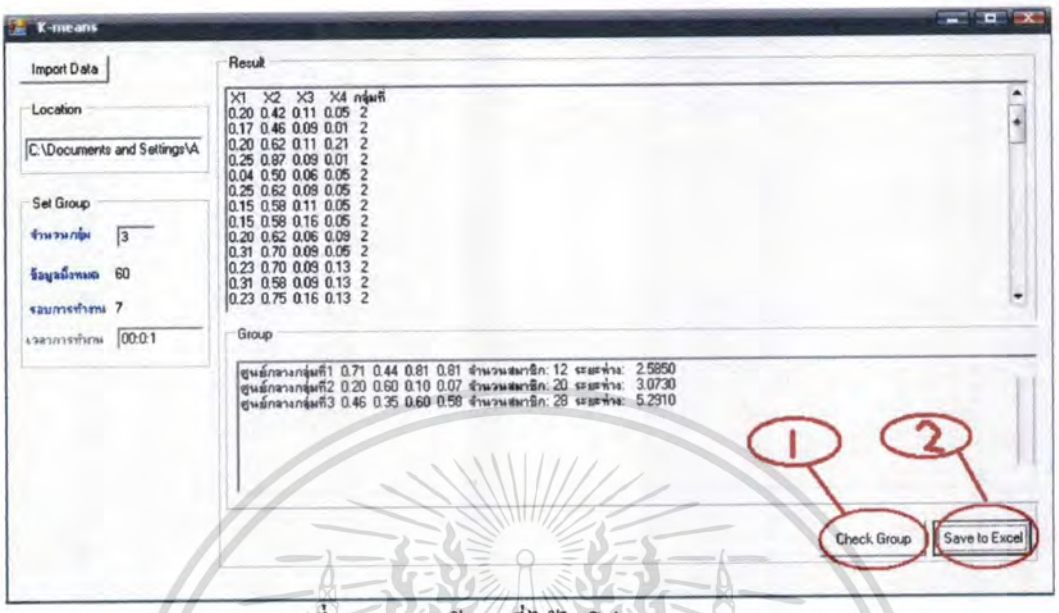


รูปที่3.24 แสดงการใช้งานของโปรแกรม

หลังจากที่นำข้อมูลเข้าสู่โปรแกรมแล้วจากรูปหมายเลข 1 จะแสดงว่าเรานำข้อมูลเข้าสู่โปรแกรมเสร็จแล้วจากนั้นผู้ใช้ต้องการจะแบ่งกลุ่มข้อมูลกี่กลุ่มสามารถจะกำหนดได้จากรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

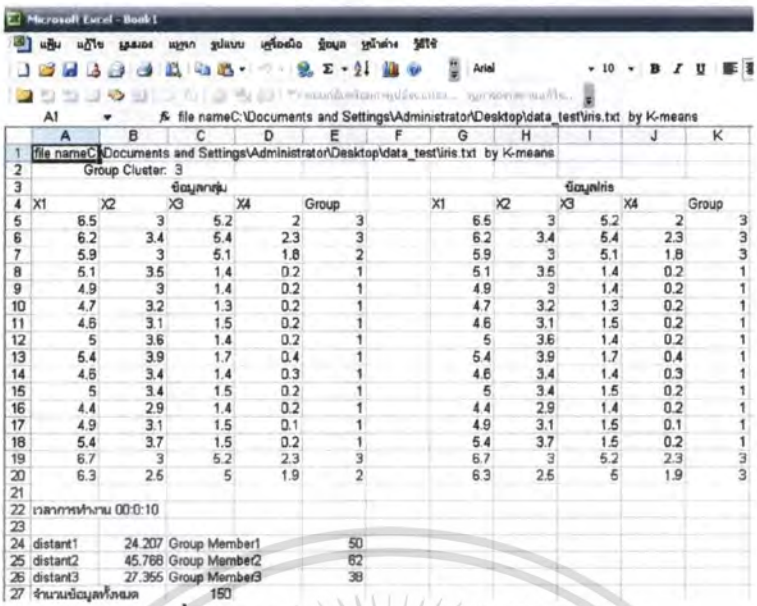
หมายเลข 2 แสดงการใส่ค่าการแบ่งกลุ่มที่ต้องการ จากนั้นกดปุ่ม K- means เพื่อทำการแบ่งกลุ่ม จากรูปหมายเลข 3



รูปที่3.25 แสดงข้อมูลที่ได้ในโปรแกรม excel

เมื่อทำการจัดกลุ่มเสร็จแล้วโปรแกรมจะแสดงข้อมูลที่จัดกลุ่มเรียบร้อยในส่วนของ results โดยจะแสดงข้อมูลที่ใช้ในการแบ่งกลุ่มและข้อมูลกลุ่มที่แบ่งได้ ในส่วนของ group จะแสดงข้อมูลศูนย์กลางของกลุ่มที่ใช้ในการจัดกลุ่ม จำนวนสมาชิกของกลุ่ม ระยะห่างรวมของข้อมูลกับศูนย์กลางกลุ่ม ในส่วนอื่นข้อหน้าต่างนี้จะแสดงข้อมูลทั้งหมดที่ใช้ในการแบ่งกลุ่ม จำนวนรอบการทำงาน ระยะเวลาการจัดกลุ่ม จากรูปหมายเลข 1 เป็นการทดสอบข้อมูลที่ใช้ในการเปรียบเทียบการทำงานโดยจะทำการเปรียบเทียบข้อมูลที่ทำการจัดกลุ่มในการทดสอบเปรียบเทียบข้อมูลเพื่อดูประสิทธิภาพที่ได้ ข้อมูลที่ได้จะแสดงในโปรแกรม excel จากรูปหมายเลข 2 จะทำการเก็บข้อมูลที่จัดกลุ่มจากโปรแกรมแสดงที่โปรแกรม excel

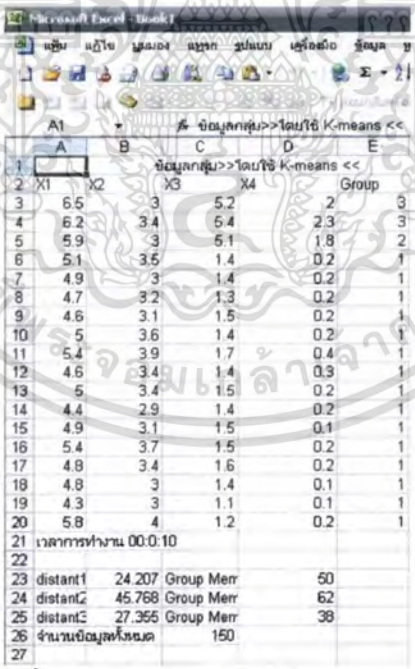
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



	A	B	C	D	E	F	G	H	I	J	K
1	file nameC:\Documents and Settings\Administrator\Desktop\data_test\iris.txt by K-means										
2	Group Cluster: 3										
3			ข้อมูลกลุ่ม						ข้อมูลIris		
4	X1	X2	X3	X4	Group		X1	X2	X3	X4	Group
5	6.5	3	5.2	2	3		6.5	3	5.2	2	3
6	6.2	3.4	5.4	2.3	3		6.2	3.4	5.4	2.3	3
7	5.9	3	5.1	1.8	2		5.9	3	5.1	1.8	3
8	5.1	3.5	1.4	0.2	1		5.1	3.5	1.4	0.2	1
9	4.9	3	1.4	0.2	1		4.9	3	1.4	0.2	1
10	4.7	3.2	1.3	0.2	1		4.7	3.2	1.3	0.2	1
11	4.6	3.1	1.5	0.2	1		4.6	3.1	1.5	0.2	1
12	5	3.6	1.4	0.2	1		5	3.6	1.4	0.2	1
13	5.4	3.9	1.7	0.4	1		5.4	3.9	1.7	0.4	1
14	4.6	3.4	1.4	0.3	1		4.6	3.4	1.4	0.3	1
15	5	3.4	1.5	0.2	1		5	3.4	1.5	0.2	1
16	4.4	2.9	1.4	0.2	1		4.4	2.9	1.4	0.2	1
17	4.9	3.1	1.5	0.1	1		4.9	3.1	1.5	0.1	1
18	5.4	3.7	1.5	0.2	1		5.4	3.7	1.5	0.2	1
19	6.7	3	5.2	2.3	3		6.7	3	5.2	2.3	3
20	6.3	2.5	5	1.9	2		6.3	2.5	5	1.9	3
21											
22	เวลาการทำงาน 00:0:10										
23											
24	distant1	24.207	Group Member1		50						
25	distant2	45.768	Group Member2		62						
26	distant3	27.355	Group Member3		38						
27	จำนวนข้อมูลทั้งหมด		150								

รูปที่3.26 การแสดงผลและข้อมูลเปรียบเทียบ

จากรูปข้อมูล Iris ที่ทำการจัดกลุ่มเสร็จแล้วมาเปรียบเทียบกับข้อมูลที่ถูกต้องเพื่อทำการทดสอบประสิทธิภาพในการจัดกลุ่ม



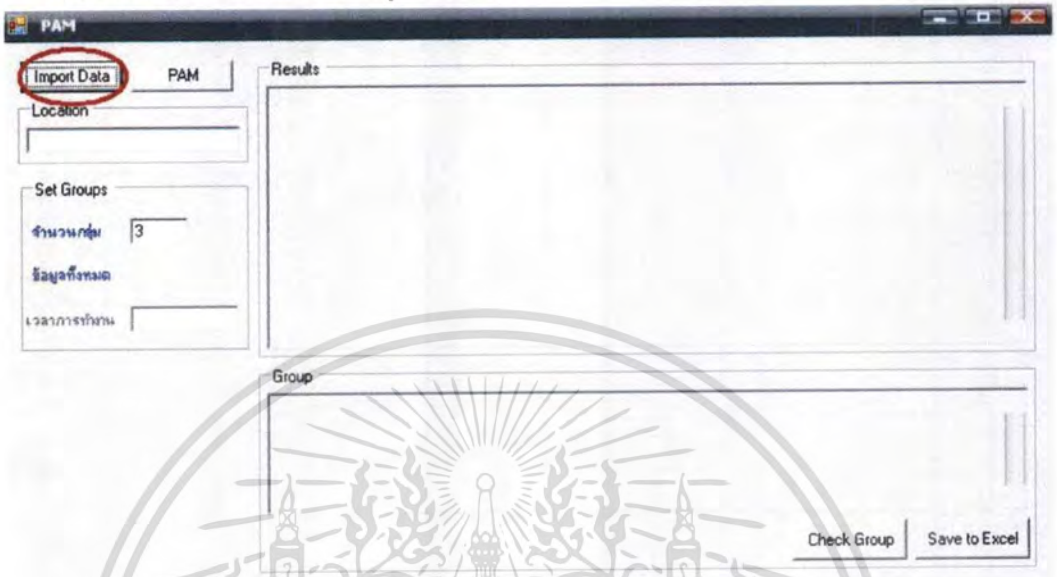
	A	B	C	D	E
1			ข้อมูลกลุ่ม>>โดยใช้ K-means <<		
2	X1	X2	X3	X4	Group
3	6.5	3	5.2	2	3
4	6.2	3.4	5.4	2.3	3
5	5.9	3	5.1	1.8	2
6	5.1	3.5	1.4	0.2	1
7	4.9	3	1.4	0.2	1
8	4.7	3.2	1.3	0.2	1
9	4.6	3.1	1.5	0.2	1
10	5	3.6	1.4	0.2	1
11	5.4	3.9	1.7	0.4	1
12	4.6	3.4	1.4	0.3	1
13	5	3.4	1.5	0.2	1
14	4.4	2.9	1.4	0.2	1
15	4.9	3.1	1.5	0.1	1
16	5.4	3.7	1.5	0.2	1
17	4.8	3.4	1.6	0.2	1
18	4.8	3	1.4	0.1	1
19	4.3	3	1.1	0.1	1
20	5.8	4	1.2	0.2	1
21	เวลาการทำงาน 00:0:10				
22					
23	distant1	24.207	Group Mem	50	
24	distant2	45.768	Group Mem	62	
25	distant3	27.355	Group Mem	38	
26	จำนวนข้อมูลทั้งหมด		150		
27					

รูปที่3.27 การแสดงผลในโปรแกรมexcel

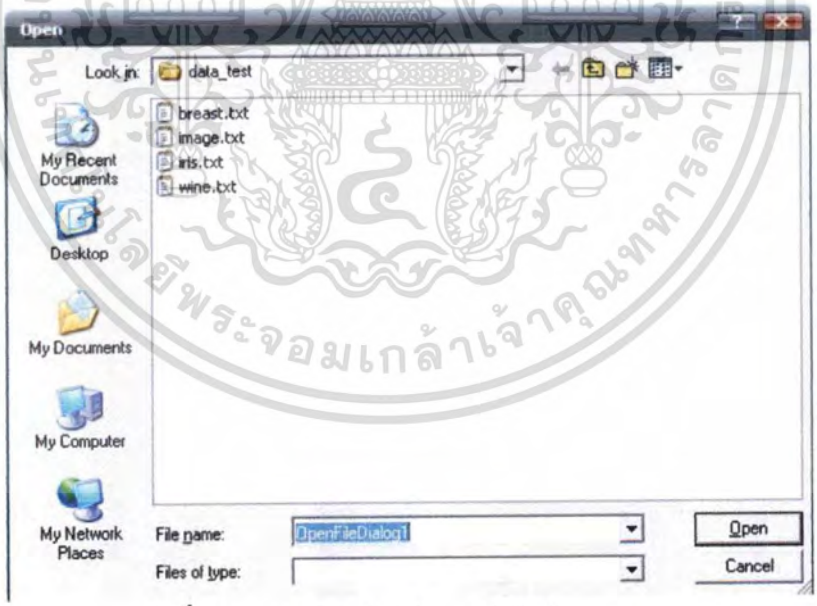
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปข้อมูลที่ทำการจัดกลุ่มเสร็จแล้วผู้ใช้สามารถจัดเก็บข้อมูลที่จัดกลุ่มเสร็จแล้วในโปรแกรม excel เพื่อใช้ในการวิเคราะห์ข้อมูลต่อหรือสามารถนำไปต่อยอดการใช้งานได้ต่อไป

3.5.2 การใช้งานโปรแกรมจัดกลุ่มโดยใช้ PAM



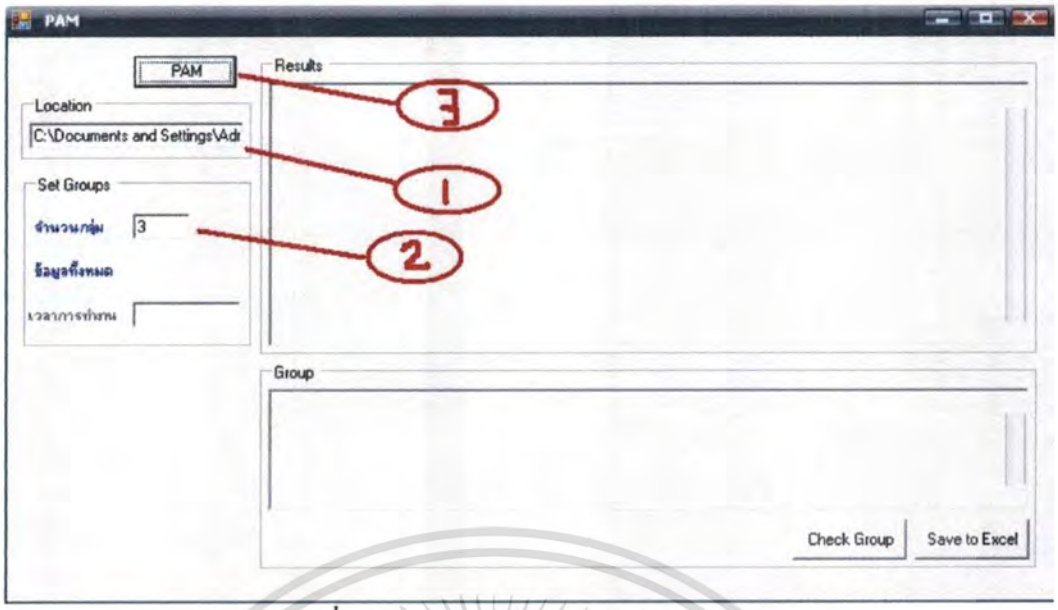
รูปที่3.28 แสดงการนำข้อมูลเข้าสู่โปรแกรม



รูปที่3.29 แสดงการเลือกข้อมูลเข้าสู่โปรแกรม

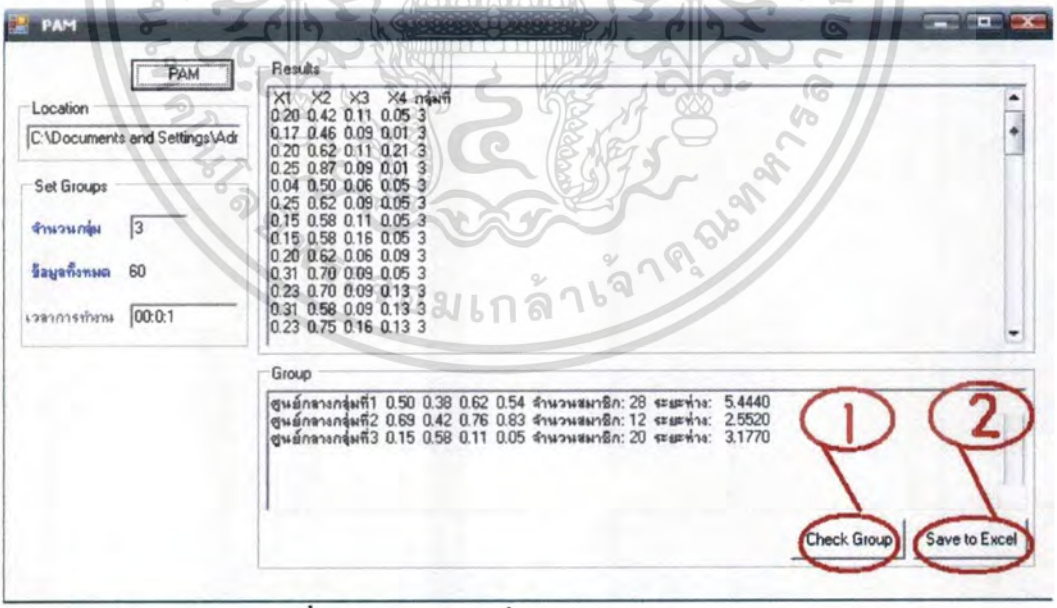
การใช้งานโปรแกรมการจัดกลุ่มโดยใช้ PAM ในส่วนแรกเริ่มจากการนำข้อมูลที่ต้องการจัดกลุ่มเข้าสู่โปรแกรมโดยการกดปุ่ม import data จากนั้นเลือกข้อมูลที่ต้องการจัดกลุ่มดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่3.30 แสดงการใช้งานของโปรแกรม

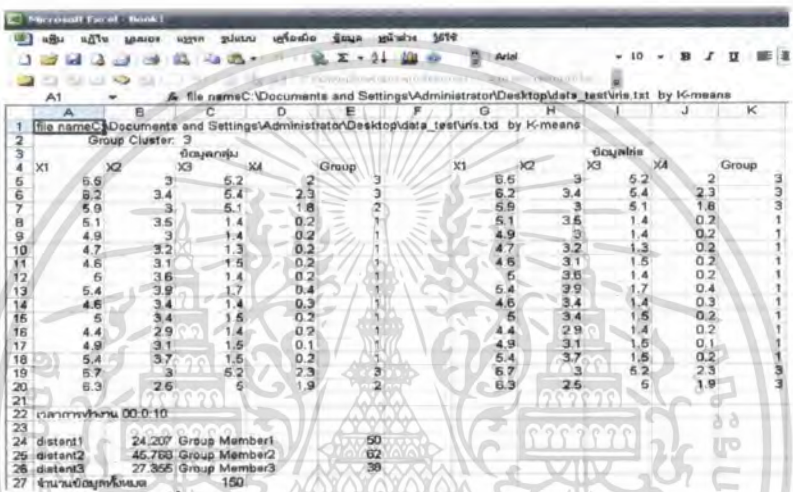
หลังจากที่นำข้อมูลเข้าสู่โปรแกรมแล้วจากรูปหมายเลข 1จะแสดงว่าเรานำข้อมูลเข้าสู่โปรแกรมเสร็จแล้วจากนั้นผู้ใช้ต้องการจะแบ่งกลุ่มข้อมูลกี่กลุ่มสามารถจะกำหนดได้จากรูปหมายเลข 2 แสดงการใส่ค่าการแบ่งกลุ่มที่ต้องการ จากนั้นกดปุ่ม PAM เพื่อทำการแบ่งกลุ่มจากรูปหมายเลข 3



รูปที่3.31 แสดงข้อมูลที่ได้ในโปรแกรม excel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

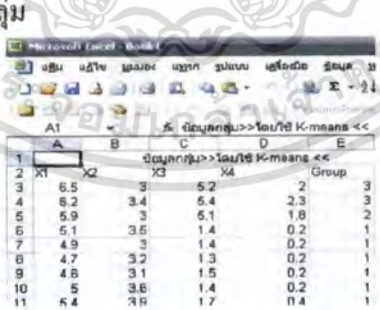
เมื่อทำการจัดกลุ่มเสร็จแล้วโปรแกรมจะแสดงข้อมูลที่จัดกลุ่มเรียบร้อยในส่วนของ results โดยจะแสดงข้อมูลที่ใช้ในการแบ่งกลุ่มและข้อมูลกลุ่มที่แบ่งได้ ในส่วนของ group จะแสดงข้อมูลศูนย์กลางของกลุ่มที่ใช้ในการจัดกลุ่ม จำนวนสมาชิกของกลุ่ม ระยะห่างรวมของข้อมูลกับศูนย์กลางกลุ่ม ในส่วนอื่นข้อหน้าต่างนี้จะแสดงข้อมูลทั้งหมดที่ใช้ในการแบ่งกลุ่ม ระยะเวลาการจัดกลุ่ม จากรูปหมายเลข 1 เป็นการทดสอบข้อมูลที่ใช้ในการเปรียบเทียบการทำงานโดยจะทำการเปรียบเทียบข้อมูลที่ทำการจัดกลุ่ม ในการทดสอบเปรียบเทียบข้อมูลเพื่อดูประสิทธิภาพที่ได้ ข้อมูลที่ได้จะแสดงในโปรแกรม excel จากรูปหมายเลข 2 จะทำการเก็บข้อมูลที่จัดกลุ่มจากโปรแกรมแสดงที่โปรแกรม excel



	A	B	C	D	E	F	G	H	I	J	K
1	file name:C:\Documents and Settings\Administrator\Desktop\data_test\iris.txt by K-means										
2	Group Cluster: 3										
3			ข้อมูลกลุ่ม						ข้อมูลIris		
4	X1	X2	X3	X4	Group		X1	X2	X3	X4	Group
5	6.5	3	5.2	2	3		6.5	3	5.2	2	3
6	6.2	3.4	5.4	2.3	3		6.2	3.4	5.4	2.3	3
7	5.9	3	5.1	1.8	2		5.9	3	5.1	1.8	3
8	5.1	3.5	1.4	0.2	1		5.1	3.5	1.4	0.2	1
9	4.9	3	1.4	0.2	1		4.9	3	1.4	0.2	1
10	4.7	3.2	1.3	0.2	1		4.7	3.2	1.3	0.2	1
11	4.6	3.1	1.5	0.2	1		4.6	3.1	1.5	0.2	1
12	5	3.6	1.4	0.2	1		5	3.6	1.4	0.2	1
13	5.4	3.9	1.7	0.4	1		5.4	3.9	1.7	0.4	1
14	4.6	3.4	1.4	0.3	1		4.6	3.4	1.4	0.3	1
15	5	3.4	1.5	0.2	1		5	3.4	1.5	0.2	1
16	4.4	2.9	1.4	0.2	1		4.4	2.9	1.4	0.2	1
17	4.9	3.1	1.5	0.1	1		4.9	3.1	1.5	0.1	1
18	5.4	3.7	1.5	0.2	1		5.4	3.7	1.5	0.2	1
19	6.7	3	5.2	2.3	3		6.7	3	5.2	2.3	3
20	6.3	2.5	5	1.9	2		6.3	2.5	5	1.9	3
21											
22	เวลาการทำงาน 00:0:10										
23											
24	distant1	24.207	Group Member1		50						
25	distant2	45.768	Group Member2		62						
26	distant3	27.355	Group Member3		38						
27	จำนวนข้อมูลทั้งหมด		150								

รูปที่3.32 การแสดงผลและข้อมูลเปรียบเทียบ

จากรูปข้อมูล Iris ที่ทำการจัดกลุ่มเสร็จแล้วมาเปรียบเทียบกับข้อมูลที่ถูกต้องเพื่อทำการทดสอบประสิทธิภาพในการจัดกลุ่ม

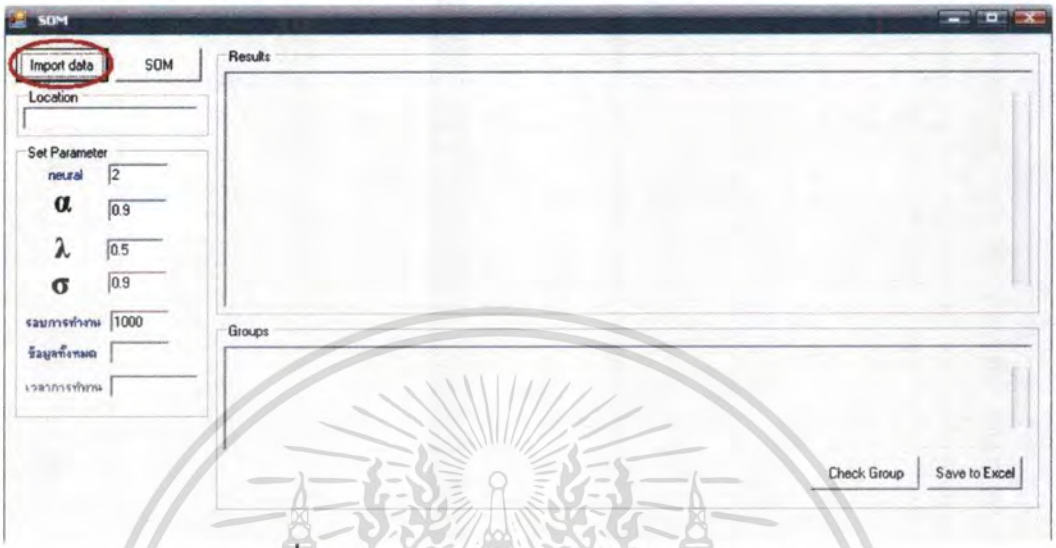


	A	B	C	D	E
1			ข้อมูลกลุ่ม>>โดยใช้ K-means <<		
2	X1	X2	X3	X4	Group
3	6.5	3	5.2	2	3
4	6.2	3.4	5.4	2.3	3
5	5.9	3	5.1	1.8	2
6	5.1	3.5	1.4	0.2	1
7	4.9	3	1.4	0.2	1
8	4.7	3.2	1.3	0.2	1
9	4.6	3.1	1.5	0.2	1
10	5	3.6	1.4	0.2	1
11	5.4	3.9	1.7	0.4	1

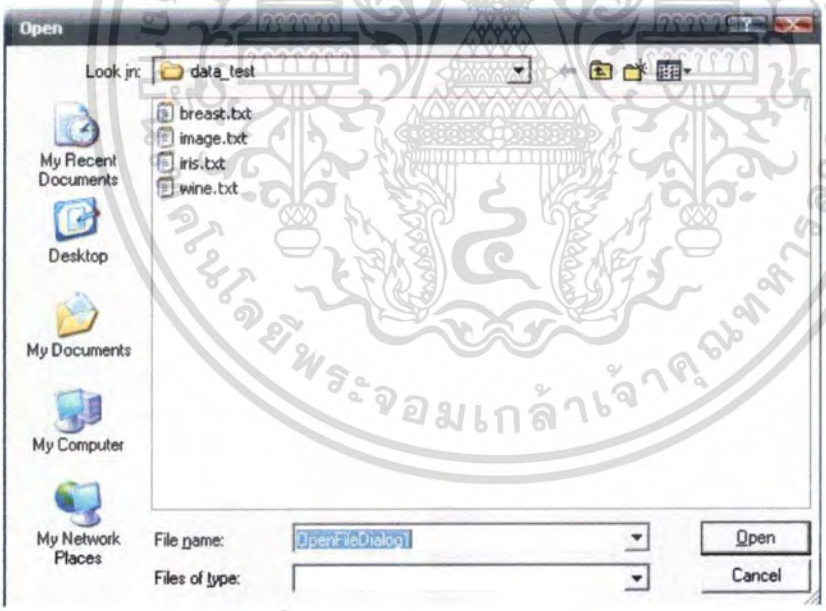
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปข้อมูลที่ทำการจัดกลุ่มเสร็จแล้วผู้ใช้สามารถจัดเก็บข้อมูลที่จัดกลุ่มเสร็จแล้วในโปรแกรม excel เพื่อใช้ในการวิเคราะห์ข้อมูลต่อหรือสามารถนำไปต่อยอดการใช้งานได้ต่อไป

3.5.3 การใช้งานโปรแกรมจัดกลุ่มโดยใช้ SOM



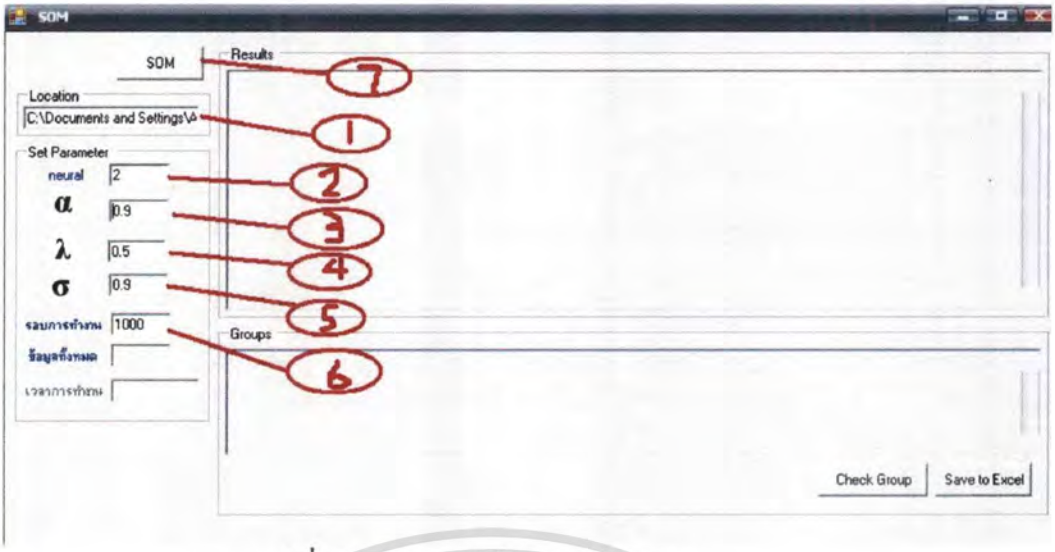
รูปที่3.34 แสดงการนำข้อมูลเข้าสู่โปรแกรม



รูปที่3.35 แสดงการเลือกข้อมูลเข้าสู่โปรแกรม

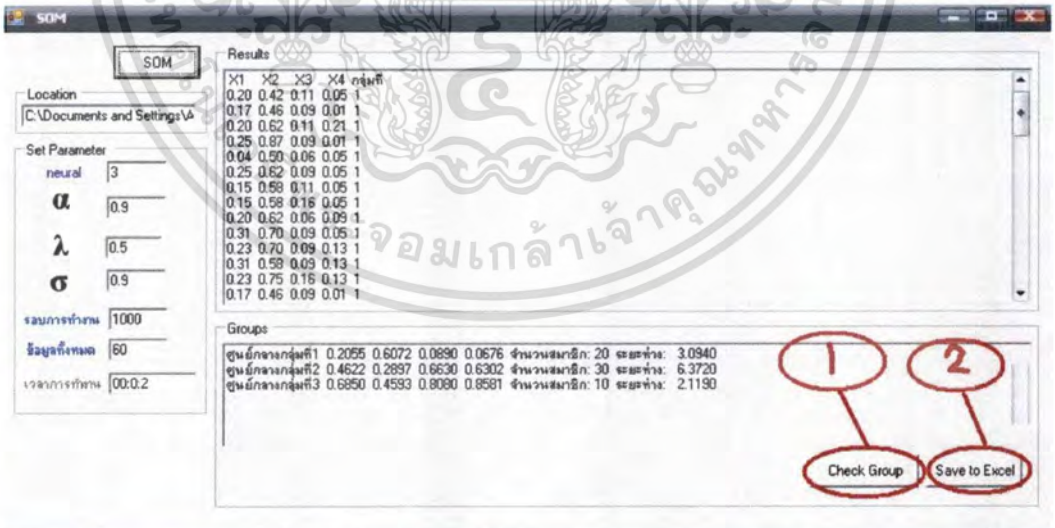
การใช้งานโปรแกรมการจัดกลุ่มโดยใช้ SOM ในส่วนแรกเริ่มจากการนำข้อมูลที่ต้องการจัดกลุ่มเข้าสู่โปรแกรมโดยการกดปุ่ม import data จากนั้นเลือกข้อมูลที่ต้องการจัดกลุ่มดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่3.36 แสดงการใช้งานของโปรแกรม

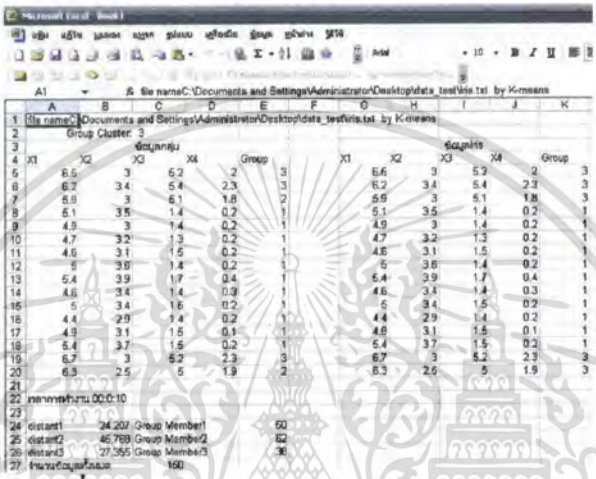
หลังจากที่นำข้อมูลเข้าสู่โปรแกรมแล้วจากรูปหมายเลข 1จะแสดงว่าเรานำข้อมูลเข้าสู่โปรแกรมเสร็จแล้วจากนั้นผู้ใช้ต้องการจะแบ่งกลุ่มข้อมูลกี่กลุ่มสามารถจะกำหนดได้จากรูปหมายเลข 2 แสดงการใส่ค่าการแบ่งกลุ่มที่ต้องการ ผู้ใช้สามารถกำหนดอัตราการเรียนรู้(α)จากรูปหมายเลข 3 ระยะห่างในการปรับน้ำหนักการเรียนรู้ (σ) จากรูปหมายเลข 4 อัตราการเรียนรู้โหนดข้างเคียง (λ) จากรูปหมายเลข 5 และจำนวนรอบการทำงานจากรูปหมายเลข 6 จากนั้นกดปุ่ม SOM เพื่อทำการแบ่งกลุ่มจากรูปหมายเลข 7



รูปที่3.37 แสดงข้อมูลที่ได้ในโปรแกรม excel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อทำการจัดกลุ่มเสร็จแล้วโปรแกรมจะแสดงข้อมูลที่จัดกลุ่มเรียบร้อยในส่วนของ results โดยจะแสดงข้อมูลที่ใช้ในการแบ่งกลุ่มและข้อมูลกลุ่มที่แบ่งได้ ในส่วนของ group จะแสดงข้อมูลศูนย์กลางของกลุ่มที่ใช้ในการจัดกลุ่ม จำนวนสมาชิกของกลุ่ม ระยะห่างรวมของข้อมูลกับศูนย์กลางกลุ่ม ในส่วนอื่นของหน้าต่างนี้จะแสดงข้อมูลทั้งหมดที่ใช้ในการแบ่งกลุ่ม ระยะเวลาการจัดกลุ่ม จากรูปหมายเลข 1 เป็นการทดสอบข้อมูลที่ใช้ในการเปรียบเทียบการทำงาน โดยจะทำการเปรียบเทียบข้อมูลที่ทำการจัดกลุ่มในการทดสอบเปรียบเทียบข้อมูลเพื่อดูประสิทธิภาพที่ได้ ข้อมูลที่ได้จะแสดงในโปรแกรม excel จากรูปหมายเลข 2 จะทำการเก็บข้อมูลที่จัดกลุ่มจากโปรแกรมแสดงที่โปรแกรม excel



รูปที่3.38 การแสดงผลและข้อมูลเปรียบเทียบ

จากรูปข้อมูล Iris ที่ทำการจัดกลุ่มเสร็จแล้วมาเปรียบเทียบกับข้อมูลที่ถูกต้องเพื่อทำการทดสอบประสิทธิภาพในการจัดกลุ่ม

	A	B	C	D	E
1			ข้อมูลกลุ่ม>>โดยใช้ K-means <<		
2	X1	X2	X3	X4	Group
3	6.5	3	5.2	2	3
4	6.2	3.4	5.4	2.3	3
5	5.9	3	5.1	1.8	2
6	5.1	3.5	1.4	0.2	1
7	4.9	3	1.4	0.2	1
8	4.7	3.2	1.3	0.2	1
9	4.6	3.1	1.5	0.2	1
10	5	3.6	1.4	0.2	1
11	5.4	3.9	1.7	0.4	1
12	4.6	3.4	1.4	0.3	1
13	5	3.4	1.5	0.2	1
14	4.4	2.9	1.4	0.2	1
15	4.9	3.1	1.5	0.1	1
16	5.4	3.7	1.5	0.2	1
17	4.8	3.4	1.6	0.2	1
18	4.8	3	1.4	0.1	1
19	4.3	3	1.1	0.1	1
20	5.8	4	1.2	0.2	1
21	เวลาการทำงาน 00:0:10				
22					
23	distant1	24.207	Group Mem	50	
24	distant2	45.768	Group Mem	62	
25	distant3	27.355	Group Mem	38	
26	จำนวนข้อมูลทั้งหมด		150		
27					

รูปที่3.39 การแสดงผลในโปรแกรม excel

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปข้อมูลที่ทำการจัดกลุ่มเสร็จแล้วผู้ใช้สามารถจัดเก็บข้อมูลที่จัดกลุ่มเสร็จแล้วในโปรแกรม excel เพื่อใช้ในการวิเคราะห์ข้อมูลต่อหรือสามารถนำไปต่อยอดการใช้งานได้ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การวิเคราะห์ผลการทดลอง

ในบทนี้จะกล่าวถึงการแสดงผลการจัดกลุ่มที่ได้จากโปรแกรมและวิเคราะห์เปรียบเทียบผลการจัดกลุ่มของอัลกอริทึมทั้ง 3 โดยใช้ข้อมูล 4 ข้อมูลในการทดสอบได้แก่ Iris, Diabetes, Breast Cancer, Wine

4.1 ข้อมูลที่ใช้กับโปรแกรม

ข้อมูลที่นำมาใช้ในการทดสอบกับโปรแกรม จะถูกจัดเก็บอยู่ในรูปของ Text File

4.1.1 Iris Data

ข้อมูลดอกไม้ Iris มีลักษณะข้อมูลทั้งหมด 4 ลักษณะที่แตกต่างกัน คือ

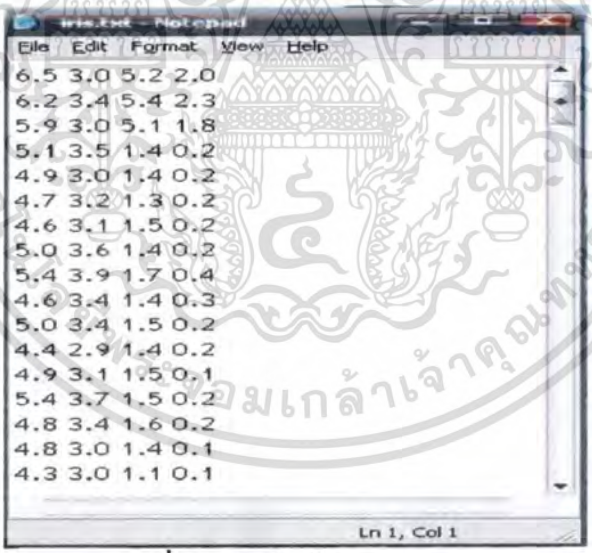
-ความยาวกลีบเลี้ยงดอกไม้

-ค่าความกว้างกลีบเลี้ยงดอกไม้

-ค่าความยาวของกลีบดอกไม้

-ค่าความกว้างของกลีบดอกไม้

โดยชนิดของดอกไม้ Iris จะมีทั้งหมด 3 ชนิด โดยมีจำนวนข้อมูลทั้งหมด 150 ข้อมูล

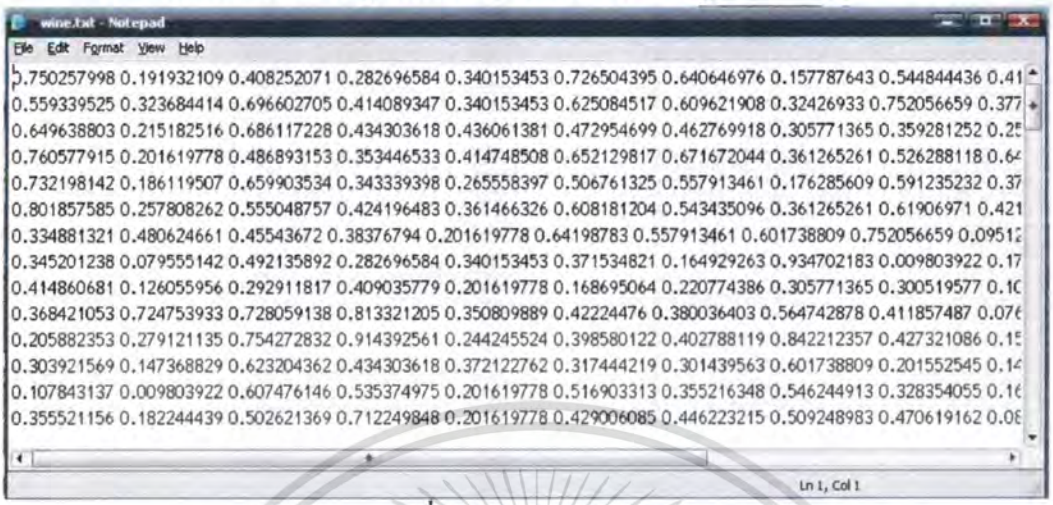


รูปที่ 4.1 ตัวอย่างข้อมูล Iris Data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 Wine Data

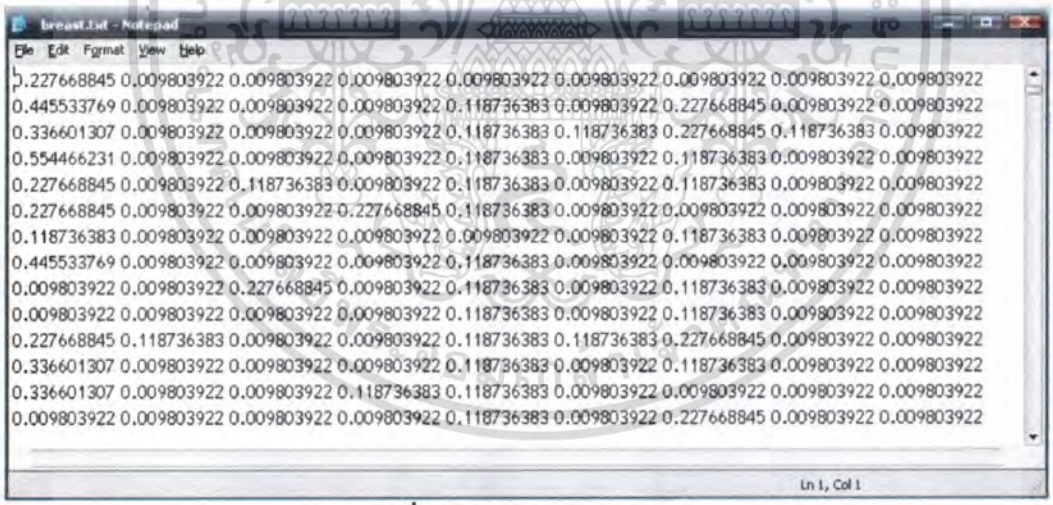
Wine Data มีลักษณะข้อมูลทั้งหมด 13 ลักษณะ 178 ข้อมูล



รูปที่ 4.2 ข้อมูล Wine Data

4.1.3 Diabetes Data

Diabetes Data มีลักษณะข้อมูลทั้งหมด 8 ลักษณะ 768 ข้อมูล

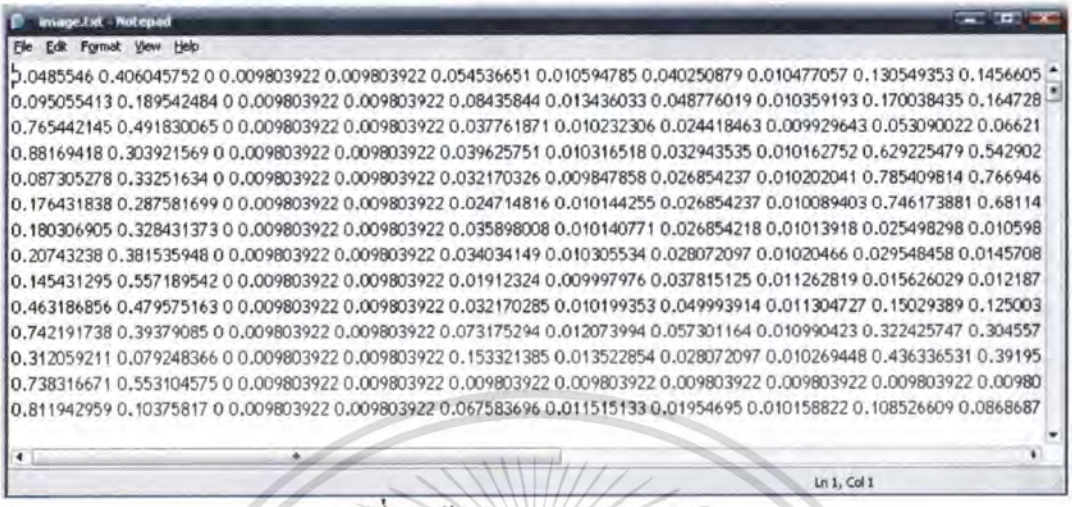


รูปที่ 4.3 ข้อมูล Diabetes Data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.4 Breast Cancer Data

Breast Cancer Data มีลักษณะข้อมูลทั้งหมด 9 ลักษณะ 699 ข้อมูล



รูปที่ 4.4 ข้อมูล Breast Cancer Data

4.2 ทดสอบค่าที่ใช้ในการจัดกลุ่มของ SOM

ในการทดลองจัดกลุ่มข้อมูล โดยใช้อัลกอริทึม 3 แบบโดย 2 แบบแรก K-means , PAM มีการกำหนดค่ากลุ่มและทำการเก็บค่าที่ได้ จำนวน 5 ครั้งเพื่อใช้ในการทดสอบประสิทธิภาพของการจัดกลุ่ม แบบสุดท้าย SOM จะทำการทดสอบโดยการทดลองเปรียบเทียบอัตราการเรียนรู้ (α) ระยะห่างโหนดข้างเคียง (λ) และค่าซิกม่า(σ) โดยจะทำการเปรียบเทียบผลที่ได้ใกล้เคียงกับ 2 แบบแรกเพื่อทำการวัดประสิทธิภาพการจัดกลุ่มโดยกำหนด ในการทดลองเราจะทำการหาค่าที่ใช้ในการจัดกลุ่มของ SOM โดยค่าที่จะนำมาใช้จะอยู่ระหว่าง 0 – 1 โดยที่เรา จะกำหนดค่า 3 ตัวเพื่อใช้ในการทดสอบ ได้แก่ 0.9 , 0.5 , 0.1 เราจะทำการสลับ ค่าที่ 3 ค่า ทำการทดสอบ 3 ครั้งจากนั้นเลือกค่าที่สามารถจัดกลุ่มได้ดีมาทดสอบกับอัลกอริทึมที่ใช้ในการจัดกลุ่มทั้ง 2 วิธี ในการทดสอบประสิทธิภาพ โดยตารางจะแสดงค่าที่ทำการทดสอบอัตราการเรียนรู้ (α) ระยะห่างโหนดข้างเคียง (λ) และค่าซิกม่า(σ)

4.2.1 ทดสอบค่าในข้อมูล Iris Data

ข้อมูลที่เรานำมาทดสอบนี้ข้อมูลทั้งหมดมี 150 ชุด กลุ่มข้อมูลทั้งหมด 3 กลุ่ม ทำการเลือกข้อมูลมาทดสอบกลุ่มละ 30 ชุด ข้อมูลที่ใช้ทดสอบทั้งหมด 90 ชุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ผลการทดสอบค่า Iris Data

α , λ, σ	ระยะห่างจุดศูนย์กลางรวมโดยเฉลี่ย	จัดกลุ่มถูกต้องโดยเฉลี่ย	จัดกลุ่มไม่ถูกต้องโดยเฉลี่ย
0.9 , 0.9 , 0.9	12.213	50	10
0.9 , 0.9 , 0.5	13.346	44	16
0.9 , 0.9 , 0.1	13.346	44	16
0.9 , 0.5 , 0.9	11.585	50	10
0.9 , 0.5 , 0.5	12.473	50	10
0.9 , 0.5 , 0.1	13.539	44	16
0.9 , 0.1 , 0.9	11.007	52	8
0.9 , 0.1 , 0.5	11.772	52	8
0.9 , 0.1 , 0.1	35.548	52	8
0.5 , 0.9 , 0.9	11.585	50	10
0.5 , 0.9 , 0.5	12.473	50	10
0.5 , 0.9 , 0.1	13.539	44	16
0.5 , 0.5 , 0.9	11.219	51	9
0.5 , 0.5 , 0.5	11.221	51	9
0.5 , 0.5 , 0.1	11.221	51	9
0.5 , 0.1 , 0.9	10.973	52	8
0.5 , 0.1 , 0.5	11.767	52	8
0.5 , 0.1 , 0.1	12.560	52	8
0.1 , 0.9 , 0.9	11.007	52	8
0.1 , 0.9 , 0.5	11.557	52	8
0.1 , 0.9 , 0.1	11.849	52	8
0.1 , 0.5 , 0.9	10.973	52	8
0.1 , 0.5 , 0.5	11.766	52	8
0.1 , 0.5 , 0.1	11.766	52	8
0.1 , 0.1 , 0.9	10.955	52	8
0.1 , 0.1 , 0.5	11.717	52	8
0.1 , 0.1 , 0.1	11.005	52	8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 ทดสอบค่าในข้อมูล Diabetes Data

ข้อมูลที่เรานำมาทดสอบนี้ข้อมูลทั้งหมดมี 768 ชุด กลุ่มข้อมูลทั้งหมด 2 กลุ่ม ทำการเลือกข้อมูลมาทดสอบกลุ่มละ50 ชุด ข้อมูลที่ใช้ทดสอบทั้งหมด 100 ชุด

ตารางที่ 4.2 ผลการทดสอบค่า Diabetes Data

α , λ, σ	ระยะห่างจุดศูนย์กลางรวมโดยเฉลี่ย	จัดกลุ่มถูกต้องโดยเฉลี่ย	จัดกลุ่มไม่ถูกต้องโดยเฉลี่ย
0.9 , 0.9 , 0.9	44.034	78	22
0.9 , 0.9 , 0.5	44.035	78	22
0.9 , 0.9 , 0.1	44.035	78	22
0.9 , 0.5 , 0.9	40.303	78	22
0.9 , 0.5 , 0.5	40.302	78	22
0.9 , 0.5 , 0.1	40.302	78	22
0.9 , 0.1 , 0.9	38.206	78	22
0.9 , 0.1 , 0.5	38.206	78	22
0.9 , 0.1 , 0.1	38.206	78	22
0.5 , 0.9 , 0.9	40.303	78	22
0.5 , 0.9 , 0.5	40.302	78	22
0.5 , 0.9 , 0.1	40.302	78	22
0.5 , 0.5 , 0.9	39.116	80	20
0.5 , 0.5 , 0.5	39.115	80	20
0.5 , 0.5 , 0.1	43.509	80	20
0.5 , 0.1 , 0.9	37.951	82	18
0.5 , 0.1 , 0.5	39.383	82	18
0.5 , 0.1 , 0.1	38.266	82	18
0.1 , 0.9 , 0.9	38.206	82	18
0.1 , 0.9 , 0.5	39.784	82	18
0.1 , 0.9 , 0.1	41.362	82	18
0.1 , 0.5 , 0.9	37.951	82	18
0.1 , 0.5 , 0.5	39.383	82	18
0.1 , 0.5 , 0.1	40.815	82	18
0.1 , 0.1 , 0.9	37.675	82	18
0.1 , 0.1 , 0.5	37.698	82	18
0.1 , 0.1 , 0.1	40.967	82	18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.3 ทดสอบค่าในข้อมูล Wine Data

ข้อมูลที่เรานำมาทดสอบนี้ข้อมูลทั้งหมดมี 178 ชุด กลุ่มข้อมูลทั้งหมด 3 กลุ่ม ทำการเลือกข้อมูลมาทดสอบกลุ่มละ30 ชุด ข้อมูลที่ใช้ทดสอบทั้งหมด 90 ชุด

ตารางที่ 4.3 ผลการทดสอบค่า Wine Data

α , λ, σ	ระยะห่างจุดศูนย์กลางรวมโดยเฉลี่ย	จัดกลุ่มถูกต้องโดยเฉลี่ย	จัดกลุ่มไม่ถูกต้องโดยเฉลี่ย
0.9 , 0.9 , 0.9	47.507	76	14
0.9 , 0.9 , 0.5	47.507	76	14
0.9 , 0.9 , 0.1	47.507	67	23
0.9 , 0.5 , 0.9	45.688	78	12
0.9 , 0.5 , 0.5	45.691	80	10
0.9 , 0.5 , 0.1	47.979	79	11
0.9 , 0.1 , 0.9	45.688	80	10
0.9 , 0.1 , 0.5	45.691	81	9
0.9 , 0.1 , 0.1	47.979	81	9
0.5 , 0.9 , 0.9	43.965	81	9
0.5 , 0.9 , 0.5	44.583	80	10
0.5 , 0.9 , 0.1	48.275	80	10
0.5 , 0.5 , 0.9	45.688	80	10
0.5 , 0.5 , 0.5	45.691	80	10
0.5 , 0.5 , 0.1	47.986	80	10
0.5 , 0.1 , 0.9	43.906	81	9
0.5 , 0.1 , 0.5	44.321	81	9
0.5 , 0.1 , 0.1	48.080	81	9
0.1 , 0.9 , 0.9	43.965	81	9
0.1 , 0.9 , 0.5	44.25	81	9
0.1 , 0.9 , 0.1	48.485	81	9
0.1 , 0.5 , 0.9	43.906	81	9
0.1 , 0.5 , 0.5	44.321	81	9
0.1 , 0.5 , 0.1	48.080	81	9
0.1 , 0.1 , 0.9	44.129	81	9
0.1 , 0.1 , 0.5	47.440	81	9
0.1 , 0.1 , 0.1	47.954	81	9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.4 ทดสอบค่าในข้อมูล Breast Cancer Data

ข้อมูลที่เรานำมาทดสอบนี้ข้อมูลทั้งหมดมีจำนวน 699 ชุด กลุ่มข้อมูลทั้งหมด 2 กลุ่ม ทำการเลือกข้อมูลมาทดสอบกลุ่มละ30 ชุด ข้อมูลที่ใช้ทดสอบทั้งหมด 60 ชุด

ตารางที่ 4.4 ผลการทดสอบค่า Breast Cancer Data

α , λ, σ	ระยะห่างจุดศูนย์กลางรวมโดยเฉลี่ย	จัดกลุ่มถูกต้องโดยเฉลี่ย	จัดกลุ่มไม่ถูกต้องโดยเฉลี่ย
0.9 , 0.9 , 0.9	41.596	55	5
0.9 , 0.9 , 0.5	41.585	55	5
0.9 , 0.9 , 0.1	41.585	55	5
0.9 , 0.5 , 0.9	39.332	55	5
0.9 , 0.5 , 0.5	37.325	55	5
0.9 , 0.5 , 0.1	37.325	55	5
0.9 , 0.1 , 0.9	34.425	55	5
0.9 , 0.1 , 0.5	34.420	55	5
0.9 , 0.1 , 0.1	34.420	55	5
0.5 , 0.9 , 0.9	37.332	55	5
0.5 , 0.9 , 0.5	37.325	55	5
0.5 , 0.9 , 0.1	37.325	55	5
0.5 , 0.5 , 0.9	36.828	55	5
0.5 , 0.5 , 0.5	36.828	55	5
0.5 , 0.5 , 0.1	36.828	55	5
0.5 , 0.1 , 0.9	33.978	55	5
0.5 , 0.1 , 0.5	33.974	55	5
0.5 , 0.1 , 0.1	33.974	55	5
0.1 , 0.9 , 0.9	34.425	55	5
0.1 , 0.9 , 0.5	34.420	55	5
0.1 , 0.9 , 0.1	34.420	55	5
0.1 , 0.5 , 0.9	33.978	55	5
0.1 , 0.5 , 0.5	33.974	55	5
0.1 , 0.5 , 0.1	33.974	55	5
0.1 , 0.1 , 0.9	33.607	55	5
0.1 , 0.1 , 0.5	33.607	55	5
0.1 , 0.1 , 0.1	33.607	55	5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการทดสอบข้อมูลทั้ง 4 ชุด ทำให้เรากำหนด ตัวแปรที่ใช้ในการจัดกลุ่มโดยวิธีการของ SOM ได้ดังต่อไปนี้

Iris กำหนด $\alpha = 0.1$, $\lambda = 0.1$, $\sigma = 0.9$

Diabetes กำหนด $\alpha = 0.1$, $\lambda = 0.1$, $\sigma = 0.9$

Breast กำหนด $\alpha = 0.1$, $\lambda = 0.1$, $\sigma = 0.9$

Wine กำหนด $\alpha = 0.5$, $\lambda = 0.1$, $\sigma = 0.9$

4.3 ผลการทดลอง

4.3.1 Iris Data

ทำการทดสอบข้อมูล Iris จำนวน 150 ข้อมูล ค่า k ที่กำหนดเป็นจำนวนกลุ่มของข้อมูลเพื่อใช้ในทดสอบความถูกต้องในการจัดกลุ่ม ค่าการเรียนรู้ (α) ระยะห่างโหนดข้างเคียง (λ) และค่าซิกม่า (σ) ได้จากการทดสอบกับตัวอย่างข้อมูล Iris จำนวน 90 ตัวอย่าง โดยทำการทดลอง 5 ครั้ง

กำหนด

K-means , PAM

- ค่า k = 3

SOM

- (α=0.1),(σ=0.1), (λ=0.9) รอบการทำงาน 1000 รอบ

ตารางที่ 4.5 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 1

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	150	150	150
เวลาการทำงาน	00.00.12	00.00.12	00.00.13
ระยะทางรวม	28.704	29.165	28.714
จัดกลุ่มถูกต้อง	138	135	132
จัดกลุ่มไม่ถูกต้อง	12	15	18

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 2

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	150	150	150
เวลาการทำงาน	00.00.12	00.00.12	00.00.13
ระยะทางรวม	34.749	29.156	28.714
จัดกลุ่มถูกต้อง	110	134	133
จัดกลุ่มไม่ถูกต้อง	40	16	17

ตารางที่ 4.7 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 3

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	150	150	150
เวลาการทำงาน	00.00.12	00.00.12	00.00.13
ระยะทางรวม	28.704	29.156	28.714
จัดกลุ่มถูกต้อง	136	135	133
จัดกลุ่มไม่ถูกต้อง	14	15	17

ตารางที่ 4.8 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 4

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	150	150	150
เวลาการทำงาน	00.00.12	00.00.12	00.00.13
ระยะทางรวม	28.704	29.165	28.714
จัดกลุ่มถูกต้อง	132	135	133
จัดกลุ่มไม่ถูกต้อง	18	15	17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.9 ผลการทดลองจัดกลุ่มข้อมูล Iris ครั้งที่ 5

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	150	150	150
เวลาการทำงาน	00.00.12	00.00.12	00.00.13
ระยะทางรวม	28.704	29.165	28.714
จัดกลุ่มถูกต้อง	133	135	132
จัดกลุ่มไม่ถูกต้อง	17	15	18

ตารางที่ 4.10 ผลการทดลองจัดกลุ่มข้อมูล Iris โดยเฉลี่ย

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	150	150	150
เวลาการทำงานเฉลี่ย	00.00.12	00.00.12	00.00.13
ระยะทางรวมมากที่สุด	34.749	29.165	28.714
ระยะทางรวมน้อยที่สุด	28.704	29.156	28.714
ระยะทางรวมเฉลี่ย	29.913	29.1614	28.714
จัดกลุ่มถูกต้องโดยเฉลี่ย	130	135	133
จัดกลุ่มไม่ถูกต้องโดยเฉลี่ย	20	15	17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2 Diabetes Data

ทำการทดสอบข้อมูล diabetes จำนวน 768 ชุด ค่า k ที่กำหนดเป็นจำนวนกลุ่มของข้อมูลเพื่อใช้ในทดสอบความถูกต้องในการจัดกลุ่ม ค่าการเรียนรู้ (α) ระยะห่างโหนดข้างเคียง (λ) และค่าซิกม่า (σ) ได้จากการทดสอบกับตัวอย่างข้อมูล diabetes จำนวน 1000 ตัวอย่าง โดยทำการทดลอง 5 ครั้ง

โดยกำหนด

-K-means , PAM

ค่า k=3

โดยกำหนด

SOM

- (α=0.1),(σ=0.1), (λ=0.9) รอบการทำงาน 1000 รอบ

ตารางที่ 4.11 ผลการทดลองจัดกลุ่มข้อมูล Diabetes ครั้งที่ 1

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	768	768	768
เวลาการทำงาน	00.03.60	00.13.25	00.13.32
ระยะทางรวม	186.068	289.074	277.358
จัดกลุ่มถูกต้อง	723	718	718
จัดกลุ่มไม่ถูกต้อง	45	50	50

ตารางที่ 4.12 ผลการทดลองจัดกลุ่มข้อมูล Diabetes ครั้งที่ 2

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	768	768	768
เวลาการทำงาน	00.03.60	00.13.26	00.13.33
ระยะทางรวม	186.068	289.481	277.358
จัดกลุ่มถูกต้อง	726	713	718
จัดกลุ่มไม่ถูกต้อง	42	55	50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.13 ผลการทดลองจัดกลุ่มข้อมูล Diabetes ครั้งที่ 3

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	768	768	768
เวลาการทำงาน	00.03.60	00.13.25	00.13.49
ระยะทางรวม	186.068	389.481	277.358
จัดกลุ่มถูกต้อง	726	716	720
จัดกลุ่มไม่ถูกต้อง	42	52	48

ตารางที่ 4.14 ผลการทดลองจัดกลุ่มข้อมูล Diabetes ครั้งที่ 4

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	768	768	768
เวลาการทำงาน	00.03.60	00.13.23	00.13.35
ระยะทางรวม	186.068	303.813	277.358
จัดกลุ่มถูกต้อง	727	716	718
จัดกลุ่มไม่ถูกต้อง	43	52	50

ตารางที่ 4.15 ผลการทดลองจัดกลุ่มข้อมูล Diabetes ครั้งที่ 5

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	768	768	768
เวลาการทำงาน	00.03.60	00.13.25	00.13.46
ระยะทางรวม	186.068	303.813	277.358
จัดกลุ่มถูกต้อง	723	716	718
จัดกลุ่มไม่ถูกต้อง	45	52	50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.16 ผลการทดลองจัดกลุ่มข้อมูล Diabetes โดยเฉลี่ย

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	768	768	768
เวลาการทำงานเฉลี่ย	00.13.20	00.13.25	00.13.39
ระยะทางรวมมากที่สุด	277.182	389.481	277.358
ระยะทางรวมน้อยที่สุด	277.174	289.074	277.358
ระยะทางรวมเฉลี่ย	277.175	315.132	277.358
จัดกลุ่มถูกต้องโดยเฉลี่ย	725	716	714
จัดกลุ่มไม่ถูกต้องโดยเฉลี่ย	43	52	50

4.3.3 Wine Data

ทำการเปรียบเทียบความถูกต้องข้อมูล wine จำนวน 178 ข้อมูล ค่า k ที่กำหนดเป็นจำนวนกลุ่มของข้อมูลเพื่อใช้ในทดสอบความถูกต้องในการจัดกลุ่ม ค่าการเรียนรู้ (α) ระยะห่างโหนดข้างเคียง (λ) และค่าซิกม่า (σ) ได้จากการทดสอบกับตัวอย่างข้อมูล Iris จำนวน 90 ตัวอย่าง โดยทำการทดลอง 5 ครั้ง

โดยกำหนด

-K-means , PAM

ค่า k=3

โดยกำหนด

SOM

- (α=0.5),(σ=0.1), (λ=0.9) รอบการทำงาน 1000 รอบ

ตารางที่ 4.17 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 1

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	178	178	178
เวลาการทำงาน	00.01.34	00.01.34	00.01.36
ระยะทางรวม	86.955	96.704	87.090
จัดกลุ่มถูกต้อง	172	162	168
จัดกลุ่มไม่ถูกต้อง	6	16	10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.18 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 2

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	178	178	178
เวลาการทำงาน	00.01.33	00.04.41	00.01.36
ระยะทางรวม	87.028	97.873	87.090
จัดกลุ่มถูกต้อง	169	163	168
จัดกลุ่มไม่ถูกต้อง	9	15	10

ตารางที่ 4.19 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 3

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	178	178	178
เวลาการทำงาน	00.01.32	00.01.34	00.01.36
ระยะทางรวม	86.974	97.983	87.090
จัดกลุ่มถูกต้อง	169	163	169
จัดกลุ่มไม่ถูกต้อง	9	15	9

ตารางที่ 4.20 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 4

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	178	178	178
เวลาการทำงาน	00.01.32	00.01.34	00.01.36
ระยะทางรวม	87.028	96.632	87.090
จัดกลุ่มถูกต้อง	169	163	168
จัดกลุ่มไม่ถูกต้อง	9	15	10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.21 ผลการทดลองจัดกลุ่มข้อมูล wine data ครั้งที่ 5

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	178	178	178
เวลาการทำงาน	00.01.32	00.01.34	00.01.36
ระยะทางรวม	86.957	96.632	87.090
จัดกลุ่มถูกต้อง	173	163	168
จัดกลุ่มไม่ถูกต้อง	5	15	10

ตารางที่ 4.22 ผลการทดลองจัดกลุ่มข้อมูล wine data โดยเฉลี่ย

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	178	178	178
เวลาการทำงาน	00.01.32	00.02.00	00.01.36
ระยะทางรวมมากที่สุด	87.028	97.983	87.090
ระยะทางรวมน้อยที่สุด	86.955	96.632	87.090
ระยะทางรวมเฉลี่ย	86.988	97.164	87.090
จัดกลุ่มถูกต้องโดยเฉลี่ย	170	163	168
จัดกลุ่มไม่ถูกต้องโดยเฉลี่ย	8	15	10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.4 Breast Cancer Data

ทำการเปรียบเทียบความถูกต้องข้อมูล breast cancer จำนวน 699 ชุด ค่า k ที่กำหนดเป็นจำนวนกลุ่มของข้อมูลเพื่อใช้ในทดสอบความถูกต้องในการจัดกลุ่ม ค่าการเรียนรู้ (α) ระยะห่างโหนดข้างเคียง (λ) และค่าซิกม่า (σ) ได้จากการทดสอบกับตัวอย่างข้อมูล Iris จำนวน 90 ตัวอย่าง โดยทำการทดลอง 5 ครั้ง

โดยกำหนด

-K-means , PAM

ค่า k = 2

โดยกำหนด

SOM

- (α=0.1),(σ=0.1), (λ=0.9) รอบการทำงาน 1000 รอบ

ตารางที่ 4.23 ผลการทดลองจัดกลุ่มข้อมูล breast cancer ครั้งที่ 1

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	699	699	699
เวลาการทำงาน	00.13.11	00.13.14	00.13.20
ระยะทางรวม	332.453	344.188	332.672
จัดกลุ่มถูกต้อง	649	644	646
จัดกลุ่มไม่ถูกต้อง	50	55	53

ตารางที่ 4.24 ผลการทดลองจัดกลุ่มข้อมูล breast cancer ครั้งที่ 2

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	699	699	699
เวลาการทำงาน	00.13.12	00.13.13	00.13.21
ระยะทางรวม	332.453	344.188	332.672
จัดกลุ่มถูกต้อง	651	642	646
จัดกลุ่มไม่ถูกต้อง	48	53	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.25 ผลการทดลองจัดกลุ่มข้อมูล breast cancer ครั้งที่ 3

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	699	699	699
เวลาการทำงาน	00.13.12	00.13.12	00.13.24
ระยะทางรวม	332.453	344.188	332.672
จัดกลุ่มถูกต้อง	649	644	646
จัดกลุ่มไม่ถูกต้อง	50	55	53

ตารางที่ 4.26 ผลการทดลองจัดกลุ่มข้อมูล breast cancer ครั้งที่ 4

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	699	699	699
เวลาการทำงาน	00.13.12	00.13.15	00.13.24
ระยะทางรวม	332.453	344.188	332.672
จัดกลุ่มถูกต้อง	646	644	643
จัดกลุ่มไม่ถูกต้อง	53	55	50

ตารางที่ 4.27 ผลการทดลองจัดกลุ่มข้อมูล breast cancer ครั้งที่ 5

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	699	699	699
เวลาการทำงาน	00.13.12	00.13.18	00.13.24
ระยะทางรวม	332.453	344.188	332.672
จัดกลุ่มถูกต้อง	649	642	643
จัดกลุ่มไม่ถูกต้อง	50	57	53

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.28 ผลการทดลองจัดกลุ่มข้อมูล breast cancer โดยเฉลี่ย

อัลกอริทึม	K-means	PAM	SOM
จำนวนข้อมูล	699	699	699
เวลาการทำงาน	00.13.12	00.13.20	00.13.20
ระยะทางรวมมากที่สุด	332.453	344.188	332.672
ระยะทางรวมน้อยที่สุด	332.453	344.188	332.672
ระยะทางรวมเฉลี่ย	332.453	344.188	332.672
จัดกลุ่มถูกต้องโดยเฉลี่ย	649	644	646
จัดกลุ่มไม่ถูกต้องโดยเฉลี่ย	50	55	53

4.4 วิเคราะห์ผลการทดลอง

จากผลการทดลองการจัดกลุ่มจากข้อมูลทั้ง 4 ชุดคือ Iris , Breast Cancer , Wine , Diabetes

Iris จากการทดสอบ ค่าระยะห่างจุดศูนย์กลางของ SOM มีค่าน้อยที่สุด แต่ความถูกต้องในการจัดกลุ่มโดยใช้ PAM มีความถูกต้องน้อยที่สุด การจัดกลุ่มโดยใช้ k-means ค่าของระยะทางรวมจะไม่เท่ากันดังนั้นถ้าใช้วิธีการจัดกลุ่มวิธีนี้เราต้องทำการจัดกลุ่มหลายครั้งเพื่อหาค่าเฉลี่ยเพื่อใช้ในวิเคราะห์การจัดกลุ่ม

Breast Cancer จากการทดสอบ ค่าระยะห่างจุดศูนย์กลางของ k-means มีค่าน้อยที่สุด และความถูกต้องมากที่สุด การจัดกลุ่มโดยใช้ SOM ค่าระยะห่างจุดศูนย์กลางมากและความถูกต้องน้อยกว่า k-means สุดท้าย การจัดกลุ่มโดยใช้ PAM ค่าระยะห่างจุดศูนย์กลางมากและความถูกต้องน้อยที่สุด

Wine จากการทดสอบ ค่าระยะห่างจุดศูนย์กลางของ k-means มีค่าน้อยที่สุด และความถูกต้องมากที่สุด การจัดกลุ่มโดยใช้ SOM ค่าระยะห่างจุดศูนย์กลางมากและความถูกต้องน้อยกว่า k-means สุดท้าย การจัดกลุ่มโดยใช้ PAM ค่าระยะห่างจุดศูนย์กลางมากและความถูกต้องน้อยที่สุด

Diabetes จากการทดสอบ ค่าระยะห่างจุดศูนย์กลางของ k-means มีค่าน้อยที่สุด และความถูกต้องมากที่สุด การจัดกลุ่มโดยใช้ SOM ค่าระยะห่างจุดศูนย์กลางมากและความถูกต้องน้อยกว่า k-means สุดท้าย การจัดกลุ่มโดยใช้ PAM ค่าระยะห่างจุดศูนย์กลางมากและความถูกต้องน้อยที่สุด

จากผลการทดลองการจัดกลุ่มจากข้อมูลทั้ง 4 ชุดคือ Iris , Breast Cancer , Wine , Diabetes จะเห็นได้ว่าการใช้อัลกอริทึม k-means ค่าของระยะทางรวมจะไม่เท่ากันดังนั้นถ้าใช้วิธีการจัดกลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีนี้เราต้องทำการจัดกลุ่มหลายครั้งเพื่อหาค่าเฉลี่ยเพื่อใช้ในวิเคราะห์การจัดกลุ่ม จัดกลุ่มโดยใช้อัลกอริทึม PAM ในการจัดกลุ่มนั้นจะใช้ข้อมูลเป็นจุดศูนย์กลางทำให้วนหาจุดศูนย์กลางที่ดีที่สุด การจัดกลุ่มข้อมูลแบบแบ่งส่วนข้อมูล (Partitioning Method) จุดศูนย์กลางเกิดจากการสุ่มทำให้ผลที่ได้จะต้องทำการหาค่าเฉลี่ย การทำงานของ SOM การใช้ข้อมูลทั้งหมดถ่วงน้ำหนักในการจัดกลุ่มทำให้ระยะห่างไม่เปลี่ยนแปลง ค่าระยะทางรวมน้อยที่สุดไม่สามารถบอกว่าการจัดกลุ่มดีที่สุดเสมอไปจากการทดสอบการจัดกลุ่มค่าระยะห่างมากแต่การจัดกลุ่มมีความถูกต้องมากกว่าค่าระยะห่างน้อยทำให้ในการทดลองควรมีตัววัดประสิทธิภาพอื่นเพื่อจะทำให้การจัดกลุ่มมีประสิทธิภาพเพิ่มขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการศึกษาและข้อเสนอแนะ

จากการศึกษาการจัดกลุ่มข้อมูลโดยใช้ K- means อัลกอริทึม และ PAM (Partitioning around Medoids) และ SOM (Self-organizing map) ในโครงการนี้สามารถสรุปผลการดำเนินงานและสรุปผลการทดลองรวมถึงข้อเสนอแนะดังต่อไปนี้

5.1 สรุปผลการศึกษา

จากการจัดกลุ่มโดยใช้ข้อมูล4 ชุดคือ Iris , Breast Cancer , Wine , Diabetes ทำการทดลอง 5 ครั้งสามารถสรุปได้ดังต่อไปนี้

อัลกอริทึม k-means ในการจัดกลุ่ม

-จุดศูนย์กลางกลางของข้อมูลมาจากการสุ่มและใช้ค่าเฉลี่ยในการหาจุดศูนย์กลางทำให้ ผลจากการจัดกลุ่มแต่ละครั้งไม่เท่ากัน

- ค่าระยะทางรวมที่ได้จะ ไม่คงที่
- มีประสิทธิภาพดีถ้าข้อมูลมีขนาดไม่มากและมีจำนวนกลุ่มน้อย
- ใช้เวลาการทำงานน้อยที่สุด
- จากการทดลอง มีประสิทธิภาพในการจัดกลุ่มข้อมูลทั้ง 4 ข้อมูลมีประสิทธิภาพดีที่สุด

อัลกอริทึม PAM ในการจัดกลุ่ม

- ค่าระยะทางที่ได้มีค่ามากที่สุด
- เวลาการทำงานมากที่สุด
- ระยะทางรวมมีการเปลี่ยนแปลงน้อย
- จุดศูนย์กลางจะคงที่

อัลกอริทึม SOM ในการจัดกลุ่ม

- การทำงานของ SOM การใช้ข้อมูลทั้งหมดถ่วงน้ำหนักในการจัดกลุ่มทำให้ระยะทางรวมคงที่
- การกำหนดตัวแปรของ SOM มีผลต่อประสิทธิภาพในการจัดกลุ่ม
- จากการทดลองการจัดกลุ่มจะมีประสิทธิภาพลองมาจาก K-means
- ใช้เวลาการทำงานมากกว่า K-means แต่น้อยกว่า PAM

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 ประโยชน์ที่ได้รับจากการศึกษาและพัฒนาระบบ

1. ทำให้เข้าใจทฤษฎี และหลักการทำงานของอัลกอริทึมต่างๆในการทำ Cluster ข้อมูลมากขึ้นรวมถึงทำให้ทราบปัญหาที่พบบ่อยในการทำ Cluster
2. ทำให้ได้โปรแกรมในการเปรียบเทียบประสิทธิภาพในการทำ Cluster ที่ทำให้ทราบว่าข้อมูลชุดนั้นเหมาะสมกับอัลกอริทึมแบบใด ทำให้เราสามารถเลือกใช้อัลกอริทึมได้ตามวัตถุประสงค์

5.3 ข้อเสนอแนะ

เพื่อปรับปรุงโปรแกรมประยุกต์นี้ในอนาคต ผู้ศึกษามีความคิดเห็นว่า เพื่อเพิ่มความถูกต้องในการจัดกลุ่มผลที่ได้รับมาจากโปรแกรมผู้ใช้ต้องทำการตรวจสอบข้อมูลที่ได้รับเองทำให้อาจเกิดการเสียเวลาถ้าข้อมูลมีขนาดใหญ่มาก และการตั้งค่าอัตราการเรียนรู้ที่เหมาะสมกับข้อมูลทำค่อนข้างยากผู้ใช้ต้องมีความชำนาญในการใช้โปรแกรมจึงจะทำให้ใช้งานโปรแกรมให้มีประสิทธิภาพเพิ่มขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

Cabena, I. et al. 1997. **Discovering Data Mining from Concept to Implemetation**. New Jersy:Prentice Hall.

Han, J. and Kamber, M.2001.**Data Mining: concept and techniques**. United States of America: Academic Press.

Zhexue Huang. **Clustering large data sets with mixed numeric and categorical values**.[Online].Available: http://www.act.cmis.csiro.au/gjw/papers/apkdd.pdf

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อผู้เขียน นายณัฐพงษ์ ฉัตรแก้วไพธิ์ทอง

วันที่เกิด วันที่ 2 ธันวาคม 2527

วุฒิการศึกษาระดับ ปริญญาตรี บริหารธุรกิจบัณฑิต

สถานที่สำเร็จการศึกษา คณะบริหารธุรกิจสาขาคอมพิวเตอร์ธุรกิจ มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ

ปีที่สำเร็จการศึกษา 2548

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้