การหาจุดศูนย์กลางเริ่มต้นของอัลกอริทึมเคมีน

INITIAL CENTER POINTS FOR K-MEANS ALGORITHM

จุฑามาศ จันทร์ทุ่ง

ฉัตรวี จิตรีศิลป์

สุนิสา ภู่ทิม

ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาวิทยาศาสตรบัณฑิต (คณิตศาสตร์ประยุกต์)

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2560

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

INITIAL CENTER POINTS FOR K-MEANS ALGORITHM

JUTHAMAS JANTHUNG

CHATRAWEE JITTREESIL

SUNISA POOTIM

A SPECIAL PROBLEM SUBMITTED INPARTIAL FULFILLMENT OF

THE REQUIREMENT FOR

THE DEGREE OF BACHELOR OF SCIENCE DEPARTMENT OF

APPLIED MATHEMATICS, FACULTY OF SCIENCE

KINGMONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

ACADEMIC YEAR 2017

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ การหาจุดศูนย์กลางเริ่มต้นของอัลกอริทึมเคมีน

Initial Center Points for K-means Algorithm

ชื่อนักศึกษา นางสาวจุฑามาศ จันทร์ทุ่ง รหัสนักศึกษา 57050026

นางสาวฉัตรวี จิตรีศิลป์ รหัสนักศึกษา 57050030

นางสาวสุนิสา ภู่ทิม รหัสนักศึกษา 57050150

ปริญญา วิทยาศาสตรบัณฑิต (คณิตศาสตร์ประยุกต์)

ภาควิชา คณิตศาสตร์

ปีการศึกษา 2560

อาจารย์ที่ปรึกษา รองศาสตราจารย์ไพรบูลย์ พันธรักษ์พงษ์

คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง (สจล.) อนุมัติให้ปัญหาพิเศษนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต (คณิตศาสตร์ประยุกต์) ประจำปีการศึกษา 2560

คณะกรรมการสอบ	ลายมือชื่อ
รศ.ดร.ฉัฐไชย์ ลีนาวงศ์ ประธานกรรมการ	
ดร.บุษยมาส พิมพ์พรรณชาติ กรรมการ	
รศ.ไพรบูลย์ พันธรักษ์พงษ์ กรรมการและอาจารย์ที่ปรึกษา	

ลิขสิทธิ์ของคณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อปัญหาพิเศษ การหาจุดศูนย์กลางเริ่มต้นของอัลกอริทึมเคมีน
Initial Center Points for K-means Algorithm

ชื่อนักศึกษา นางสาวจุฑามาศ จันทร์ทุ่ง รหัสนักศึกษา 57050026
นางสาวฉัตรวี จิตรีศิลป์ รหัสนักศึกษา 57050030
นางสาวสุนิสา ภู่ทิม รหัสนักศึกษา 57050150

ปริญญา วิทยาศาสตรบัณฑิต (คณิตศาสตร์ประยุกต์)

ภาควิชา คณิตศาสตร์

คณะ วิทยาศาสตร์

ปีการศึกษา 2560

อาจารย์ที่ปรึกษา รองศาสตราจารย์ไพรบูลย์ พันธรักษ์พงษ์

บทคัดย่อ

ปัญหาพิเศษนี้นำเสนอผลการศึกษาเพื่อหาวิธีเลือกค่าเริ่มต้นสำหรับการแบ่งกลุ่มแบบเคมีน (K-means) วิธีการแบ่งกลุ่มแบบเคมีนต้องมีการกำหนดค่าจำนวนกลุ่มที่ต้องการแบ่งกลุ่ม (K) และจุดศูนย์กลางเริ่มต้น (Initial Centroid) ดังนั้นผู้ศึกษาได้ศึกษาวิธีเลือกค่าเริ่มต้นจำนวน 4 วิธีคือ 1) เลือกค่าเริ่มต้นโดยวิธีการสุ่ม 2) เลือกค่าเริ่มต้นโดยการคำนวณหาค่ากึ่งกลางของช่วง 3) เลือกค่าเริ่มต้นโดยการสุ่มหลายค่าและ 4) เลือกค่าเริ่มต้นโดยใช้ความถี่สูงสุดของข้อมูล โดยทดลองกับเซตข้อมูลจำนวน 3 เซตข้อมูลบน $\mathbb{R}^2$ คือ 1) เซตข้อมูลไอริส (Iris-setosa) 2) เซตข้อมูลไวน์ (Wine) และ 3) เซตข้อมูลไฟป่า (Forest Fires) ผลการศึกษาพบว่าการเลือกค่าเริ่มต้นโดยการคำนวณหาค่ากึ่งกลางของช่วงทำให้การคำนวณของเคมีนมีจำนวนวนซ้ำน้อยที่สุด

คำสำคัญ : การแบ่งกลุ่มวิธี K-means,ระยะทางยุคลิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Title	Initial Center Points for K-means Algorithm
Students	Miss Juthamas Janthung 57050026
	Miss Chatrawee Jittreesil 57050030
	Miss Sunisa Pootim 57050150
Degree	Bachelor of Science (Applied Mathematics)
Department	Mathematics
Faculty	Science
University	King Mongkut's Institute of Technology Ladkrabang (KMITL)
Academic Year	2017
Advisor	Assoc.Prof.Praiboon Pantaragphong

Abstract

This special problem presents the method to choose the initial points for K-means clustering. K-means must be identified the number of clusters and centroids. We are to studying four method to choose the initial points such as 1) by random single value,2) by calculated the middle value of the range, 3) by random the multiple values and 4) by the frequency of data. We are experiment these method with three data sets on $\mathbb{R}^2$ such as 1) Iris Data Set, 2) Wine Data Set and 3) Forest Fires Data Set. The experiment show that the initial points by the middle value of the range make to the least of number of iterations of K-means.

Keywords: K-means clustering, Euclidean distance

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปัญหาพิเศษเล่มนี้สำเร็จลุล่วงได้ด้วยดี เนื่องด้วยได้รับความกรุณาเป็นอย่างสูงจาก รศ.ไพรบูลย์ พันธรักษ์พงษ์ อาจารย์ที่ปรึกษาปัญหาพิเศษ ซึ่งได้กรุณาให้วิชาความรู้ คำปรึกษา คำแนะนำ ตลอดจนตรวจสอบปรับปรุงแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่อย่างดียิ่ง คณะผู้จัดทำตระหนักถึงความตั้งใจจริงและความทุ่มเทของอาจารย์เสมอมา และขอกราบขอบพระคุณเป็นอย่าสูงไว้ ณ ที่นี้

ขอขอบคุณ รศ.ดร.ฉัฐไชย์ ลีนาวงศ์ และ ดร.บุษยมาส พิมพ์พรรณชาติ ประธานและกรรมการสอบ ปัญหาพิเศษ ตลอดจนคณาจารย์ภาควิชาคณิตศาสตร์ ที่ให้ความรู้ต่างๆ แก่คณะผู้จัดทำเสมอมา รวมถึงขอบคุณเจ้าหน้าที่ประจำภาควิชาคณิตศาสตร์ทุกท่านที่คอยช่วยเหลือในด้านการอำนวยความสะดวกเกี่ยวกับอุปกรณ์ที่จำเป็นต่างๆ ตลอดจนรุ่นพี่นักศึกษาปริญญาเอกทุกท่านที่ให้ข้อมูลต่างๆ ที่เอื้อต่อการทำปัญหาพิเศษเล่มนี้ จนทำให้ปัญหาพิเศษนี้สำเร็จลุล่วงไปได้ด้วยดี

อนึ่ง คณะผู้จัดทำหวังว่าปัญหาพิเศษเล่มนี้จะมีประโยชน์อยู่ไม่น้อย จึงขอมอบส่วนดีทั้งหมดนี้ ให้แก่เหล่าคณาจารย์ ที่ได้ประสิทธิประสาทวิชาจนทำให้ได้ผลงานปัญหาพิเศษนี้เป็นประโยชน์ต่อผู้ที่เกี่ยวข้อง และขอมอบความกตัญญูกตเวทิตาคุณ แด่บิดามารดา และผู้มีพระคุณทุกท่าน สำหรับข้อบกพร่องต่างๆ ที่อาจจะเกิดขึ้น คณะผู้จัดทำขอน้อมรับผิดแต่เพียงผู้เดียว และยินดีที่จะรับฟังคำชี้แนะจากทุกท่านที่ได้เข้ามาศึกษา เพื่อเป็นประโยชน์แก่การปรับปรุงแก้ไขและพัฒนาให้ดียิ่งขึ้นต่อไป

จุฑามาศ จันทร์ทุ่ง
ฉัตรวี จิตรีศิลป์
สุนิสา ภู่ทิม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

เรื่อง **หน้า**

บทคัดย่อภาษาไทย ..... ก

บทคัดย่อภาษาอังกฤษ ..... ข

กิตติกรรมประกาศ ..... ค

สารบัญ ..... ง

สารบัญตาราง ..... จ

สารบัญรูปภาพ ..... ฉ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการทำ 1

1.3 ขอบเขตของปัญหา 1

1.4 ประโยชน์ที่คาดว่าจะได้รับ 2

1.5 ขั้นตอนในการดำเนินงาน 2

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง 4

2.1 การแบ่งกลุ่ม 4

2.2 การแทนข้อมูลด้วยเมทริกซ์ 4

2.3 ระยะทางแบบยุคลิด 6

2.4 การแบ่งกลุ่มแบบเคมีน 7

2.5 ผลรวมกำลังสองของค่าความผิดพลาด 8

2.6 วิธีการเอลโบ 8

2.7 เซตข้อมูลที่ใช้ในการทดลอง 10

บทที่ 3 ขั้นตอนวิธีเลือกค่าเริ่มต้น 12

3.1 วิธีที่จะศึกษา 12

3.1 1.วิธีการหาค่าเริ่มต้นของกลุ่มแบบสุ่ม 13

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.12. การหาค่ากลางเริ่มต้นของกลุ่มโดยการหาค่ากึ่งกลางของช่วง 13
3.1.3 การหาค่ากลางเริ่มต้นของกลุ่มโดยการคำนวณแบบสุ่มหลายค่า 16
3.1.4 การหาค่ากลางเริ่มต้นของกลุ่มโดยใช้ความถี่สูงสุดของข้อมูล 19
บทที่ 4 ผลการศึกษา 23
4.1 เซตข้อมูลไอริส 23
4.1.1 จำนวนกลุ่มที่เหมาะสมของเซตข้อมูลไอริส 24
4.1.2 วิธีการหาค่าเริ่มต้นของกลุ่มโดยการสุ่มของเซตข้อมูลไอริส 25
4.1.3 การหาค่ากลางเริ่มต้นของกลุ่มโดยการหาค่ากึ่งกลางของช่วงของเซตข้อมูลไอริส 27
4.1.4 การหาค่ากลางเริ่มต้นของกลุ่มโดยการคำนวณสุ่มหลายค่า 28
4.1.5 การหาค่ากลางเริ่มต้นของกลุ่มโดยการเลือกจากความถี่สูงสุดของเซตข้อมูลไอริส 30
4.1.6 จำนวนรอบการคำนวณของเซตข้อมูลไอริส 31
4.2 เซตข้อมูลไวน์ 32
4.2.1 จำนวนกลุ่มที่เหมาะสมของเซตข้อมูลไวน์ 32
4.2.2 วิธีการหาค่าเริ่มต้นของกลุ่มโดยการสุ่มของเซตข้อมูลไวน์ 34
4.2.3 การหาค่ากลางเริ่มต้นของกลุ่มโดยการหาค่ากึ่งกลางของช่วงของเซตข้อมูลไวน์ 35
4.2.4 การหาค่ากลางเริ่มต้นของกลุ่มโดยการคำนวณสุ่มหลายค่า 36
4.2.5 การหาค่ากลางเริ่มต้นของกลุ่มโดยการสุ่มจากความถี่สูงสุดของข้อมูล 38
เซตข้อมูล Wine จำนวน 178 ข้อมูลดังตาราง 38
4.2.6 จำนวนรอบการคำนวณของเซตข้อมูลไวน์ 39
4.3 เซตข้อมูลไฟป่า 39
4.3.1 จำนวนกลุ่มที่เหมาะสมของเซตข้อมูลไฟป่า 40
4.3.2 วิธีการหาค่าเริ่มต้นของกลุ่มโดยการสุ่มของเซตข้อมูลไฟป่า 41
4.3.3 การหาค่ากลางเริ่มต้นของกลุ่มโดยการหาค่ากึ่งกลางของช่วงของเซตข้อมูลไฟป่า 42
4.3.5 การหาค่ากลางเริ่มต้นของกลุ่มโดยการเลือกจากความถี่สูงสุดของเซตข้อมูลไฟป่า 46
4.3.6 จำนวนรอบการคำนวณของเซตข้อมูลไฟป่า 47

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 จำนวนรอบการคำนวณเคมีนของแต่ละวิธี 48

บทที่ 5 สรุปผลและข้อเสนอแนะ 49

5.1 สรุปผล 49

5.2 ข้อเสนอแนะ 49

เอกสารอ้างอิง 50

ภาคผนวก 51

ภาคผนวก ก 52

ภาคผนวก ข 57

ภาคผนวก ค 62

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ **หน้า**

ตารางที่ 1.1 แผนการดำเนินงาน 3
ตารางที่ 4.1 แสดงค่าความชันของแต่ละจุด 24
ตารางที่ 4.2 แสดงผลการใช้คำสั่งสุ่ม 1000 ครั้ง 25
ตารางที่ 4.3 แสดงจำนวนรอบการคำนวณโดยใช้คำสั่งสุ่มของเซตข้อมูลไอริส 26
ตารางที่ 4.4 แสดงจุดศูนย์กลางและรอบบการคำนวณของเซตข้อมูลไวน์ 28
ตารางที่ 4.5 แสดงจำนวนรอบการคำนวณเคมีนของฐานข้อมูลไอริส 28
ตารางที่ 4.6 แสดงจุดศูนย์กลางและจำนวนรอบในการคำนวณการสุ่มจำนวน 1000 รอบ 30
ตารางที่ 4.7 แสดงจำนวนรอบการคำนวณวิธีหาค่ากึ่งกลางของช่วง 30
ตารางที่ 4.8 แสดงความถี่ของข้อมูลที่ซ้ำกันของเซตข้อมูลไอริส 30
ตารางที่ 4.9 แสดงรอบการคำนวณของแต่ละวิธีของเซตข้อมูลไอริส 31
ตารางที่ 4.10 แสดงค่าความชันของแต่ละจุด 33
ตารางที่ 4.11 แสดงจุดศูนย์กลางของเซตข้อมูลไวน์จากการสุ่มจำนวน 1000 รอบ 34
ตารางที่ 4.12 แสดงจำนวนรอบการคำนวณโดยใช้คำสั่งสุ่มของเซตข้อมูลไวน์ 35
ตารางที่ 4.13 แสดงจุดศูนย์กลางและจำนวนรอบของการคำนวณของเซตข้อมูลไวน์ 36
ตารางที่ 4.14 แสดงการสุ่มหลายค่าของเซตข้อมูลไวน์ 37
ตารางที่ 4.15 แสดงจุดศูนย์กลางและจำนวนรอบในการคำนวณจากการสุ่มจำนวน 1000 รอบ 37
ตารางที่ 4.16 แสดงค่าเฉลี่ยของวิธีการสุ่มของเซตข้อมูลไวน์จากการสุ่มจำนวน 1000 รอบ 38
ตารางที่ 4.17 เซตข้อมูลไวน์ 38
ตารางที่ 4.18 แสดงรอบการคำนวณของแต่ละวิธีของเซตข้อมูลไวน์ 40
ตารางที่ 4.19 แสดงค่าความชันของแต่ละจุดและค่าผลต่างของความชัน 40
ตารางที่ 4.20 แสดงจุดศูนย์กลางและจำนวนรอบการคำนวณจากการสุ่มจำนวน 1000 รอบ 41
ตารางที่ 4.21 แสดงค่าเฉลี่ยของวิธีการสุ่มของเซตข้อมูลไฟป่า 42
ตารางที่ 4.22 แสดงจุดศูนย์กลางและจำนวนรอบการคำนวณของเซตข้อมูลไฟป่า 44
ตารางที่ 4.23 แสดงผลการสุ่มของวิธีสุ่มแบบหลายค่าและจำนวนรอบการคำนวณเซตข้อมูลไฟป่า . 44
ตารางที่ 4.24 แสดงจุดศูนย์กลางและจำนวนวนรอบการคำนวณจากการสุ่มจำนวน 1000 รอบ 45
ตารางที่ 4.25 แสดงค่าเฉลี่ยของวิธีการสุ่มของเซตข้อมูลไฟป่าจากการสุ่มจำนวน 1000 รอบ 46
ตารางที่ 4.26 แสดงความถี่ของข้อมูลฐานข้อมูลไฟป่า 46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.27 แสดงจุดศูนย์กลางและจำนวนรอบการคำนวณของเซตข้อมูลไฟป่า 47
ตารางที่ 4.28 แสดงรอบการคำนวณของแต่ละวิธีของเซตข้อมูลไฟป่า.....47
ตารางที่ 4.29 แสดงจำนวนรอบแต่ละวิธีของแต่ละเซตข้อมูล.....48

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ

รูปที่ **หน้า**

รูปที่ 2.1 เมทริกซ์ของข้อมูล 4
รูปที่ 2.2 แสดงตัวอย่างเมทริกซ์ข้อมูลของเซตข้อมูลไอริส 5
รูปที่ 2.3 การหาระยะทางแบบยุคลิด 6
รูปที่ 2.4 ดอกไอริส 11
รูปที่ 3.1 คำสั่งในการคำนวณเคมีน 12
รูปที่ 3.2 การใช้คำสั่งในการสุ่มจุดศูนย์กลางเริ่มต้นในการคำนวณ 13
รูปที่ 3.3 การใช้คำสั่งในการคำนวณแบบกึ่งกลางของช่วง 14
รูปที่ 3.4 แสดงการหาค่ากึ่งกลางของแต่ละช่วงของฐานข้อมูล 16
รูปที่ 3.5 การใช้คำสั่งในการคำนวณแบบสุ่มหลายค่า 17
รูปที่ 3.6 คำสั่งใช้ในการคำนวณหาความถี่ของข้อมูล 20
รูปที่ 3.7 แสดงความถี่ของข้อมูล 22
รูปที่ 4.1 กราฟแสดงความยาวกลีบดอก-ความกว้างกลีบดอกเซตข้อมูลไอริส 23
รูปที่ 4.2 แสดงค่าผลรวมกำลังสองของค่าความผิดพลาดของเซตข้อมูลไอริส 24
รูปที่ 4.3 แสดงการหาจุดกึ่งกลางของแต่ละช่วงของเซตข้อมูลไอริส 27
รูปที่ 4.4 แสดงความถี่ของข้อมูลเซตข้อมูลไอริส 30
รูปที่ 4.5 กราฟแสดงกรดคาร์บอลิก-สารต้านอนุมูลอิสระของเซตข้อมูลไวน์ 32
รูปที่ 4.6 กราฟแสดงค่าผลรวมกำลังสองของค่าความผิดพลาดของเซตข้อมูลไวน์ 32
รูปที่ 4.7 แสดงการหาจุดกึ่งกลางของแต่ละช่วงของเซตข้อมูลไวน์ 36
รูปที่ 4.8 กราฟแสดงค่าความชื้นของเชื้อเพลิง-ดัชนีดีซีของเซตข้อมูลไฟป่า 39
รูปที่ 4.9 แสดงค่าผลรวมกำลังสองของค่าความผิดพลาดของเซตข้อมูลไฟป่า 40
รูปที่ 4.10 แสดงการหาจุดกึ่งกลางของแต่ละช่วงของเซตข้อมูลไฟป่า 43

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การแบ่งกลุ่ม (Clustering) เป็นวิธีการวิเคราะห์ข้อมูลเพื่อจัดกลุ่มข้อมูล โดยจะพิจารณาข้อมูลที่มีความคล้ายหรือเหมือนกันจัดไว้ในกลุ่มเดียวกัน ซึ่งในปัจจุบันนี้มีวิธีการเกี่ยวกับการแบ่งกลุ่มอยู่มากมาย เช่น การแบ่งกลุ่มแบบเคมีน การแบ่งกลุ่มแบบลำดับขั้น การแบ่งกลุ่มแบบสเปกตรัมและกราฟ เป็นต้น

สำหรับวิธีการแบ่งกลุ่มแบบเคมีน (K-means) จะจัดข้อมูลที่มีลักษณะเหมือนกัน หรือคล้ายกันให้อยู่กลุ่มเดียวกัน โดยการวัดระยะห่างระหว่างค่าของข้อมูลกับค่ากลางของกลุ่ม จากนั้นจะเลือกนำข้อมูลจัดไว้ในกลุ่มที่ได้ค่าระยะห่างน้อยที่สุด และคำนวณค่ากลางของกลุ่มใหม่ จะทำเช่นนี้จนกระทั่งค่ากลางของกลุ่มไม่เปลี่ยนแปลงหรือเปลี่ยนแปลงในระดับที่ยอมรับได้ วิธีการแบ่งกลุ่มแบบเคมีนจะต้องมีการกำหนดค่า K (จำนวนกลุ่ม) และจุดศูนย์กลางของกลุ่ม (Centroid) แบบสุ่ม งานที่ทำของขั้นตอนวิธีเคมีนจะเป็น $O(nKt)$ เมื่อ n เป็นจำนวนข้อมูลทั้งหมด t เป็นจำนวนรอบ และ K เป็นจำนวนกลุ่ม จากการศึกษากับพบว่า จุดศูนย์กลางเริ่มต้น ถ้าสุ่มแล้วได้ค่าใกล้เคียงกับจุดศูนย์กลาง จำนวนรอบการทำซ้ำจะน้อยครั้ง ถ้าสุ่มได้ค่าที่ห่างจากจุดศูนย์กลางจำนวนรอบจะมาก

ในปัญหาพิเศษนี้จึงสนใจศึกษาเกี่ยวกับการเลือกค่าเริ่มต้นที่เป็นจุดศูนย์กลางของการแบ่งกลุ่มแบบเคมีน

1.2 วัตถุประสงค์ของการทำ

1) ศึกษาวิธีการเลือกค่าเริ่มต้นที่เป็นค่ากลางของกลุ่มสำหรับวิธีการแบ่งกลุ่มแบบเคมีน

1.3 ขอบเขตของปัญหา

1) ศึกษาขั้นตอนวิธีการหาค่ากลางของการแบ่งกลุ่มแบบเคมีน

2) เซตข้อมูลที่ใช้ศึกษาได้แก่เซตข้อมูลไอริส (Iris) เซตข้อมูลไวน์ (wine) และเซตข้อมูลไฟป่า (Forest Fires)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้วิธีการที่หาค่าเริ่มต้นที่เป็นค่ากลางของกลุ่มสำหรับการแบ่งกลุ่มแบบเคมีน

2) ทำให้การแบ่งกลุ่มของเคมีนมีประสิทธิภาพยิ่งขึ้น

1.5 ขั้นตอนในการดำเนินงาน

1) ศึกษาขั้นตอนวิธีของวิธีการแบ่งกลุ่มข้อมูลแบบเคมีน

2) ศึกษาวิธีการที่เกี่ยวข้องกับการกำหนดค่าเริ่มต้นของเคมีน

3) ทดสอบวิธีกับข้อมูลหลายๆเซตข้อมูล

4) ปรับปรุงและแก้ไขในข้อผิดพลาดที่เกิดขึ้น

5) สรุปผลการดำเนินงาน พร้อมทำเอกสารประกอบในการทำปัญหาพิเศษ

6) นำเสนอปัญหาพิเศษ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สามารถแสดงกิจกรรมดำเนินงานได้ดังตารางที่ 1.1

ตารางที่ 1.1 แผนการดำเนินงาน

ขั้นตอนการดำเนินงาน	ระยะเวลาการดำเนินงาน								
	ปี 2560				ปี 2561				
	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.	พ.ค.
ศึกษาขั้นตอนวิธีของการแบ่งกลุ่มแบบเคมีน									
ศึกษาวิธีการที่เกี่ยวข้องกับการกำหนดค่าเริ่มต้นของเคมีน									
ทดสอบวิธีกับข้อมูลหลายๆเซตข้อมูล									
ปรับปรุงและแก้ไขในข้อผิดพลาดที่เกิดขึ้น									
สรุปผลการดำเนินงาน พร้อมทำเอกสารประกอบในการทำปัญหาพิเศษ									
นำเสนอปัญหาพิเศษ									

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

2.1 การแบ่งกลุ่ม

การแบ่งกลุ่ม (Clustering) เป็นวิธีการวิเคราะห์ข้อมูล โดยจะแบ่งชุดข้อมูลออกเป็นกลุ่ม (Cluster) โดยนำข้อมูลที่มีคุณลักษณะเหมือนกันหรือคล้ายกันจัดไว้ในกลุ่มเดียวกัน ขั้นตอนวิธีที่ใช้ในการแบ่งกลุ่มจะอาศัยความเหมือน (Similarity) หรือความใกล้ชิด (Proximity) โดยคำนวณจากการวัดระยะห่างระหว่างข้อมูล โดยใช้การวัดระยะห่างแบบต่างๆ เช่น การวัดระยะห่างแบบยุคลิด (Euclidean distance) การวัดระยะห่างแบบแมนฮัตตัน (Manhattan distance) การวัดระยะห่างแบบเชบิเชฟ (Chebyshev distance)

การแบ่งกลุ่มในปัจจุบันนั้นมีหลายวิธี เช่น การแบ่งกลุ่มแบบเคมีน (K-means Clustering) การแบ่งกลุ่มแบบลำดับขั้น (Hierarchical Clustering) การแบ่งกลุ่มสเปกตรัมและกราฟ (Spectral and Graph Clustering) เป็นต้น ในปัญหาพิเศษนี้ผู้ศึกษาจะศึกษาวิธีการแบ่งกลุ่มแบบเคมีนโดยใช้การวัดระยะห่างแบบยุคลิด (Euclidean distance)

2.2 การแทนข้อมูลด้วยเมทริกซ์

ข้อมูลมักถูกแสดงหรืออยู่ในรูปเมทริกซ์ข้อมูลที่มีขนาด $n \times d$ นั่นคือ n แถวและ d หลักหรือคอลัมน์ เมื่อแถวคือคือสิ่งที่มีอยู่ในฐานข้อมูลและคอลัมน์แทนแอตทริบิวต์หรือคุณลักษณะของสิ่งที่สนใจ แต่ละแถวในเมทริกซ์ข้อมูลเก็บค่าของแอตทริบิวต์ที่สังเกต เมทริกซ์ข้อมูลแสดงดังรูปต่อไปนี้

$$D = \begin{array}{c|cccc} & X_1 & X_2 & \cdots & X_d \\ \hline x_1 & x_{11} & x_{12} & \cdots & x_{1d} \\ x_2 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n & x_{n1} & x_{n2} & \cdots & x_{nd} \end{array}$$

รูปที่ 2.1 เมทริกซ์ของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ x_{11} คือ ค่าของข้อมูลในแถวที่ 1 และหลักที่ 1

x_{12} คือ ค่าของข้อมูลในแถวที่ 1 และหลักที่ 2

$\vdots$

x_{1d} คือ ค่าของข้อมูลในแถวที่ 1 และหลักที่ d

$\vdots$

x_{nd} คือ ค่าของข้อมูลในแถวที่ n และหลักที่ d

ตัวอย่าง เซตข้อมูล Iris จำนวน 150 ข้อมูลมี 4 คอลัมน์ได้แก่ Sepal Length , Sepal Width , Petal Length และ Petal Width ดังนั้นเมทริกซ์ข้อมูลของเซตข้อมูล Iris จะมีขนาด 150×4 สามารถเขียนได้ ดังรูปต่อไปนี้

	Sepal Length X_1	Sepal Width X_2	Petal Length X_3	Petal Width X_4
x_1	5.1	3.5	1.4	0.2
x_2	4.9	3	1.4	0.2
x_3	4.7	3.2	1.3	0.2
x_4	7	3.2	4.7	1.4
x_5	5.3	3.7	1.5	0.2
x_6	6.4	3.2	4.5	1.5
x_7	5	2	3.5	1
x_8	6.3	3.3	6	2.5
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{149}	7.1	3	5.9	2.1
x_{150}	6.7	2.5	5.8	1.8

รูปที่ 2.2 แสดงตัวอย่างเมทริกซ์ข้อมูลของเซตข้อมูล Iris

เมื่อ x_{11} คือ 5.1 ซึ่งอยู่ในแถวที่ 1 และหลักที่ 1

x_{12} คือ 3.5 ซึ่งอยู่ในแถวที่ 1 และหลักที่ 2

$\vdots$

x_{1503} คือ 5.8 ซึ่งอยู่ในแถวที่ 150 และหลักที่ 3

x_{1504} คือ 1.8 ซึ่งอยู่ในแถวที่ 150 และหลักที่ 4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3 ระยะทางแบบยุคลิด

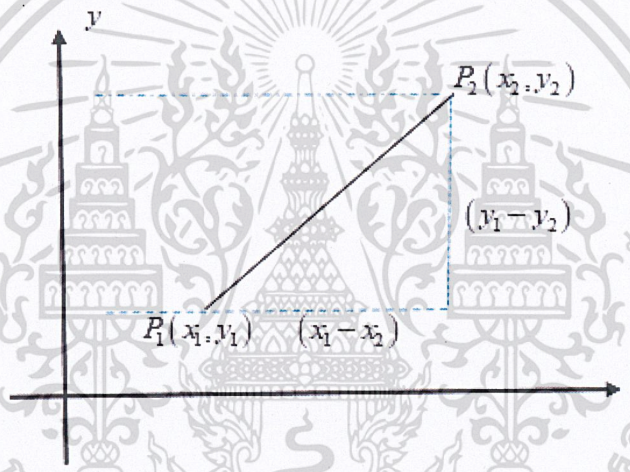
ระยะทางแบบยุคลิด (Euclidean distance) เป็นการหาระยะห่างจากรากที่สองของผลต่างระหว่างจุดสองจุดยกกำลังสอง

ดังนั้น ข้อมูลบน $\mathbb{R}^n$ สามารถหาระยะห่างแบบยุคลิด ระหว่าง จุด P_1 กับจุด P_2 ได้ดังนี้

ให้ $P_1=(x_1,x_2,...,x_n)$ และ $P_2=(y_1,y_2,...,y_n)$

$$D=\sqrt{(x_1-y_1)^2+(x_2-y_2)^2+\ldots+(x_n-y_n)^2}$$

และ ข้อมูลบน $\mathbb{R}^2$ สามารถหาระยะห่างแบบยุคลิด ระหว่าง จุด P_1 กับจุด P_2 ได้ดังรูปนี้



รูปที่ 2.3 การหาระยะทางแบบยุคลิด

จากรูปให้ $P_1=(x_1,y_1)$ และ $P_2=(x_2,y_2)$

ดังนั้น ข้อมูลบน $\mathbb{R}^2$ สามารถหาระยะห่างแบบยุคลิด ระหว่าง จุด P_1 กับจุด P_2 ได้ดังนี้

$$D=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 การแบ่งกลุ่มแบบเคมีน

วิธีแบ่งกลุ่มแบบเคมีน (K-means) จะนำข้อมูลที่มีลักษณะเหมือนกัน หรือคล้ายกัน อยู่กลุ่มเดียวกัน โดยการวัดระยะห่างจากค่าข้อมูลทุกจุดกับค่าจุดศูนย์กลางของกลุ่ม จากนั้นพิจารณาค่าระยะห่างของข้อมูลแต่ล่ะตัวเทียบกับทุกกลุ่มโดยจะเลือกค่าระยะห่างที่มีค่าน้อยที่สุด จัดไว้ในกลุ่มนั้น และคำนวณค่าจุดศูนย์กลางของกลุ่มใหม่ จะทำเช่นนี้จนกระทั่งข้อมูลไม่เปลี่ยนกลุ่มหรือค่าจุดศูนย์กลางของกลุ่มไม่เปลี่ยนแปลง ซึ่งการแบ่งกลุ่มแบบเคมีนจะต้องมีการกำหนดจำนวนกลุ่มที่ต้องการแบ่ง (K) และ จุดศูนย์กลางของกลุ่ม (Centroid) ที่เกิดจากการสุ่ม ก่อนเริ่มวิเคราะห์การจัดกลุ่ม ซึ่งถ้าสุ่มดี จำนวนการทำงานก็จะน้อย แต่ถ้าสุ่มไม่ดี จำนวนการทำงานจะมาก และ ข้อมูลขนาดใหญ่จะใช้เวลาในการคำนวณนาน นอกจากนี้ ขั้นตอนวิธีเคมีนนี้ ถือได้ว่าเป็นขั้นตอนที่มีการทำซ้ำมาก

ขั้นตอนการจัดกลุ่มแบบเคมีนมีดังนี้

ขั้นตอนที่ 1 : กำหนดค่าจุดศูนย์กลางเริ่มต้น (Centroid) และ ค่า K (จำนวนกลุ่มที่เราต้องการแบ่งกลุ่ม) ซึ่งจำนวนของค่าจุดศูนย์กลางเริ่มต้นจะเท่ากับ ค่า K

ขั้นตอนที่ 2 : คำนวณหาระยะห่างระหว่างค่าของข้อมูลทุกจุดกับจุดศูนย์กลางของกลุ่ม (Centroid) ที่กำหนดไว้ในขั้นตอนที่ 1 โดยคำนวณหาระยะทางแบบยุคลิด (Euclidean distance) ดังสมการต่อไปนี้

$$D=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$$

ขั้นตอนที่ 3 : พิจารณาค่าระยะห่างของข้อมูลแต่ละตัวที่คำนวณได้จากขั้นตอนที่ 2 เทียบกับทุกกลุ่มโดยจะเลือกค่าระยะห่างที่มีค่าน้อยที่สุด จัดไว้ในกลุ่มนั้น

ขั้นตอนที่ 4 : คำนวณหาค่าจุดศูนย์กลางของแต่ละกลุ่มใหม่ โดยใช้การคำนวณแบบคิดค่าเฉลี่ย

ขั้นตอนที่ 5 : วนกลับไปทำขั้นตอนที่ 2 และ ขั้นตอนที่ 3 จนกระทั่งข้อมูลไม่เกิดการเปลี่ยนกลุ่มหรือจุดศูนย์กลางของกลุ่มไม่เปลี่ยนแปลง จบการทำงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 ผลรวมกำลังสองของค่าความผิดพลาด

SSE (Sum of Square Error) คือ การคำนวณหาค่าความผิดพลาดระหว่างค่าของข้อมูลทุกจุดกับค่าเฉลี่ยของข้อมูล ในการวิเคราะห์ค่าผลรวมกำลังสองของค่าความผิดพลาด ดังนั้น ค่า K ที่เหมาะสม คือ จุด SSE ที่มีอัตราการเปลี่ยนแปลงมากที่สุดเมื่อเทียบกับจุดก่อนหน้า สามารถคำนวณ SSE ได้ดังสูตรนี้

$$SSE = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

2.6 วิธีการเอลโบ

วิธีการเอลโบ (Elbow Method) เป็นวิธีการตรวจสอบค่า K ว่าเหมาะสมที่จะนำค่า K นั้นไปแบ่งกลุ่มข้อมูลแบบวิธีการแบ่งกลุ่มแบบเคมีนหรือไม่ โดยจะตรวจสอบจากค่าความชันของเส้นในแนวระดับจากจุดเริ่มต้นไปยังจุดสุดท้ายและค่าความชันของเส้นโค้งจากค่า SSE สูตรในการหาความชันของเส้นแนวระดับคือ

$$m = \frac{\Delta y}{\Delta x}$$

เมื่อ m คือค่าของความชัน

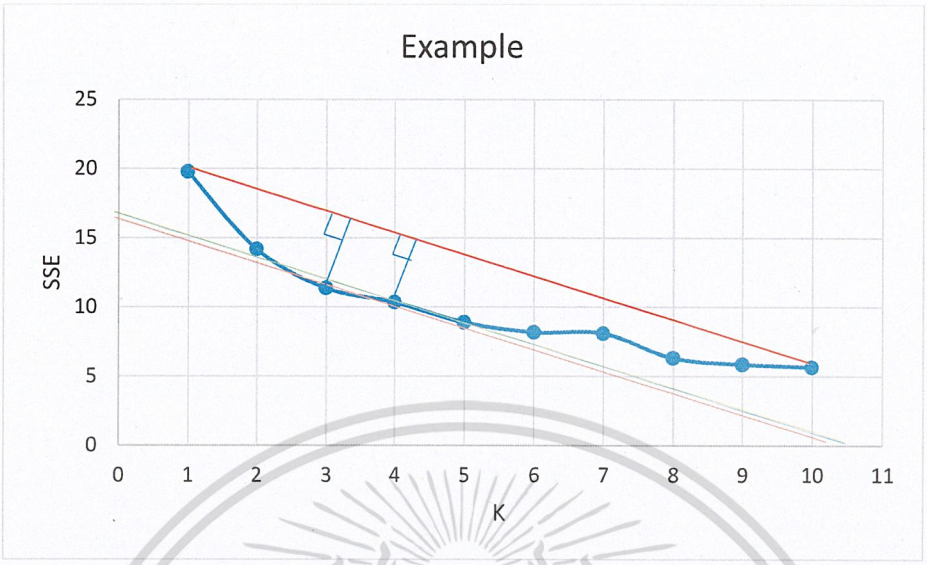
Δy คือผลต่างของค่า SSE จุดแรกและจุดสุดท้าย

Δx คือผลต่างของค่า K โดยที่ K =1 จนถึง K =10

เมื่อคำนวณหาค่าความชันจากนั้นหาผลต่างของความชันแต่ละจุดกับความชันของเส้นแนวระดับเมื่อค่าความชันมีการเปลี่ยนแปลงจากค่าลบเป็นค่าบวกแสดงว่าจุดนั้นคือจุดเปลี่ยนเว้าของเส้นความชันนั่นคือค่า K ที่เหมาะสมของเซตข้อมูลนั้นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างการคำนวณเอลโบ



คำนวณความชันเส้นแนวระดับจากสูตร $m = \frac{\Delta y}{\Delta x}$

$$m = \frac{5.6614 - 19.77}{10 - 1} = -1.57$$

จากนั้นคำนวณค่าความชันของแต่ละจุดดังตาราง

จำนวนกลุ่ม (K)	ค่า SSE	ความชัน	ผลต่างของความชันกับเส้นแนวระดับ
1	19.77677	-	-
2	14.17918	-5.59759	-4.02759
3	11.36477	-2.81441	-1.24441
4	10.34039	-1.02438	0.54562
5	8.89384	-1.44655	0.12345
6	8.17527	-0.71857	0.85143
7	8.09622	-0.07905	1.49095
8	6.32088	-1.77534	-0.20534
9	5.86105	-0.45983	1.11017
10	5.6614	-0.19965	1.37035

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางข้างต้นจะเห็นได้ว่าเมื่อ K =3 ผลต่างค่าความชันมีค่าเป็นลบและเมื่อ K =4 ผลต่างค่าความชันมีค่าเป็นบวกแสดงว่า K =3 คือจุดเปลี่ยนเว้าความชัน ดังนั้นค่า K ที่เหมาะสม คือ K =3

2.7 เซตข้อมูลที่ใช้ในการทดลอง

2.7.1 เซตข้อมูลดอกไอริส (Iris-setosa)

รูปที่ 2.4 ดอกไอริส (Iris)

เซตข้อมูลไอริส (Iris) เป็นเซตข้อมูลเกี่ยวกับการจำแนกชนิดของดอก iris ซึ่งประกอบไปด้วยจำนวนทั้งหมด 150 ข้อมูลและคุณลักษณะภายนอก 4 ลักษณะของดอก iris

1) ความยาวกลีบเลี้ยง (sepal length)
2) ความกว้างกลีบเลี้ยง (sepal width)
3) ความยาวกลีบดอก (petal length)
4) ความกว้างกลีบดอก (petal width)

โดยในฐานข้อมูลนี้ได้จำแนกชนิดของดอก iris ออกเป็น 3 ชนิด ได้แก่ Iris-Setosa, Iris-Versicolor และ Iris-Virginia ซึ่งลักษณะของดอก Iris ทั้ง 3 ชนิดจะมีลักษณะคล้ายคลึงกัน

ที่มา https://archive.ics.uci.edu/ml/datasets/iris

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.2 เซตข้อมูลไวน์ (Wine)

เซตข้อมูลไวน์เป็นเซตข้อมูลเกี่ยวผลของการวิเคราะห์ทางเคมีของไวน์ที่ปลูกในภูมิภาคเดียวกันในอิตาลีแต่ได้มาจากสามพันธุ์ที่แตกต่างกัน ซึ่งวิเคราะห์จาก 13 องค์ประกอบในแต่ละประเภทของไวน์ ได้แก่

1) Alcohol	2) Malic acid	3) Ash
4) Alcalinity of ash	5) Total phenols	6) Magnesium
7) Flavonoid	8) Neoflavanoid phenols	9) Hue
10) Color intensity	11) OD280/OD315 of diluted wines	
12) Proline	13) Proanthocyanins	

ที่มา http://archive.ics.uci.edu/ml/datasets/wine

2.7.3 ฐานข้อมูลไฟป่า (Forest Fires)

เป็นเซตข้อมูลที่มีวัตถุประสงค์เพื่อทำนายพื้นที่ ที่เกิดไฟไหม้ป่าในภาคตะวันออกเฉียงเหนือของโปรตุเกสโดยใช้ข้อมูลทางอุตุนิยมวิทยาและข้อมูลอื่น ๆ ส่วนประกอบที่ใช้ในการทำนายมีดังนี้

1. X - แกนพิกัดเชิงพื้นที่ภายในแผนที่ Montesinho Park: 1 ถึง 9
2. พิกัดเชิงพื้นที่ y - แกนแกนภายในแผนที่ Montesinho Park: 2 ถึง 9
3. เดือน - เดือนของปี: 'Jan' ถึง 'Dec'
4. วัน - วันของสัปดาห์: 'mon' to 'sun'
5. FFMC - ดัชนี FFMC จากระบบ FWI: 18.7 ถึง 96.20
6. DMC - ดัชนี DMC จากระบบ FWI: 1.1 ถึง 291.3
7. ดัชนี DC - DC จากระบบ FWI: 7.9 ถึง 860.6
8. ดัชนี ISI - ISI จากระบบ FWI: 0.0 ถึง 56.10
9. อุณหภูมิ - องศาเซลเซียส: 2.2 ถึง 33.30
10. RH - ความชื้นสัมพัทธ์ใน%: 15.0 ถึง 100
11. ความเร็วลม - ความเร็วลมเป็นกิโลเมตร / ชม.: 0.40 ถึง 9.40

ที่มา https://archive.ics.uci.edu/ml/index.php

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3

ขั้นตอนวิธีเลือกค่าเริ่มต้น

3.1 วิธีที่จะศึกษา

จากขั้นตอนการคำนวณเคมีนในบทที่ 2 สามารถเขียนเป็นคำสั่งได้ดังรูป

```
for (int i=0; i< _dataSize; i++){
     x=_data[i].x;
         y=_data[i].y;
     for (int c=0; c<_numClusters; c++){
         centroidsDist[c]=Math.sqrt(Math.pow(x-_centroids[c].x,2)
                 +Math.pow(y-_centroids[c].y,2));
         }
     min= centroidsDist[0];
     mincluster=0;
     for (int c=1; c<_numClusters; c++){
         if (min>centroidsDist[c]){
             min=centroidsDist[c];
             mincluster=c;
            }
         }
    _data[i].label= mincluster;
```

รูปที่ 3.1 คำสั่งในการคำนวณเคมีน

เมื่อ _data[i].x และ _data[i].y คือคู่อันดับของข้อมูลที่ถูกเก็บในอาเรย์ และ _data[i].label คือ แสดงว่าข้อมูลนั้นอยู่กลุ่มไหน

จากรูปที่ 3.1 คำสั่งในการคำนวณเคมีนโดยคำนวณหาค่าระยะห่างระหว่างข้อมูลแต่ละข้อมูลกับจุดศูนย์กลางเริ่มต้นที่ได้เลือกมานั้น คำนวณระยะห่างโดยวิธียุคลิดและเทียบว่าระยะห่างจากกลุ่มไหนน้อยที่สุดจัดไว้กลุ่มนั้น เมื่อทำการคำนวณแล้วจะทราบจำนวนรอบในการคำนวณของการเลือกค่าเริ่มต้นแต่ละวิธี ก่อนที่จะเริ่มใช้คำสั่งในการคำนวณเคมีนจะต้องมีการกำหนดจุดศูนย์กลางเริ่มต้นที่ใช้ในการคำนวณ โดยจะศึกษาวิธีที่จะเลือกค่าเริ่มต้นในการคำนวณดังวิธีต่อไปนี้เพื่อลดจำนวนรอบการคำนวณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1 วิธีการหาค่าเริ่มต้นของกลุ่มแบบสุ่ม

สุ่มค่าเริ่มต้นในการคำนวณโดยใช้คำสั่ง Math.random ดังรูป

```
private double [] _data,_datay;

while ( i<_numClusters ){
      c = (int) (Math.random()*_dataSize);

      _centroids[i] = _data[c];
          _centroidsy[i] = _datay[c];
```

รูปที่ 3.2 การใช้คำสั่งในการสุ่มจุดศูนย์กลางเริ่มต้นในการคำนวณ

เมื่อ _NumClusters คือค่า K ที่เหมาะสมของแต่ละฐานข้อมูล และ _dataSize คือจำนวนข้อมูลในแต่ละฐานข้อมูล

จากรูปที่ 3.2 คำสั่ง Math.random จะส่งค่าที่เป็นประเภท double ที่มีค่าเป็นบวกซึ่งจะมีค่ามากกว่าหรือเท่ากับ 0 แต่น้อยกว่า 1 จากกคำสั่งในรูปที่ 3.7 ตัวอย่างเช่น กำหนดให้จำนวนข้อมูล (_dataSize) เท่ากับ 150 การทำงานของคำสั่ง random คือ ถ้าสุ่มได้ค่า 0.59*150=88.500 จากนั้นคำสั่ง int จะแปลงค่าที่ได้เป็นจำนวนเต็มจะได้เท่ากับ 88 จึงนำค่าที่ได้เป็นตำแหน่งของข้อมูลซึ่งเก็บในอาเรย์ที่มีตำแหน่ง 0-149 จะได้จุดศูนย์กลางเริ่มต้นที่มีจำนวนเท่ากับจำนวนกลุ่มจากนั้นนำไปทำการคำนวณเคมีนเพื่อที่จะทราบจำนวนรอบการคำนวณ

3.1.2. การหาค่ากลางเริ่มต้นของกลุ่มโดยการหาค่ากึ่งกลางของช่วง

กึ่งกลางของช่วง หรือค่ากลางคือตำแหน่งตรงกลางของข้อมูลที่มีการเรียงจากน้อยไปมาก วิธีการคำนวณแบบกึ่งกลางของช่วง ของแต่ล่ะฐานข้อมูลบน $\mathbb{R}^2$ เกิดจากแนวคิดที่คาดว่าจุดศูนย์กลางของข้อมูลจะอยู่อยู่ตำแหน่งกลางของแต่ละช่วงข้อมูล ภายใต้สมมติฐานข้อมูลที่มีการกระจายสม่ำเสมอ มีขั้นตอนดังนี้

กำหนดให้ n = จำนวนข้อมูล

K = จำนวนกลุ่มที่ต้องการแบ่ง

max = ค่าของข้อมูลที่มากที่สุด

min = ค่าของข้อมูลที่น้อยที่สุด

$interval$ = ช่วงความกว้างของข้อมูล

$median$ = จุดศูนย์กลางเริ่มต้นในการคำนวณเคมีน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 1. หาความกว้างของช่วงข้อมูลโดยแบ่งตามจำนวนกลุ่มที่ต้องการแบ่ง

$$interval = \frac{\max - \min}{K}$$

ขั้นตอนที่ 2. เมื่อทราบความกว้างของแต่ละช่วงจากนั้นหาค่ากึ่งกลางของแต่ละช่วง

$$median = \frac{\max - \min}{2}$$

ขั้นตอนที่ 3. เมื่อได้ค่า *median* ในช่วงแรก ผลรวมของค่า *median* ในช่วงแรกและ *interval* จะได้ค่า *median* ในช่วงถัดไป

จากขั้นตอนดังกล่าวสามารถเขียนเป็นคำสั่งได้ดังรูป

```
double minx=_data[0].x;
double maxx=_data[0].x;
double miny=_data[0].y;
double maxy=_data[0].y;
for (int i=1; i< _dataSize; i++){
            if (maxx<_data[i].x)  maxx=_data[i].x;
            if (minx>_data[i].x)  minx=_data[i].x;
            if (maxy<_data[i].y)  maxy=_data[i].y;
            if (miny>_data[i].y)  miny=_data[i].y;
        }
double intervalx= (maxx-minx)/_numClusters;
    double lowerx=minx;
double intervaly= (maxy-miny)/_numClusters;
    double lowery=miny;
int []centroidsfrom = new int[_numClusters];
for (int k=0; k<_numClusters; k++){
    _centroids[k].x = lowerx+(intervalx)/2;
            lowerx=lowerx+intervalx;
    _centroids[k].y = lowery+(intervaly)/2;
            lowery=lowery+intervaly;
            }
```

รูปที่ 3.3 การใช้คำสั่งในการคำนวณแบบ กึ่งกลางของช่วง

จากรูปที่ 3.3 จะได้จุดศูนย์กลางเริ่มต้นในการคำนวณเคมีน โดยการเลือกจุดศูนย์กลางเริ่มต้นแต่ละจุดของแต่ละเซตข้อมูลจากนั้นนำจุดที่คำนวณได้ไปคำนวณเคมีนโดยใช้คำสั่งในรูปที่ 3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างการคำนวณค่าเริ่มต้นวิธีค่ากึ่งกลางของช่วง บน $\mathbb{R}^2$ ของข้อมูลจำนวน 10 ข้อมูล เมื่อ K =3

ID	X	Y
1	2	10
2	2	5
3	8	4
4	5	8
5	7	5
6	6	4
7	1	2
8	4	9
9	3	6
10	5	7

จากข้อมูลจะเห็นได้ว่าค่า $\text{Min}_x = 1$ $\text{Max}_x = 8$

และ $\text{Min}_y = 2$ $\text{Max}_y = 10$

กำหนดให้ ค่า k = 3 ; $\text{interval}_x = \frac{\text{max}_x - \text{min}_x}{k} = \frac{8-1}{3} = 2.33$

$\text{interval}_y = \frac{\text{max}_y - \text{min}_y}{k} = \frac{10-2}{3} = 2.67$

จากการคำนวณข้างต้น

จะได้ความกว้างของข้อมูลแกน X ช่วงที่ 1 คือ $\text{Min}_x + \text{interval}_x$ = 1+2.33 =3.33

จะได้ความกว้างของข้อมูลแกน X ช่วงที่ 2 คือ $\text{Lower}_x + \text{interval}_x$ = 3.33+2.33=5.66

จะได้ความกว้างของข้อมูลแกน X ช่วงที่ 3 คือ $\text{Lower}_x + \text{interval}_x$ = 5.66+2.33=7.99

และ จะได้ความกว้างของข้อมูลแกน y ช่วงที่ 1 คือ $\text{Min}_y + \text{interval}_y$ = 2+2.67 = 4.67

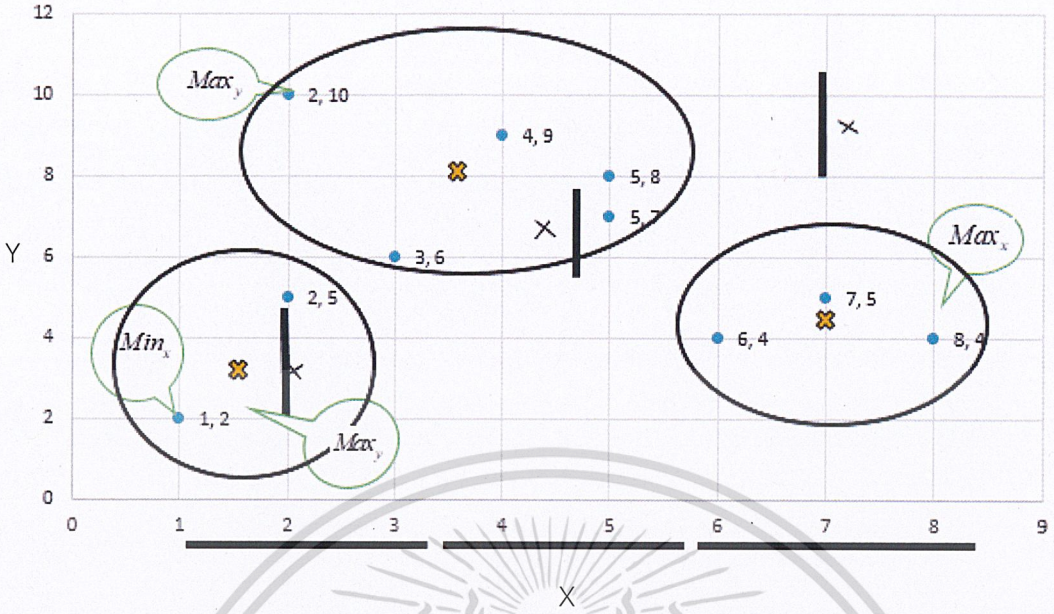
จะได้ความกว้างของข้อมูลแกน y ช่วงที่ 2 คือ $\text{Lower}_y + \text{interval}_y$ = 4.67+2.67=7.34

จะได้ความกว้างของข้อมูลแกน y ช่วงที่ 3 คือ $\text{Lower}_y + \text{interval}_y$ = 7.34+2.67 = 10.01

ค่า Median ในช่วงแรกคือ $\text{Min}_x + \frac{3.33-1}{2} = 2.16$, $\text{Min}_y + \frac{4.67-2}{2} = 3.33$

ค่าเริ่มต้นที่คำนวณได้คือ (2.16,3.33) (4.49,6.00) และ (6.82,8.67) จากนั้นนำไปคำนวณเคมีนได้รอบการคำนวณเคมีนเท่ากับ 2 รอบ จุดศูนย์กลางสุดท้ายคือ (1.50,3.50) (3.80,8.00) และ (7.00,4.33)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.4 แสดงการหาจุด Median ของแต่ล่ะช่วงของฐานข้อมูล

เมื่อ x คือ จุดศูนย์กลางเริ่มต้นในการคำนวณแบบ กึ่งกลางของช่วงและ

⊠ คือ จุดศูนย์กลาสุดท้ายของฐานข้อมูล

3.1.3 การหาค่ากลางเริ่มต้นของกลุ่มโดยการคำนวณแบบสุ่มหลายค่า

จากข้อเสียของการสุ่มจุดศูนย์กลางเริ่มต้นของวิธีเคมีนคือ อาจจะมีการสุ่มจุดที่อยู่ในกลุ่มเดียวกันทำให้มีจำนวนรอบของการคำนวณเคมีนมาก วิธีการนี้เป็นการนำค่าเฉลี่ยของข้อมูลมาเป็นค่ากลางเริ่มต้นในการคำนวณเคมีน ภายใต้ข้อสันนิษฐานว่าค่าเฉลี่ยของข้อมูลหลายๆข้อมูลอาจจะเป็นตัวแทนของกลุ่มหรือค่าเฉลี่ยที่คำนวณได้อาจจะใกล้เคียงจุดศูนย์กลางสุดท้ายโดยจะสุ่มจาก 10% ของจำนวนข้อมูลทั้งหมดของแต่ละฐานข้อมูลบน $\mathbb{R}^2$ มีขั้นตอนดังนี้

กำหนดให้ n = จำนวนข้อมูล

K = จำนวนกลุ่มที่ต้องการแบ่ง

$TotalMember$ = จำนวนที่สุ่มขึ้นมาทั้งหมด

$Each$ = จำนวนข้อมูลที่สุ่มของแต่ละกลุ่ม

$interval$ = ระยะของตำแหน่งที่ถูกสุ่มของข้อมูลแต่ละตัวที่ถูกสุ่ม

$centroid$ = ค่าเริ่มต้นในการคำนวณเคมีน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 1 คำนวณหาจำนวนข้อมูลโดยคำนวณจาก 10 % ของจำนวนข้อมูลทั้งหมด

$$TotalMember = \frac{10}{100} \times n$$

ขั้นตอนที่ 2 เมื่อทราบจำนวนข้อมูลที่จะสุ่มจากนั้นนำไปแบ่งตามกลุ่ม จากนั้นจะทราบว่าแต่ละกลุ่มต้องสุ่มข้อมูลกลุ่มละกี่ข้อมูล

$$Each = \frac{\text{TotalMember}}{\text{k}}$$

ขั้นตอนที่ 3 หาค่าเฉลี่ยของผลรวมข้อมูลในแต่ละช่วงหารด้วยจำนวนข้อมูลที่สุ่มได้ในแต่ละกลุ่ม

$$\text{centroid} = \frac{\sum_{i=1}^{\text{each}} \text{interval}_i}{\text{each}}$$

ขั้นตอนที่ 4 ทำการคำนวณเคมีน วัดระยะห่างระหว่างข้อมูลทุกตัวกับจุดศูนย์กลางเริ่มต้นที่หาได้ในขั้นตอนที่

จากขั้นตอนวิธีการข้างต้นสามารถนำมาเขียนเป็นคำสั่งได้ดังรูปต่อไปนี้

```
int each =a/_numClusters;
a=  (int) (0.1*_dataSize);
for (int j=0; j<_numClusters; j++){
     int c=0;
     int count=0;
     Double sumX=0.0; Double sumY; sumY = 0.0;
     Double AvgX=0.0; Double AvgY=0.0;
   for(int i=0; i<each; i++){
        c = (int) (Math.random()*_dataSize);
        outFile.printf("%4d %5.3f %5.3f",c, _data[c].x,_data[c].y);
        sumX=sumX+_data[c].x;
        sumY=sumY+_data[c].y;
        count++;
   }
    _centroids[j].x = sumX/count;
   _centroids[j].y = sumY/count;
```

รูปที่ 3.5 การใช้คำสั่งในการคำนวณแบบสุ่มหลายค่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.5 ใช้คำสั่ง Math.random ให้ทำงานสุ่มตามจำนวนครั้งเท่ากับจำนวนข้อมูลที่ต้องการสุ่ม จากนั้นคำนวณเคมีนโดยคำนวณหาค่าระยะห่างระหว่างข้อมูลแต่ละข้อมูลกับจุดศูนย์กลางเริ่มต้นที่ได้คำนวณมานั้น โดยคำนวณค่าระยะห่างโดยวิธียุคลิดและเทียบว่าห่างจากกลุ่มไหนน้อยที่สุดจะได้อยู่กลุ่มนั้น เมื่อทำการคำนวณแล้วจะทราบจำนวนรอบในการคำนวณของการสุ่มแบบสุ่มหลายตัว

ตัวอย่างการคำนวณสุ่มหลายค่า

เมื่อมีข้อมูล 20 ข้อมูลบน $\mathbb{R}^2$ และ K =2

ข้อมูล	A	B
1	22	35
2	15	20
3	21	43
4	78	41
5	34	56
6	23	21
7	42	41
8	13	65
9	65	34
10	55	75
11	22	54
12	14	32
13	15	78
14	28	54
15	34	32
16	20	21
17	42	12
18	56	30
19	11	40
20	18	37

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สุ่มจำนวน 10% ของจำนวนข้อมูลดังนั้นจึงสุ่มทั้งหมด 2 ค่าซึ่งแต่ละกลุ่มจะสุ่มเพียง 1 ค่า

$Cluster_1$			$Cluster_2$		
ตำแหน่งที่ถูกสุ่ม	A	B	ตำแหน่งที่ถูกสุ่ม	A	B
3	21	34	2	15	20
7	42	41	11	22	54

จากนั้นหาผลรวมค่าของข้อมูลในตำแหน่งที่ถูกสุ่มและหาค่าเฉลี่ยจะได้เท่ากับ ค่าเฉลี่ยของ $Cluster_1$ เท่ากับ $\frac{21+42}{2}=31.50$ และ $\frac{34+41}{2}=37.50$ ดังนั้นจุดศูนย์กลางเริ่มต้นของ $Cluster_1$ คือ (31.50,37.50) Cluster 2 คิดเช่นเดียวกับ $Cluster_1$ จุดศูนย์กลางเริ่มต้นของ $Cluster_2$ คือ (18.50,37.00) มีการคำนวณเคมีนทั้งหมด 4 รอบและได้จุดศูนย์กลางสุดท้ายของกลุ่มคือ (40.8,52.8) และ (22.0,29.3)

3.1.4 การหาค่ากลางเริ่มต้นของกลุ่มโดยใช้ความถี่สูงสุดของข้อมูล

วิธีการเลือกค่าเริ่มต้นจากความถี่ของข้อมูลที่สูงสุดบน $\mathbb{R}^2$ ความถี่ของข้อมูลเกิดจากมีข้อมูลที่เหมือนกันและซ้ำกันซึ่งคาดว่าการเลือกค่าเริ่มต้นจากกันซ้ำกันของข้อมูลนั้นค่าที่เลือกมาอาจจะใกล้เคียงกับจุดศูนย์กลางจริงเพื่อลดจำนวนรอบของการคำนวณเคมีน มีขั้นตอนดังนี้

1) สำรวจข้อมูลคำนวณความถี่ของข้อมูลที่ซ้ำกัน
2) เลือกจุดที่มีความถี่สูงสุดและรองลงมาเป็นค่าเริ่มต้นในการคำนวณเคมีน
3) คำนวณเคมีนโดยใช้คำสั่งในรูปที่ 3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
public void tally() {
    int jump = 0;
    DATA[] _newData = {};
    for (int j = 0; j <= _dataSize; j = j + 1) {
        int index = findIndex(_newData, _data[j]);
        if (index == -1) {
            _newData = addDATA(_newData, _data[j]);
            _newData[_newData.length - 1].count = 1;
        } else {
            _newData[index].count = _newData[index].count + 1;
        }
    }
    _data = _newData;
}
public DATA[] addDATA(DATA[] arr, DATA add) {
    DATA[] _newArr = new DATA[arr.length + 1];
    for (int i = 0; i < arr.length; i++) {
        _newArr[i] = arr[i];
    }
    _newArr[arr.length] = add;
    return _newArr;
}

public void sortArray() {//go through the array and sort from high to small
   DATA temp;
   int jump;
   for (int i = 0; i < _data.length; i = i + jump) {
     jump = 1;
     for (int j = 1; j < (_data.length - i); j++) {
      if (_data[j - 1].count < _data[j].count) {
          temp = _data[j - 1];
          _data[j - 1] = _data[j];
          _data[j] = temp;
      if(_data[j].x == _data[j + jump].x && _data[j].y == _data[j + jump].y){
                jump++;
                 }
              }
          }
      }
   for (int j = 0; j < _data.length; j++) {
      if (_data[j].count != 0) {
          outFile.printf("%5.3f %5.3f %4d \n"
                  , _data[j].x, _data[j].y, _data[j].count);
          }
      }
```

รูปที่ 3.6 คำสั่งคำนวณหาความถี่ของข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

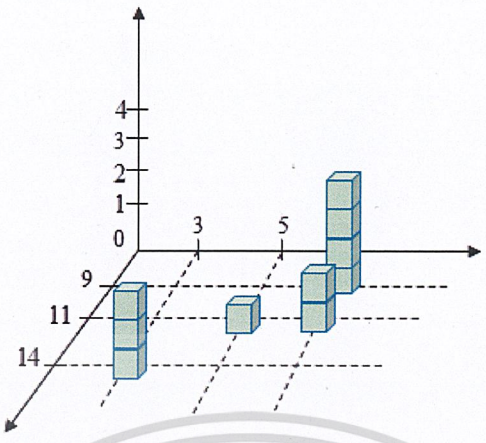
ตัวอย่างข้อมูลบน $\mathbb{R}^2$ ของข้อมูลจำนวน 20 ข้อมูล เมื่อค่า K =2

ข้อมูลตัวที่	X	Y
1	14	5
2	14	3
3	14	3
4	14	3
5	11	6
6	11	6
7	9	6
8	9	6
9	9	6
10	9	6

ตารางแสดงความถี่ของข้อมูลที่ซ้ำกัน

X	Y	ความถี่
9	6	4
11	6	2
14	3	3
14	5	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.7 แสดงความถี่ของข้อมูล

ค่าเริ่มต้นในการคำนวณเคมีนที่เลือกจากความถี่ที่มากสุดและรองลงมา คือ (9,6) และ (14,3) นำค่าที่เลือกไป คำนวณเคมีน เกิดการคำนวณเคมีน 2 รอบ และได้จุดศูนย์กลางสุดท้ายคือ (11.17,4.33) และ (9,6)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

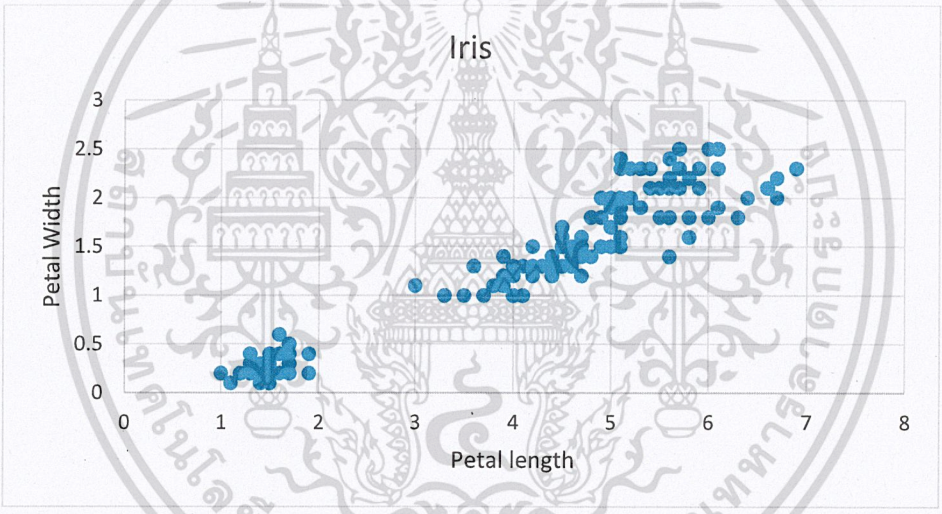
บทที่ 4

ผลการศึกษา

จากการทดลองการเลือกจุดศูนย์กลางเริ่มต้นในแบบต่างๆในบทที่ 3 โดยทดลองกับเซตข้อมูล 3 เซตข้อมูลคือ 1) เซตข้อมูลไอริส (Iris) 2) เซตข้อมูลไวน์ (Wine) และ 3) เซตข้อมูลไฟป่า (Forest Fires) ได้ผลดังนี้

4.1 เซตข้อมูลไอริส

เซตข้อมูลไอริส (Iris) จำนวน 150 ข้อมูลโดยผู้ศึกษาจะเลือกใช้แอตทริบิวต์ Petal Length (ความยาวกลีบดอก) และ Petal Width (ความกว้างกลีบดอก) เนื่องจากลักษณะข้อมูลของทั้งสองแอตทริบิวต์สามารถแสดงดังรูปที่ 4.1

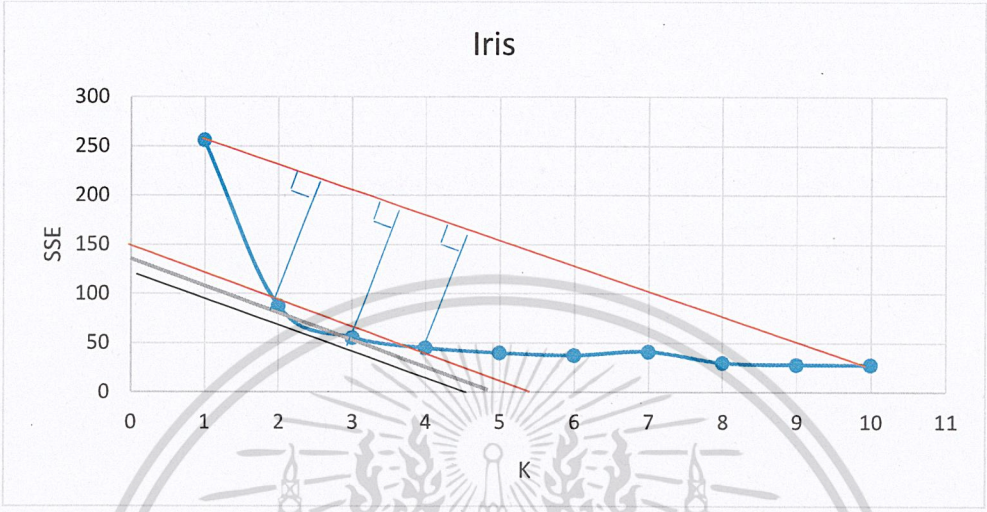


รูปที่ 4.1 กราฟแสดงข้อมูล Petal length-Petal width ของเซตข้อมูล Iris

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.1 จำนวนกลุ่มที่เหมาะสมของเซตข้อมูล Iris

คำนวณค่า SSE ของเซตข้อมูล Iris เมื่อ K =1 ถึง K =10 และคำนวณหาจุดเปลี่ยนเว้าจากค่าความชันของเส้นแนวระดับและแต่ละจุด K



รูปที่ 4.2 กราฟแสดงค่า SSE ของเซตข้อมูล Iris

คำนวณค่าความชันของเส้นแนวระดับจะได้เท่ากับ

$$m = \frac{\Delta y}{\Delta x}$$

$$m = \frac{27.649 - 256.104}{10 - 1}$$

$$m = -25.383$$

จากนั้นคำนวณความของแต่ละจุดและหาผลต่างของความชันที่จุดนั้นๆและความชันของเส้นแนวระดับดังตาราง

ตารางที่ 4.1 แสดงค่าความชันของแต่ละจุด

จำนวนกลุ่ม (K)	ค่า SSE	ความชัน	ผลต่างของความชันกับเส้นแนวระดับ
1	256.104	-	-
2	87.58	-168.524	-143.141
3	55.314	-32.266	-6.883

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จำนวนกลุ่ม (K)	ค่า SSE	ความชัน	ผลต่างของความชันกับเส้นแนวระดับ
4	44.883	-10.431	14.952
5	39.902	-4.981	20.402
6	37.531	-2.371	23.012
7	41.231	3.7	29.083
8	30.001	-11.23	14.153
9	28.108	-1.893	23.49
10	27.649	-0.459	24.924

จากตารางที่ 4.1 จะเห็นได้ว่าเมื่อ K =3 มีค่าความชันเป็นลบและเมื่อ K =4 ค่าความชันเป็นบวกแสดงว่าค่า K ที่เหมาะสมของเซตข้อมูล Iris คือ K =3

4.1.2 วิธีการหาค่าเริ่มต้นของกลุ่มโดยการสุ่มของเซตข้อมูล Iris

จากในบทที่ 3 โดยวิธีการสุ่มจุดศูนย์กลางเริ่มต้นโดยการสุ่มจำนวน 1000 ครั้งได้ผลการทดลองของเซตข้อมูล Iris ซึ่งจะแสดงจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายและรอบในการคำนวณดังนี้

ตารางที่ 4.2 แสดงผลการสุ่ม 1000 ครั้ง

รอบ	จุดศูนย์กลาง	$Cluster_1$			$Cluster_2$			$Cluster_3$		
		petal length	petal width	ตำแหน่ง	petal length	petal width	ตำแหน่ง	petal length	petal width	ตำแหน่ง
1	จุดศูนย์กลางเริ่มต้น	3.3	1	58	5.9	2.1	103	1.6	0.2	30
	จุดศูนย์กลางสุดท้าย	4.26923	1.34231		5.59583	2.0375		1.464	0.244	
	จำนวนรอบ	6								
2	จุดศูนย์กลางเริ่มต้น	4.8	1.8	127	4.2	1.3	95	1.6	0.6	44
	จุดศูนย์กลางสุดท้าย	5.59583	2.0375		4.26923	1.34231		1.464	0.244	
	จำนวนรอบ	6								
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รอบ	จุดศูนย์กลาง	$Cluster_1$			$Cluster_2$			$Cluster_3$		
		petal length	petal width	ตำแหน่ง	petal length	petal width	ตำแหน่ง	petal length	petal width	ตำแหน่ง
997	จุดศูนย์กลางเริ่มต้น	3.7	1	82	1.4	0.1	13	5.3	2.3	116
	จุดศูนย์กลางสุดท้าย	4.26923	1.34231		1.464	0.244		5.59583	2.0375	
	จำนวนรอบ	6								
988	จุดศูนย์กลางเริ่มต้น	4.8	1.4	77	1.3	0.2	3	3.7	1	82
	จุดศูนย์กลางสุดท้าย	5.59583	2.0375		1.464	0.244		4.26923	1.34231	
	จำนวนรอบ	7								
999	จุดศูนย์กลางเริ่มต้น	4.4	1.4	66	4.8	1.8	139	5.8	2.2	105
	จุดศูนย์กลางสุดท้าย	1.464	0.244		4.29259	1.35926		5.62609	2.04783	
	จำนวนรอบ	6								
1000	จุดศูนย์กลางเริ่มต้น	4.5	1.5	85	4	1.3	72	4.8	1.8	127
	จุดศูนย์กลางสุดท้าย	4.29259	1.35926		1.464	0.244		5.62609	2.04783	
	จำนวนรอบ	9								

จากผลข้างต้นสามารถสรุปจำนวนรอบที่ดีที่สุด จำนวนรอบที่แย่ที่สุดและค่าเฉลี่ยของการทำงาน 1000 ครั้งดังตาราง

ตารางที่ 4.3 แสดงจำนวนรอบการคำนวณโดยการสุ่ม 1000 ครั้งของเซตข้อมูล Iris

เซตข้อมูล	จำนวนรอบที่ดีที่สุด	จำนวนรอบที่แย่ที่สุด	จำนวนรอบเฉลี่ย
Iris	2	18	8.34

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3 การหาค่ากลางเริ่มต้นของกลุ่มโดยการหาค่ากึ่งกลางของช่วงของเซตข้อมูล Iris

จากการคำนวณหาค่ากึ่งกลางของช่วงโดยใช้คำสั่งในรูปที่ 3.5 ได้ผลการทดลองของเซตข้อมูล Iris ดังนี้

1) หาความกว้างของข้อมูลแต่ละช่วงของแกน x

$$interval_x = \frac{\max_x - \min_x}{K} = \frac{6.9-1}{3} = 1.96$$

ดังนั้นแต่ละช่วงจะมีความกว้างเท่ากับ 1.96 ดังนั้นจุดแบ่งช่วงแรกคือ 1+1.96=2.96 ช่วงถัดไปคือ 4.92 และช่วงสุดท้ายคือ 6.92

หาความกว้างของข้อมูลแต่ละช่วงของแกน y

$$interval_y = \frac{\max_y - \min_y}{K} = \frac{2.5-0.1}{3} = 0.8$$

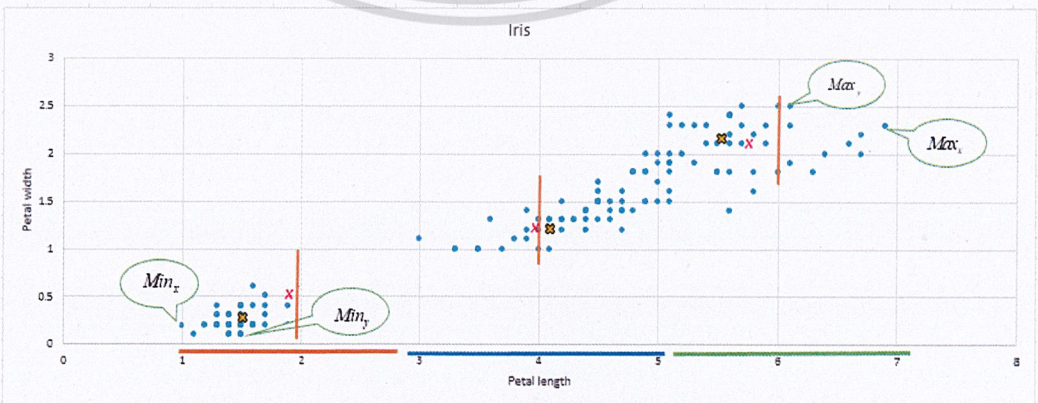
ดังนั้นแต่ละช่วงจะมีความกว้างเท่ากับ 0.8 ดังนั้นจุดแบ่งช่วงแรกคือ 0.1+0.8=0.9 ช่วงถัดไปคือ 1.7 และช่วงสุดท้ายคือ 2.5

2) หาค่ากึ่งกลางของแต่ละช่วงของแกน x โดยนำค่า $\frac{\max - \min}{2} + \min$ ของแต่ละช่วงทั้งแกน x และแกน y

ดังนั้นค่า *median* ช่วงแรกของแกน x คือ $\frac{2.96-1}{2} = 0.98+1 = 1.98$ และ

median ช่วงแรกของแกน y คือ $\frac{0.9-0.1}{2} = 0.4+0.1 = 0.5$

ดังนั้นค่าเริ่มต้นในการคำนวณของช่วงแรกคือ (1.98,0.5) ช่วงที่สองและช่วงสุดท้ายคิดเช่นเดียวกันจะได้ (3.95,1.3) และ (5.59,2.10) แสดงดังรูป



รูปที่ 4.3 แสดงการหาจุดกึ่งกลางของแต่ละช่วงของเซตข้อมูล Iris

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ x คือจุดศูนย์กลางเริ่มต้นในการคำนวณแบบกึ่งกลางของช่วง และ

❌ คือจุดศูนย์กลางสุดท้ายของเซตข้อมูล

จากการคำนวณข้างต้นสามารถสรุปค่าจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายและจำนวนรอบการคำนวณเคมีนแสดงดังตาราง

ตารางที่ 4.4 แสดงจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายและจำนวนรอบของการคำนวณเคมีนของเซตข้อมูล Wine

รอบ	จุดศูนย์กลาง	$Cluster_1$		$Cluster_2$		$Cluster_3$	
		Petal length	Petal width	Petal length	Petal width	Petal length	Petal width
1	จุดศูนย์กลางเริ่มต้น	1.98333	0.5	3.95	1.3	5.91667	2.1
	จุดศูนย์กลางสุดท้าย	1.462	0.246	4.26923	1.34231	5.59583	2.0375
	จำนวนรอบ	2					

4.1.4 การหาค่ากลางเริ่มต้นของกลุ่มโดยการคำนวณสุ่มหลายค่า

การคำนวณเคมีน โดยใช้วิธีการคำนวณแบบสุ่มหลายค่า โดยจะนำค่าเฉลี่ยของข้อมูลมาเป็นค่ากลางเริ่มต้นในการคำนวณเคมีนโดยจะสุ่มจาก 10% ของจำนวนข้อมูลทั้งหมด คำนวณจากในบทที่ 3 ได้ผลออกมาดังนี้

ตารางที่ 4.5 แสดงผลการสุ่มของวิธีสุ่มแบบหลายค่าและจำนวนรอบการคำนวณเคมีนของฐานข้อมูล Iris

เซตข้อมูล	Iris
ค่า K	3
จำนวนข้อมูล	150
จำนวนที่ต้องสุ่มทั้งหมด	15
จำนวนที่สุ่มแต่ละกลุ่ม	5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 แสดงจุดศูนย์กลางและจำนวนรอบในการคำนวณเคมีนของวิธีการสุ่มหลายค่าจากการสุ่มจำนวน 1000 ครั้ง

รอบ	จุดศูนย์กลาง	$Cluster_1$		$Cluster_2$		$Cluster_3$	
		Petal length	Petal width	Petal length	Petal width	Petal length	Petal width
1	จุดศูนย์กลางเริ่มต้น	3.36	0.96	4.4	1.4	4.24	1.36
	จุดศูนย์กลางสุดท้าย	1.464	0.2439	5.595	2.037	4.269	1.342
	จำนวนรอบ	7					
2	จุดศูนย์กลางเริ่มต้น	3.12	0.94	4.66	1.62	4.08	1.52
	จุดศูนย์กลางสุดท้าย	1.464	0.2439	90.623	412.42	4.269	1.342
	จำนวนรอบ	7					
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
998	จุดศูนย์กลางเริ่มต้น	4.46	1.56	2	0.44	5.04	1.78
	จุดศูนย์กลางสุดท้าย	4.269	1.342	1.464	0.2439	5.595	2.037
	จำนวนรอบ	5					
999	จุดศูนย์กลางเริ่มต้น	2.9	0.84	4.02	1.32	4.24	1.42
	จุดศูนย์กลางสุดท้าย	1.464	0.2439	4.269	1.342	5.595	2.037
	จำนวนรอบ	7					
1000	จุดศูนย์กลางเริ่มต้น	2.66	0.74	2.94	0.96	4	1.32
	จุดศูนย์กลางสุดท้าย	1.464	0.2439	4.269	1.342	5.595	2.037
	จำนวนรอบ	9					

จากผลลัพธ์ข้างต้นจะเห็นได้ว่ามีทั้งจำนวนรอบที่ดีและแย่ดังนั้นสามารถสรุปจำนวนรอบเฉลี่ยได้ดังตาราง

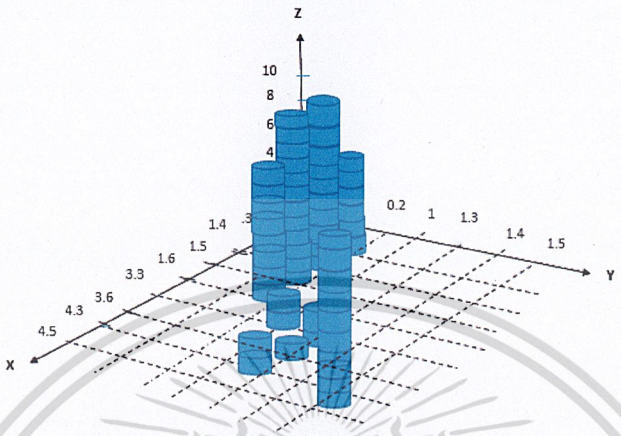
ตารางที่ 4.7 แสดงจำนวนรอบการคำนวณวิธีหาค่ากึ่งกลางของช่วง

เซตข้อมูล	จำนวนรอบที่ดีที่สุด	จำนวนรอบที่แย่ที่สุด	จำนวนรอบเฉลี่ย
Iris	2	15	7.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.5 การหาค่ากลางเริ่มต้นของกลุ่มโดยการเลือกจากความถี่สูงสุดของเซตข้อมูล Iris

จากเซตข้อมูล Iris จำนวน 150 ข้อมูลมีข้อมูลที่ซ้ำกันดังรูป



รูปที่ 4.4 แสดงความถี่ของข้อมูลเซตข้อมูล Iris

ตารางที่ 4.8 แสดงความถี่ของข้อมูลที่ซ้ำกันของเซตข้อมูล Iris

Petal length	Petal width	ความถี่
1.4	0.2	8
1.5	0.2	7
4.5	1.5	5
1.3	0.2	4
1.3	0.3	2
1.2	0.2	2

จากตารางความถี่ดังนั้นจุดศูนย์กลางเริ่มต้นในการคำนวณคือ (1.4,0.2) (1.5,0.2) และ (4.5,1.5)

เซตข้อมูล	จุดศูนย์กลาง	$Cluster_1$	$Cluster_2$	$Cluster_3$	จำนวนรอบ
Iris	จุดศูนย์กลางเริ่มต้น	(3.32,0.98)	(3.32,1.10)	(3.5,0.90)	9
	จุดศูนย์กลางสุดท้าย	(4.269,1.342)	(5.595,2.037)	(1.462,0.246)	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.6 จำนวนรอบการคำนวณของเซตข้อมูล Iris

ในการเลือกค่าเริ่มต้นในการคำนวณของวิธีต่างๆสามารถสรุปรอบการคำนวณของแต่ละวิธีได้ดังตาราง

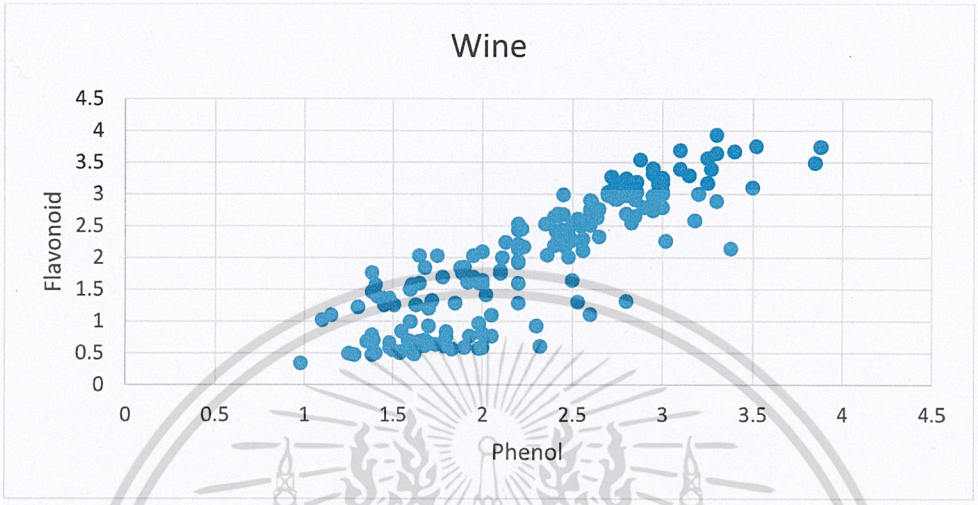
ตารางที่ 4.9 แสดงรอบการคำนวณของแต่ละวิธีของเซตข้อมูล Iris

วิธีเลือกค่าเริ่มต้น	จำนวนการสุ่ม	จำนวนรอบการคำนวณเคมีน
สุ่ม	1000	8.34
คำนวณหาค่ากึ่งกลางของช่วง	1	2
สุ่มหลายค่า	1000	7.1
ความถี่สูงสุดของข้อมูล	1	9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 เซตข้อมูลไวน์

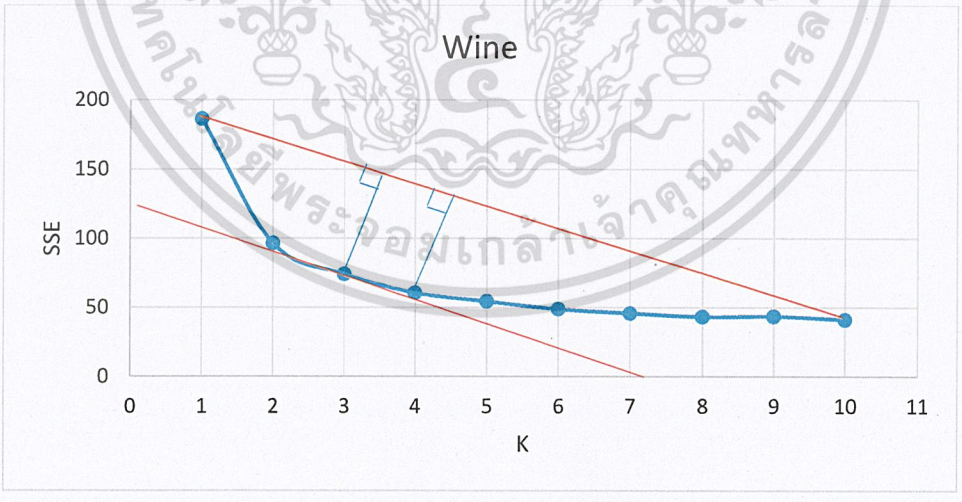
เซตข้อมูลไวน์ (Wine) จำนวน 178 ข้อมูลโดยผู้ศึกษาจะเลือกใช้แอตทริบิวต์ Phenol และ Flavonoid (สารต้านอนุมูลอิสระ) เนื่องจากลักษณะข้อมูลของทั้งสองแอตทริบิวต์สามารถแสดงดังรูป



รูปที่ 4.5 กราฟแสดงข้อมูล Phenol-Flavonoid ของเซตข้อมูล Wine

4.2.1 จำนวนกลุ่มที่เหมาะสมของเซตข้อมูล Wine

คำนวณค่า SSE ของเซตข้อมูล Wine เมื่อ K =1 ถึง K =10 และคำนวณหาจุดเปลี่ยนเว้าจากค่าความชันของเส้นแนวระดับและแต่ละจุด K



รูปที่ 4.6 กราฟแสดงค่า SSE ของเซตข้อมูล Wine

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำนวณค่าความชันของเส้นแนวระดับจะได้เท่ากับ

$$m = \frac{\Delta y}{\Delta x}$$

$$m = \frac{41.341 - 186.263}{10 - 1}$$

$$m = -16.77$$

จากนั้นคำนวณความของแต่ละจุดและหาผลต่างของความชันที่จุดนั้นๆและความชันของเส้นแนวระดับดังตาราง

ตารางที่ 4.10 แสดงค่าความชันของแต่ละจุด

จำนวนกลุ่ม (K)	ค่า SSE	ความชัน	ผลต่างของความชันกับเส้นแนวระดับ
1	186.2631	-	-
2	96.44895	-89.8142	-73.0442
3	74.23309	-22.2159	-5.44586
4	60.74152	-13.4916	3.27843
5	54.53051	-6.21101	10.55899
6	48.9512	-5.57931	11.19069
7	45.97797	-2.97323	13.79677
8	43.4115	-2.56647	14.20353
9	43.71805	0.30655	17.07655
10	41.34138	-2.37667	14.39333

จากตารางที่ 4.10 จะเห็นได้ว่าเมื่อ K =3 มีค่าความชันเป็นลบและเมื่อ K =4 ค่าความชันเป็นบวกแสดงว่าค่า K ที่เหมาะสมของเซตข้อมูล Wine คือ K =3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.2 วิธีการหาค่าเริ่มต้นของกลุ่มโดยการสุ่มของเซตข้อมูล Wine

จากในบทที่ 3 โดยวิธีการสุ่มจุดศูนย์กลางเริ่มต้นโดยสุ่มจำนวน 1000 ครั้งได้ผลการทดลองของเซตข้อมูล Wine ซึ่งจะแสดงจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายและรอบในการคำนวณดังนี้

ตารางที่ 4.11 แสดงจุดศูนย์กลางและรอบการคำนวณจากการสุ่มจำนวน 1000 ครั้ง

รอบ	จุดศูนย์กลาง	$Cluster_1$			$Cluster_2$			$Cluster_3$		
		Alcohol	Flavonoid	ตำแหน่ง	Alcohol	Flavonoid	ตำแหน่ง	Alcohol	Flavonoid	ตำแหน่ง
1	จุดศูนย์กลางเริ่มต้น	1.38	1.76	108	2.85	2.91	16	2	2.09	118
	จุดศูนย์กลางสุดท้าย	1.62125	0.83768		2.90814	2.99871		2.17627	1.94725	
	จำนวนรอบ	4								
2	จุดศูนย์กลางเริ่มต้น	1.41	1.39	149	2.6	2.68	37	2.42	2.65	66
	จุดศูนย์กลางสุดท้าย	1.64919	0.8971		2.96197	3.06311		2.26704	2.10481	
	จำนวนรอบ	7								
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
998	จุดศูนย์กลางเริ่มต้น	1.83	0.56	161	2.1	1.79	63	1.55	0.84	168
	จุดศูนย์กลางสุดท้าย	1.62125	0.83768		2.90814	2.99871		2.17627	1.94725	
	จำนวนรอบ	12								
999	จุดศูนย์กลางเริ่มต้น	1.7	0.92	166	1.39	0.7	163	3	2.79	54
	จุดศูนย์สุดท้าย	2.17627	1.94725		1.62125	0.83768		2.90814	2.99871	
	จำนวนรอบ	10								
1000	จุดศูนย์กลางเริ่มต้น	3.1	3.39	48	1.85	1.28	70	1.65	0.68	176
	จุดศูนย์กลางสุดท้าย	2.90814	2.99871		2.17627	1.94725		1.62125	0.83768	
	จำนวนรอบ	7								

จากผลข้างต้นสามารถสรุปจำนวนรอบที่ดีที่สุด จำนวนรอบที่แย่ที่สุดและค่าเฉลี่ยของการทำงาน 1000 ครั้งดังตารางที่ 4.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.12 แสดงจำนวนรอบการคำนวณโดยการสุ่มสุ่ม 1000 ครั้งของเซตข้อมูล Wine

เซตข้อมูล	จำนวนรอบที่ดีที่สุด	จำนวนรอบที่แย่ที่สุด	จำนวนรอบการคำนวณเฉลี่ย
Wine	2	18	9.7

4.2.3 การหาค่ากลางเริ่มต้นของกลุ่มโดยการหาค่ากึ่งกลางของช่วงของเซตข้อมูล Wine

จากการคำนวณหาค่ากึ่งกลางของช่วง โดยใช้คำสั่งในรูปที่ 3.6 ได้ผลการทดลองของเซตข้อมูล Wine ดังนี้

1) หาความกว้างของข้อมูลแต่ละช่วงของแกน x

$$interval_x = \frac{\max_x - \min_x}{K} = \frac{3.88 - 0.98}{3} = 0.97$$

ดังนั้นแต่ละช่วงจะมีความกว้างเท่ากับ 0.97 ดังนั้นจุดแบ่งช่วงแรกคือ 0.98+0.97=1.95 ช่วงถัดไปคือ 2.92 และช่วงสุดท้ายคือ 3.89

หาความกว้างของข้อมูลแต่ละช่วงของแกน y

$$interval_y = \frac{\max_y - \min_y}{K} = \frac{3.93 - 0.34}{3} = 1.19$$

ดังนั้นแต่ละช่วงจะมีความกว้างเท่ากับ 1.19 ดังนั้นจุดแบ่งช่วงแรกคือ 0.34+1.19=1.53 ช่วงถัดไปคือ 2.73 และช่วงสุดท้ายคือ 3.93

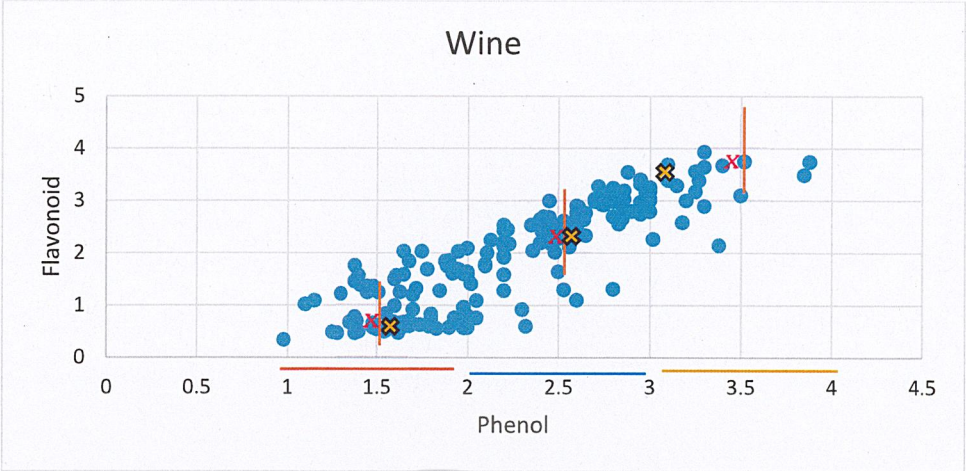
2) หาค่ากึ่งกลางของแต่ละช่วงของแกน x โดยนำค่า $\frac{\max - \min}{2} + \min$ ของแต่ละช่วงทั้งแกน x และแกน y

ดังนั้นค่า *median* ช่วงแรกของแกน x คือ $\frac{1.95 - 0.98}{2} = 0.48 + 0.98 = 1.46$ และ

median ช่วงแรกของแกน y คือ $\frac{1.53 - 0.34}{2} = 0.595 + 0.34 = 0.93$

ดังนั้นค่าเริ่มต้นในการคำนวณของช่วงแรกคือ (1.46,0.93) ช่วงที่สองและช่วงสุดท้ายคิดเช่นเดียวกันจะได้ (2.43,2.13) และ (3.39,3.33) แสดงดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.7 แสดงการหาจุดกึ่งกลางของแต่ละช่วงของเซตข้อมูล Wine

เมื่อ x คือจุดศูนย์กลางเริ่มต้นในการคำนวณแบบกึ่งกลางของช่วง และ

คือจุดศูนย์กลางสุดท้ายของเซตข้อมูล

จากการคำนวณข้างต้นสามารถสรุปค่าจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายและจำนวนรอบการคำนวณเคมีนแสดงดังตาราง

ตารางที่ 4.13 แสดงจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายและจำนวนรอบของการคำนวณเคมีนของเซตข้อมูล Wine

รอบ	จุดศูนย์กลาง	$Cluster_1$		$Cluster_2$		$Cluster_3$	
		Phenol	Flavonoid	Phenol	Flavonoid	Phenol	Flavonoid
1	จุดศูนย์กลางเริ่มต้น	1.46333	0.93833	2.43	2.135	3.39667	3.33167
	จุดศูนย์กลางสุดท้าย	1.64919	0.8971	2.26704	2.10481	2.96197	3.06311
	จำนวนรอบ	6					

4.2.4 การหาค่ากลางเริ่มต้นของกลุ่มโดยการคำนวณสุ่มหลายค่า

การคำนวณเคมีน โดยใช้วิธีการคำนวณแบบสุ่มหลายค่า โดยจะนำค่าเฉลี่ยของข้อมูลมาเป็นค่ากลางเริ่มต้นในการคำนวณเคมีนโดยจะสุ่มจาก 10% ของจำนวนเซตข้อมูล Wine ซึ่งคำนวณจากในบทที่ 3 ได้ผลออกมาดังตารางที่ 4.14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.14 แสดงการสุ่มหลายค่าของเซตข้อมูล Wine

เซตข้อมูล	Wine
ค่า K	3
จำนวนข้อมูล	178
จำนวนที่ต้องสุ่มทั้งหมด	19
จำนวนที่สุ่มของแต่ละกลุ่ม	6

ตารางที่ 4.15 แสดงจุดศูนย์กลางและจำนวนรอบในการคำนวณจากการสุ่มจำนวน 1000 รอบ

รอบ	จุดศูนย์กลาง	$Cluster_1$		$Cluster_2$		$Cluster_3$	
		Alcohol	Flavonoid	Alcohol	Flavonoid	Alcohol	Flavonoid
1	จุดศูนย์กลางเริ่มต้น	2.268	2.066	1.5979	0.9019	2.0759	1.476
	จุดศูนย์กลางสุดท้าย	1.649	0.89	2.26	2.104	2.96	3.063
	จำนวนรอบ	12					
2	จุดศูนย์กลางเริ่มต้น	2.158	1.772	2.404	2.364	2.398	2.34
	จุดศูนย์กลางสุดท้าย	2.26	2.104	1.649	0.89	2.96	3.063
	จำนวนรอบ	5					
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
998	จุดศูนย์กลางเริ่มต้น	1.682	1.132	1.99	1.55	2.164	2.0119
	จุดศูนย์กลางสุดท้าย	2.296	3.063	2.26	2.104	1.649	0.89
	จำนวนรอบ	5					
999	จุดศูนย์กลางเริ่มต้น	1.546	0.796	2.502	1.85	2.276	2.2359
	จุดศูนย์สุดท้าย	1.649	0.89	2.26	2.104	2.96	3.063
	จำนวนรอบ	6					
1000	จุดศูนย์กลางเริ่มต้น	2.739	2.974	2.192	1.6	1.914	1.316
	จุดศูนย์กลางสุดท้าย	2.96	3.063	2.26	2.104	1.649	0.89
	จำนวนรอบ	4					

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลลัพธ์ข้างต้นจะเห็นได้ว่ามีทั้งจำนวนรอบที่ดีและแย่ดังนั้นสามารถสรุปจำนวนรอบเฉลี่ยได้ดังตารางที่ 4.16

ตารางที่ 4.16 แสดงค่าเฉลี่ยจากการสุ่มจำนวน 1000 รอบ

เซตข้อมูล	จำนวนรอบที่ดีที่สุด	จำนวนรอบที่แย่ที่สุด	จำนวนรอบการคำนวณเฉลี่ย
Wine	3	16	7

4.2.5 การหาค่ากลางเริ่มต้นของกลุ่มโดยการสุ่มจากความถี่สูงสุดของข้อมูล

เซตข้อมูล Wine จำนวน 178 ข้อมูลดังตาราง

ตารางที่ 4.17 เซตข้อมูลWine

ข้อมูลที่	Phenol	Flavonoid
1	2.8	3.06
2	2.65	2.76
3	2.8	3.24
4	3.85	3.49
5	2.8	2.69
6	3.27	3.39
7	2.5	2.52
8	2.6	2.51
9	2.8	2.98
10	2.98	3.15
11	2.95	3.32
12	2.2	2.43
⋮	⋮	⋮
169	1.55	0.84
170	1.98	0.96
171	1.25	0.49
172	1.39	0.51
173	1.68	0.7
174	1.68	0.61
175	1.8	0.75
176	1.59	0.69
177	1.65	0.68
178	2.05	0.76

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.17 ของเซตข้อมูล Wine จะพบว่าไม่มีข้อมูลใดที่ซ้ำกันจึงไม่เกิดความถี่ของข้อมูลดังนั้นเซตข้อมูล Wine จึงไม่เหมาะกับวิธีการเลือกค่าเริ่มต้นจากความถี่สูงสุดของข้อมูล

4.2.6 จำนวนรอบการคำนวณของเซตข้อมูล Wine

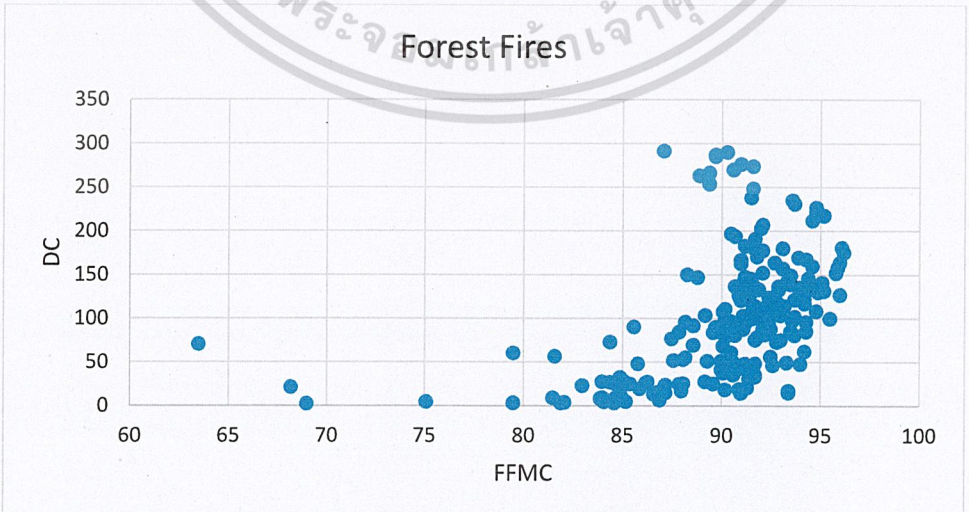
ในการเลือกค่าเริ่มต้นในการคำนวณของวิธีต่างๆสามารถสรุปรอบการคำนวณของแต่ละวิธีได้ดังตาราง

ตารางที่ 4.18 แสดงรอบการคำนวณของแต่ละวิธีของเซตข้อมูล Wine

วิธีเลือกค่าเริ่มต้น	จำนวนการสุ่ม	จำนวนรอบการคำนวณเคมีน
สุ่ม	1000	9.7
คำนวณหาค่ากึ่งกลางของช่วง	1	6
สุ่มหลายค่า	1000	7
ความถี่สูงสุดของข้อมูล	-	-

4.3 เซตข้อมูลไฟป่า

เซตข้อมูลไฟป่า (Forest Fires) จำนวน 514 ข้อมูลโดยผู้ศึกษาจะเลือกใช้แอตทริบิวต์บิวต์ FFMC (Fine Fuel Moisture Code : ค่าความชื้นของเชื้อเพลิงชั้นบนสุด) และ DC (Drought Cod : ค่าความชื้นของเชื้อเพลิงชั้นล่างสุดสามารถบอกว่าความแห้งแล้วส่งผลกระทบต่อเชื้อเพลิงอย่างไร) เนื่องจากลักษณะข้อมูลของทั้งสองแอตทริบิวต์สามารถแสดงดังรูป



รูปที่ 4.8 กราฟแสดงข้อมูล FFMC-DC ของเซตข้อมูล Forest Fires

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.1 จำนวนกลุ่มที่เหมาะสมของเซตข้อมูล Forest Fires

คำนวณค่า SSE ของเซตข้อมูล Forest Fires เมื่อ K =1 ถึง K =10 และคำนวณหาจุดเปลี่ยนเว้าจากค่าความชันของเส้นแนวระดับและแต่ละจุด K

รูปที่ 4.9 กราฟแสดงค่า SSE ของเซตข้อมูล Forest Fires

คำนวณค่าความชันของเส้นแนวระดับจะได้เท่ากับ

$$m = \frac{\Delta y}{\Delta x}$$

$$m = \frac{8594.6 - 104271.6}{10 - 1}$$

$$m = -10630.77$$

จากนั้นคำนวณความของแต่ละจุดและหาผลต่างของความชันที่จุดนั้นๆและความชันของเส้นแนวระดับดังตารางที่ 4.19

ตารางที่ 4.19 แสดงค่าความชันของแต่ละจุดและค่าผลต่างของความชัน

จำนวนกลุ่ม (K)	ค่า SSE	ความชัน	ผลต่างของความชันกับเส้นแนวระดับ
1	104271.6	-	-
2	38402.52	-65869	-55238.261
3	25516.32	-12886.2	-2255.4328

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จำนวนกลุ่ม (K)	ค่า SSE	ความชัน	ผลต่างของความชันกับเส้นแนวระดับ
4	19187.9	-6328.42	4302.3483
5	14779.76	-4408.14	6222.6303
6	13009.97	-1769.8	8860.9704
7	10690.2	-2319.77	8311.0015
8	9728.157	-962.041	9668.7283
9	8926.493	-801.664	9829.1046
10	8594.679	-331.814	10298.955

จากตารางที่ 4.19 จะเห็นได้ว่าเมื่อ K =3 มีค่าความชันเป็นลบและเมื่อ K =4 ค่าความชันเป็นบวกแสดงว่าค่า K ที่เหมาะสมของเซตข้อมูล Forest Fires คือ K =3

4.3.2 วิธีการหาค่าเริ่มต้นของกลุ่มโดยการสุ่มของเซตข้อมูล Forest Fires

จากในบทที่ 3 โดยวิธีการสุ่มจุดศูนย์กลางเริ่มต้นโดยสุ่มจำนวน 1000 ครั้งได้ผลการทดลองของเซตข้อมูล Forest Fires ซึ่งจะแสดงจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายและรอบในการคำนวณดังนี้

ตารางที่ 4.20 แสดงจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายของเซตข้อมูล Forest Fires โดยให้ทำการสุ่มจำนวน 1000 รอบ และแสดงจำนวนรอบของการคำนวณเคมีน

รอบ	จุดศูนย์กลาง	$Cluster_1$			$Cluster_2$			$Cluster_3$		
		FFMC	DC	ตำแหน่ง	FFMC	DC	ตำแหน่ง	FFMC	DC	ตำแหน่ง
1	จุดศูนย์กลางเริ่มต้น	89.2	431.6	480	85.4	349.7	279	92.4	668	27
	จุดศูนย์กลางสุดท้าย	90.6234	412.4218		86.47742	75.05806		91.725	694.19333	
	จำนวนรอบ	7								
2	จุดศูนย์กลางเริ่มต้น	91.6	474.9	292	92.7	575.8	417	90	682.6	38
	จุดศูนย์กลางสุดท้าย	86.47742	75.05806		90.62344	412.42188		91.725	694.19333	
	จำนวนรอบ	12								
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รอบ	จุดศูนย์กลาง	$Cluster_1$			$Cluster_2$			$Cluster_3$		
		FFMC	DC	ตำแหน่ง	FFMC	DC	ตำแหน่ง	FFMC	DC	ตำแหน่ง
998	จุดศูนย์กลางเริ่มต้น	91.1	738.1	440	95.2	578.8	192	92.1	745.3	357
	จุดศูนย์กลางสุดท้าย	90.62344	412.42188		86.47742	75.05806		91.725	694.19333	
	จำนวนรอบ	18								
999	จุดศูนย์กลางเริ่มต้น	91.7	80.8	17	92.9	699.6	231	91.5	608.2	121
	จุดศูนย์สุดท้าย	86.47742	75.05806		91.725	694.19333		90.62344	412.42188	
	จำนวนรอบ	10								
1000	จุดศูนย์กลางเริ่มต้น	92.3	450.2	410	88.1	67.6	135	91	25.6	470
	จุดศูนย์กลางสุดท้าย	91.725	694.19333		90.62344	412.42188		86.47742	75.05806	
	จำนวนรอบ	9								

จากผลลัพธ์ข้างต้นจะเห็นได้ว่ามีทั้งจำนวนรอบที่ดีและแย่ดังนั้นสามารถสรุปจำนวนรอบเฉลี่ยได้ดังตาราง

ตารางที่ 4.21 แสดงค่าเฉลี่ยของวิธีการสุ่ม จำนวนรอบที่ดีที่สุด และจำนวนรอบที่แย่ที่สุดของเซตข้อมูล Forest Fires

เซตข้อมูล	จำนวนรอบที่ดีที่สุด	จำนวนรอบที่แย่ที่สุด	จำนวนรอบเฉลี่ย
Forest Fires	2	18	11.6

4.3.3 การหาค่ากลางเริ่มต้นของกลุ่มโดยการหาค่ากึ่งกลางของช่วงของเซตข้อมูล Forest Fires

จากการคำนวณหาค่ากึ่งกลางของช่วง โดยใช้คำสั่งในรูปที่ 3.3 ได้ผลการทดลองของเซตข้อมูล Forest Fires ดังนี้

1) หาความกว้างของข้อมูลแต่ละช่วงของแกน x

$$interval_x = \frac{\max_x - \min_x}{K} = \frac{96.2 - 63.5}{3} = 10.90$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นแต่ละช่วงจะมีความกว้างเท่ากับ 10.90 ดังนั้นจุดแบ่งช่วงแรกคือ 63.5+10.90=74.4 ช่วงถัดไปคือ 85.3 และช่วงสุดท้ายคือ 96.2

หาความกว้างของข้อมูลแต่ละช่วงของแกน y

$$interval_y = \frac{\max_y - \min_y}{K} = \frac{860.60 - 7.90}{3} = 284.23$$

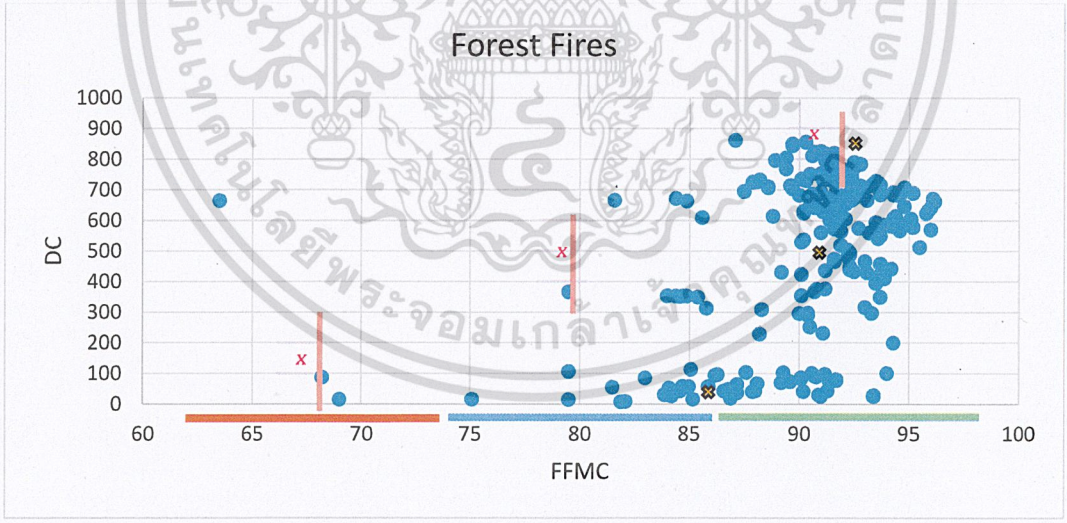
ดังนั้นแต่ละช่วงจะมีความกว้างเท่ากับ 284.23 ดังนั้นจุดแบ่งช่วงแรกคือ 7.90+284.23= 292.13 ช่วงถัดไปคือ 576.36 และช่วงสุดท้ายคือ 860.59

2) หาค่ากึ่งกลางของแต่ละช่วงของแกน x โดยนำค่า $\frac{\max - \min}{2} + \min$ ของแต่ละช่วงทั้งแกน x และแกน y

ดังนั้นค่า *median* ช่วงแรกของแกน x คือ $\frac{74.4 - 63.5}{2} = 5.45 + 63.5 = 68.95$ และ

median ช่วงแรกของแกน y คือ $\frac{292.13 - 7.90}{2} = 142.11 + 7.90 = 150.01$

ดังนั้นค่าเริ่มต้นในการคำนวณของช่วงแรกคือ (1.46,0.93) ช่วงที่สองและช่วงสุดท้ายคิดเช่นเดียวกันจะได้ (2.43,2.13) และ (3.39,3.33) แสดงดังรูป



รูปที่ 4.10 แสดงการคำนวณหาค่ากึ่งกลางแต่ละช่วงของเซตข้อมูล Forest Fires

เมื่อ x คือจุดศูนย์กลางเริ่มต้นในการคำนวณแบบกึ่งกลางของช่วง และ

✖ คือจุดศูนย์กลางสุดท้ายของเซตข้อมูล

จากการคำนวณข้างต้นสามารถสรุปค่าจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายและจำนวนรอบการคำนวณเคมีนแสดงดังตารางที่ 4.22

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.22 แสดงจุดศูนย์กลางเริ่มต้น จุดศูนย์กลางสุดท้ายและจำนวนรอบของการคำนวณ เคมีนของเซตข้อมูล Forest Fires ที่ได้จากการใช้คำสั่งในบทที่ 3

รอบ	จุดศูนย์กลาง	$Cluster_1$		$Cluster_2$		$Cluster_3$	
		FFMC	DC	FFMC	DC	FFMC	DC
1	จุดศูนย์กลางเริ่มต้น	68.95	150.01667	79.85	434.25	90.75	718.48333
	จุดศูนย์กลางสุดท้าย	87.58571	72.25495	90.62344	412.42188	91.84011	694.15877
	จำนวนรอบการคำนวณ	5					

4.3.4 การหาค่ากลางเริ่มต้นของกลุ่มโดยการคำนวณสุ่มหลายค่า

การคำนวณเคมีน โดยใช้วิธีการคำนวณแบบสุ่มหลายค่าโดยจะนำค่าเฉลี่ยของข้อมูลมาเป็นค่ากลางเริ่มต้นในการคำนวณเคมีนโดยจะสุ่มจาก 10% ของจำนวนข้อมูลเซตข้อมูล Forest Fires ซึ่งคำนวณจากในบทที่ 3 ได้ผลออกมาดังตารางที่ 4.23

ตารางที่ 4.23 แสดงผลการสุ่มของวิธีสุ่มแบบหลายค่าและจำนวนรอบการคำนวณเคมีนของเซตข้อมูล Forest Fires

เซตข้อมูล	Forest Fires
ค่า K	3
จำนวนของข้อมูล	514
จำนวนที่ต้องสุ่มทั้งหมด	51
จำนวนที่สุ่มแต่ละกลุ่ม	17

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.24 แสดงจุดศูนย์กลางและจำนวนรอบในการคำนวณจากการสุ่มจำนวน 1000 รอบ

รอบ	จุดศูนย์กลาง	$Cluster_1$		$Cluster_2$		$Cluster_3$	
		FFMC	DC	FFMC	DC	FFMC	DC
1	จุดศูนย์กลางเริ่มต้น	90.4176	448.6823	91.42357	537.3411	91.9294	641.6588
	จุดศูนย์กลางสุดท้าย	86.477	75.058	90.623	412.42	91.725	694.193
	จำนวนรอบ	11					
2	จุดศูนย์กลางเริ่มต้น	91.8647	588.0941	91.3529	516.2882	90.5176	391.8294
	จุดศูนย์กลางสุดท้าย	91.725	694.193	90.623	412.42	86.477	75.058
	จำนวนรอบ	7					
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
998	จุดศูนย์กลางเริ่มต้น	90.9647	555.3235	91.25887	543.7647	88.9823	449.7647
	จุดศูนย์กลางสุดท้าย	91.725	694.193	90.623	412.42	86.477	75.058
	จำนวนรอบ	11					
999	จุดศูนย์กลางเริ่มต้น	90.1117	584.5294	91.6941	634.2705	87.1999	556.5058
	จุดศูนย์สุดท้าย	90.623	412.42	91.725	694.193	86.477	75.058
	จำนวนรอบ	12					
1000	จุดศูนย์กลางเริ่มต้น	92.0411	457.8999	90.6999	501.3117	90.04705	519.6176
	จุดศูนย์กลางสุดท้าย	86.477	75.058	90.623	412.42	91.725	694.193
	จำนวนรอบ	7					

จากผลลัพธ์ข้างต้นจะเห็นได้ว่ามีทั้งจำนวนรอบที่ดีและแย่ดังนั้นสามารถสรุปจำนวนรอบเฉลี่ยได้ดังตารางที่ 4.25

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.25 แสดงค่าเฉลี่ยจากการสุ่มจำนวน 1000 รอบ

เซตข้อมูล	จำนวนรอบที่ดีที่สุด	จำนวนรอบที่แย่ที่สุด	จำนวนรอบเฉลี่ย
Forest Fires	4	15	10.9

4.3.5 การหาค่ากลางเริ่มต้นของกลุ่มโดยการเลือกจากความถี่สูงสุดของเซตข้อมูล Forest Fires

ตารางที่ 4.26 แสดงความถี่ของข้อมูลฐานข้อมูล Forest Fires

FFMC	DC	ความถี่
92.1	745.3	10
91	692.6	9
91.4	601.4	8
93.7	715.1	7
90.9	686.5	7
91.7	80.8	7
94.8	647.1	7
92.9	706.4	7
91.6	764	7
92.4	668	6
93.5	728.6	6
94.3	692.3	6
90.2	631.2	6
92.1	654.1	6
95.2	690	3
94.8	706.7	3
94.9	587.1	3
90.6	669.1	2
92.3	488	2
63.5	665.3	2
84.9	664.2	2
90.3	730.2	2
87.2	64.7	2
86.6	43	2
93	466.6	2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางความถี่ของฐานข้อมูล Forest Fires เมื่อ K =3 จะได้จุดศูนย์กลางเริ่มต้นคือ (92.1,745.3) (91,692.6) และ (91.4,601.4) จากนั้นคำนวณเคมีนได้จุดศูนย์กลางสุดท้ายและจำนวนรอบการคำนวณดังตารางที่ 4.27

ตารางที่ 4.27 แสดงจุดศูนย์กลางและจำนวนรอบการคำนวณของเซตข้อมูล Forest Fires

เซตข้อมูล	จุดศูนย์กลาง	$Cluster_1$	$Cluster_2$	$Cluster_3$	จำนวนรอบการคำนวณ
Forest Fires	จุดศูนย์กลางเริ่มต้น	(92.1,745.3)	(91,692.6)	(91.4,601.4)	12
	จุดศูนย์กลางสุดท้าย	(90.623,412.42)	(86.477,75.058)	(91.725,694.193)	

4.3.6 จำนวนรอบการคำนวณของเซตข้อมูล Forest Fires

ในการเลือกค่าเริ่มต้นในการคำนวณของวิธีต่างๆสามารถสรุปรอบการคำนวณของแต่ละวิธีได้ดังตารางที่ 4.28

ตารางที่ 4.28 แสดงรอบการคำนวณของแต่ละวิธีของเซตข้อมูล Forest Fires

วิธีเลือกค่าเริ่มต้น	จำนวนการสุ่ม	จำนวนรอบการคำนวณเคมีน
สุ่ม	1000	11.6
คำนวณหาก๋ึงกลางของช่วง	1	5
สุ่มหลายค่า	1000	10.9
ความถี่สูงสุดของข้อมูล	1	12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.4 จำนวนรอบการคำนวณเคมีนของแต่ละวิธี

จากการทดลองเลือกค่าเริ่มต้นในการคำนวณเคมีน 4 วิธีกับ 3 เซตข้อมูลจำนวนรอบแสดงดังตาราง

ตารางที่ 4.29 แสดงจำนวนรอบแต่ละวิธีของแต่ละเซตข้อมูล

เซตข้อมูล	วิธีที่ศึกษา	จำนวนรอบในการสุ่ม (รอบ)	จำนวนรอบเฉลี่ย (รอบ)
Iris	แบบสุ่ม	1000	8.34
	สุ่มหลายค่า	1000	7.1
	ค่ากึ่งกลางของช่วง	1	2
	เลือกจากความถี่สูงสุดของข้อมูล	1	9
Wine	แบบสุ่ม	1000	9.7
	สุ่มหลายค่า	1000	7
	ค่ากึ่งกลางของช่วง	1	6
	เลือกจากความถี่สูงสุดของข้อมูล	-	-
Forest Fires	แบบสุ่ม	1000	11.6
	สุ่มหลายค่า	1000	10.9
	ค่ากึ่งกลางของช่วง	1	5
	เลือกจากความถี่สูงสุดของข้อมูล	1	12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลและข้อเสนอแนะ

5.1 สรุปผล

จากการเลือกค่าเริ่มต้นในการคำนวณเคมีนจาก 4 วิธีคือ 1) การสุ่ม 2) วิธีคำนวณหาค่ากึ่งกลางของช่วง 3) วิธีสุ่มหลายค่า และ 4) ใช้ความถี่สูงสุดของข้อมูล บน $\mathbb{R}^2$ โดยทดลองกับเซตข้อมูล 3 เซตข้อมูลคือ 1) เซตข้อมูลไอริส (Iris-setosa) 2) เซตข้อมูลไวน์ (Wine) และ 3) เซตข้อมูลไฟป่า (Forest Fires) พบว่าจำนวนรอบของการคำนวณเคมีนแตกต่างกัน จะเห็นได้ว่าการสุ่มที่ดีนั้นขึ้นอยู่กับการกระจายและความหนาแน่นของข้อมูล ถ้าเซตข้อมูลมีความหนาแน่นและการกระจายตัวของข้อมูลมาก วิธีคำนวณหาค่ากึ่งกลางของช่วงมีรอบในการคำนวณเคมีนวนซ้ำน้อยที่สุด รองลงมาคือวิธีสุ่มหลายค่า วิธีสุ่ม และใช้ความถี่สูงสุดของข้อมูล ตามลำดับ

5.2 ข้อเสนอแนะ

1) การสุ่มวิธีต่างๆสมารถหาผลลัพธ์ของเซตข้อมูลบน $\mathbb{R}^2$ ได้แล้วในทำนองเดียวกันสามารถนำไปพัฒนาเพื่อหาผลลัพธ์การสุ่มแบบต่างๆของเซตข้อมูลบน $\mathbb{R}^3$ ได้

2) การคำนวณความถี่ของข้อมูลกับความหนาแน่นของข้อมูลอาจจะทำการคำนวณเคมีนมีประสิทธิภาพมากยิ่งขึ้น

3) สำหรับการหาความถี่ของข้อมูลบนแกน x, y ไม่สามารถมองเห็นความถี่ของข้อมูลได้ จึงควรหาวิธีการหรือฟังชันก์การฉายภาพ (projection) หรือพัฒนาเพื่อแก้ปัญหานี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

Chayanan N. เอกสารประกอบการเรียนวิชา Data Mining Analysis. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. กรุงเทพฯ.

Forest Fires Data Set (2008). สืบค้นเมื่อ 20 ธันวาคม 2560, จาก https://archive.ics.uci.edu/ml/datasets/ forest+fires

Iris Data Set (1988). สืบค้นเมื่อ 20 ธันวาคม 2560,จาก https://archive.ics.uci.edu/ml/datasets/iris

Mohammed J. Zaki,Wagner Meira Jr.,(2013). Data Minig and Analysis:Fundamental Concepts and Algorithm. Cambridge University.

Wine Data Set (1988). สืบค้นเมื่อ 20 ธันวาคม 2560, จาก https://archive.ics.uci.edu/ml/datasets/wine

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

เซตข้อมูล Iris

เซตข้อมูล Iris จำนวน 150 ข้อมูลประกอบด้วย

Sepal length คือ ความยาวกลีบเลี้ยง

Sepal width คือ ความกว้างกลีบเลี้ยง

Petal length คือ ความยาวกลีบดอก

Petal width คือ ความกว้างกลีบดอก

Sepal length	Sepal width	Petal length	Petal width	Iris
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5	3.6	1.4	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa
5.4	3.7	1.5	0.2	Iris-setosa
4.8	3.4	1.6	0.2	Iris-setosa
4.8	3	1.4	0.1	Iris-setosa
4.3	3	1.1	0.1	Iris-setosa
5.8	4	1.2	0.2	Iris-setosa
5.7	4.4	1.5	0.4	Iris-setosa
5.4	3.9	1.3	0.4	Iris-setosa
5.1	3.5	1.4	0.3	Iris-setosa
5.7	3.8	1.7	0.3	Iris-setosa
5.1	3.8	1.5	0.3	Iris-setosa
5.4	3.4	1.7	0.2	Iris-setosa
5.1	3.7	1.5	0.4	Iris-setosa
4.6	3.6	1	0.2	Iris-setosa
5.1	3.3	1.7	0.5	Iris-setosa
4.8	3.4	1.9	0.2	Iris-setosa

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

sepal length	sepal width	petal length	petal width	iris
5	3	1.6	0.2	Iris-setosa
5	3.4	1.6	0.4	Iris-setosa
5.2	3.5	1.5	0.2	Iris-setosa
5.2	3.4	1.4	0.2	Iris-setosa
4.7	3.2	1.6	0.2	Iris-setosa
4.8	3.1	1.6	0.2	Iris-setosa
5.4	3.4	1.5	0.4	Iris-setosa
5.2	4.1	1.5	0.1	Iris-setosa
5.5	4.2	1.4	0.2	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa
5	3.2	1.2	0.2	Iris-setosa
5.5	3.5	1.3	0.2	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa
4.4	3	1.3	0.2	Iris-setosa
5.1	3.4	1.5	0.2	Iris-setosa
5	3.5	1.3	0.3	Iris-setosa
4.5	2.3	1.3	0.3	Iris-setosa
4.4	3.2	1.3	0.2	Iris-setosa
5	3.5	1.6	0.6	Iris-setosa
5.1	3.8	1.9	0.4	Iris-setosa
4.8	3	1.4	0.3	Iris-setosa
5.1	3.8	1.6	0.2	Iris-setosa
4.6	3.2	1.4	0.2	Iris-setosa
5.3	3.7	1.5	0.2	Iris-setosa
5	3.3	1.4	0.2	Iris-setosa
7	3.2	4.7	1.4	Iris-versicolor
6.4	3.2	4.5	1.5	Iris-versicolor
6.9	3.1	4.9	1.5	Iris-versicolor
5.5	2.3	4	1.3	Iris-versicolor
6.5	2.8	4.6	1.5	Iris-versicolor
5.7	2.8	4.5	1.3	Iris-versicolor
6.3	3.3	4.7	1.6	Iris-versicolor
4.9	2.4	3.3	1	Iris-versicolor
6.6	2.9	4.6	1.3	Iris-versicolor
5.2	2.7	3.9	1.4	Iris-versicolor

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

sepal length	sepal width	petal length	petal width	iris
5	2	3.5	1	Iris-versicolor
5.9	3	4.2	1.5	Iris-versicolor
6	2.2	4	1	Iris-versicolor
6.1	2.9	4.7	1.4	Iris-versicolor
5.6	2.9	3.6	1.3	Iris-versicolor
6.7	3.1	4.4	1.4	Iris-versicolor
5.6	3	4.5	1.5	Iris-versicolor
5.8	2.7	4.1	1	Iris-versicolor
6.2	2.2	4.5	1.5	Iris-versicolor
5.6	2.5	3.9	1.1	Iris-versicolor
5.9	3.2	4.8	1.8	Iris-versicolor
6.1	2.8	4	1.3	Iris-versicolor
6.3	2.5	4.9	1.5	Iris-versicolor
6.1	2.8	4.7	1.2	Iris-versicolor
6.4	2.9	4.3	1.3	Iris-versicolor
6.6	3	4.4	1.4	Iris-versicolor
6.8	2.8	4.8	1.4	Iris-versicolor
6.7	3	5	1.7	Iris-versicolor
6	2.9	4.5	1.5	Iris-versicolor
5.7	2.6	3.5	1	Iris-versicolor
5.5	2.4	3.8	1.1	Iris-versicolor
5.5	2.4	3.7	1	Iris-versicolor
5.8	2.7	3.9	1.2	Iris-versicolor
6	2.7	5.1	1.6	Iris-versicolor
5.4	3	4.5	1.5	Iris-versicolor
6	3.4	4.5	1.6	Iris-versicolor
6.7	3.1	4.7	1.5	Iris-versicolor
6.3	2.3	4.4	1.3	Iris-versicolor
5.6	3	4.1	1.3	Iris-versicolor
5.5	2.5	4	1.3	Iris-versicolor
5.5	2.6	4.4	1.2	Iris-versicolor
6.1	3	4.6	1.4	Iris-versicolor
5.8	2.6	4	1.2	Iris-versicolor
5	2.3	3.3	1	Iris-versicolor
5.6	2.7	4.2	1.3	Iris-versicolor

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

sepal length	sepal width	petal length	petal width	iris
5.7	3	4.2	1.2	Iris-versicolor
5.7	2.9	4.2	1.3	Iris-versicolor
6.2	2.9	4.3	1.3	Iris-versicolor
5.1	2.5	3	1.1	Iris-versicolor
5.7	2.8	4.1	1.3	Iris-versicolor
6.3	3.3	6	2.5	Iris-virginica
5.8	2.7	5.1	1.9	Iris-virginica
7.1	3	5.9	2.1	Iris-virginica
6.3	2.9	5.6	1.8	Iris-virginica
6.5	3	5.8	2.2	Iris-virginica
7.6	3	6.6	2.1	Iris-virginica
4.9	2.5	4.5	1.7	Iris-virginica
7.3	2.9	6.3	1.8	Iris-virginica
6.7	2.5	5.8	1.8	Iris-virginica
7.2	3.6	6.1	2.5	Iris-virginica
6.5	3.2	5.1	2	Iris-virginica
6.4	2.7	5.3	1.9	Iris-virginica
6.8	3	5.5	2.1	Iris-virginica
5.7	2.5	5	2	Iris-virginica
5.8	2.8	5.1	2.4	Iris-virginica
6.4	3.2	5.3	2.3	Iris-virginica
6.5	3	5.5	1.8	Iris-virginica
7.7	3.8	6.7	2.2	Iris-virginica
7.7	2.6	6.9	2.3	Iris-virginica
6	2.2	5	1.5	Iris-virginica
6.9	3.2	5.7	2.3	Iris-virginica
5.6	2.8	4.9	2	Iris-virginica
7.7	2.8	6.7	2	Iris-virginica
6.3	2.7	4.9	1.8	Iris-virginica
6.7	3.3	5.7	2.1	Iris-virginica
7.2	3.2	6	1.8	Iris-virginica
6.2	2.8	4.8	1.8	Iris-virginica
6.1	3	4.9	1.8	Iris-virginica
6.4	2.8	5.6	2.1	Iris-virginica
7.2	3	5.8	1.6	Iris-virginica

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

sepal length	sepal width	petal length	petal width	iris
7.4	2.8	6.1	1.9	Iris-virginica
7.9	3.8	6.4	2	Iris-virginica
6.4	2.8	5.6	2.2	Iris-virginica
6.3	2.8	5.1	1.5	Iris-virginica
6.1	2.6	5.6	1.4	Iris-virginica
7.7	3	6.1	2.3	Iris-virginica
6.3	3.4	5.6	2.4	Iris-virginica
6.4	3.1	5.5	1.8	Iris-virginica
6	3	4.8	1.8	Iris-virginica
6.9	3.1	5.4	2.1	Iris-virginica
6.7	3.1	5.6	2.4	Iris-virginica
6.9	3.1	5.1	2.3	Iris-virginica
5.8	2.7	5.1	1.9	Iris-virginica
6.8	3.2	5.9	2.3	Iris-virginica
6.7	3.3	5.7	2.5	Iris-virginica
6.7	3	5.2	2.3	Iris-virginica
6.3	2.5	5	1.9	Iris-virginica
6.5	3	5.2	2	Iris-virginica
6.2	3.4	5.4	2.3	Iris-virginica
5.9	3	5.1	1.8	Iris-virginica

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

เซตข้อมูล Wine

เซตข้อมูล Wine จำนวน 178 ข้อมูลประกอบด้วย

1) Alcohol คือ แอลกอฮอล์

2) Malic acid คือ กรดเมลิก เป็นสารประกอบอินทรีย์

3) Ash คือ เถ้า

4) Alcalinity of ash คือ สภาพความเป็นด่างของเถ้า

5) Total phenols คือ ผลรวมของสารประกอบฟีนอลหรือสารประกอบฟีนอลิก

6) Magnesium คือ แร่ธาตุแมกนีเซียม

7) Flavonoid คือ สารต้านอนิมูลอิสระ

8) Neoflavanoid phenols คือ สารประกอบฟีนอล

9) Hue คือ เฉดสี

10) Color intensity คือ ความเข้มของสี

11) OD280/OD315 of diluted wines คือ ความเจือจางของโปรตีน

12) Proline คือ โพรลีนเป็นกรดอะมิโนไม่จำเป็น

13) Proanthocyanins คือ สารที่ได้จากการสกัดเมล็ดและเปลือกขององุ่น

Alcohol	Malic acid	Ash	Alcalinity of ash	Magne-sium	Total phenols	Flava-noids	Nonflava-noid-phenols	Proantho-cyanins	Color intensity	Hue	OD280 /OD315	Pro-line
14.23	1.71	2.43	15.6	127	2.8	3.06	0.28	2.29	5.64	1.04	3.92	1065
13.2	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.4	1050
13.16	2.36	2.67	18.6	101	2.8	3.24	0.3	2.81	5.68	1.03	3.17	1185
14.37	1.95	2.5	16.8	113	3.85	3.49	0.24	2.18	7.8	0.86	3.45	1480
13.24	2.59	2.87	21	118	2.8	2.69	0.39	1.82	4.32	1.04	2.93	735
14.2	1.76	2.45	15.2	112	3.27	3.39	0.34	1.97	6.75	1.05	2.85	1450
14.39	1.87	2.45	14.6	96	2.5	2.52	0.3	1.98	5.25	1.02	3.58	1290
14.06	2.15	2.61	17.6	121	2.6	2.51	0.31	1.25	5.05	1.06	3.58	1295
14.83	1.64	2.17	14	97	2.8	2.98	0.29	1.98	5.2	1.08	2.85	1045
13.86	1.35	2.27	16	98	2.98	3.15	0.22	1.85	7.22	1.01	3.55	1045
14.1	2.16	2.3	18	105	2.95	3.32	0.22	2.38	5.75	1.25	3.17	1510
14.12	1.48	2.32	16.8	95	2.2	2.43	0.26	1.57	5	1.17	2.82	1280
13.75	1.73	2.41	16	89	2.6	2.76	0.29	1.81	5.6	1.15	2.9	1320
14.75	1.73	2.39	11.4	91	3.1	3.69	0.43	2.81	5.4	1.25	2.73	1150
14.38	1.87	2.38	12	102	3.3	3.64	0.29	2.96	7.5	1.2	3	1547
13.63	1.81	2.7	17.2	112	2.85	2.91	0.3	1.46	7.3	1.28	2.88	1310
14.3	1.92	2.72	20	120	2.8	3.14	0.33	1.97	6.2	1.07	2.65	1280
13.83	1.57	2.62	20	115	2.95	3.4	0.4	1.72	6.6	1.13	2.57	1130
14.19	1.59	2.48	16.5	108	3.3	3.93	0.32	1.86	8.7	1.23	2.82	1680
13.64	3.1	2.56	15.2	116	2.7	3.03	0.17	1.66	5.1	0.96	3.36	845
14.06	1.63	2.28	16	126	3	3.17	0.24	2.1	5.65	1.09	3.71	780

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Alcohol	Malic acid	Ash	Alcalinity of ash	Magne-sium	Total phenols	Flava-noids	Nonflava-noid-phenols	Proantho-cyanins	Color intensity	Hue	OD280 /OD315	Pro-line
12.93	3.8	2.65	18.6	102	2.41	2.41	0.25	1.98	4.5	1.03	3.52	770
13.71	1.86	2.36	16.6	101	2.61	2.88	0.27	1.69	3.8	1.11	4	1035
12.85	1.6	2.52	17.8	95	2.48	2.37	0.26	1.46	3.93	1.09	3.63	1015
13.5	1.81	2.61	20	96	2.53	2.61	0.28	1.66	3.52	1.12	3.82	845
13.05	2.05	3.22	25	124	2.63	2.68	0.47	1.92	3.58	1.13	3.2	830
13.39	1.77	2.62	16.1	93	2.85	2.94	0.34	1.45	4.8	0.92	3.22	1195
13.3	1.72	2.14	17	94	2.4	2.19	0.27	1.35	3.95	1.02	2.77	1285
13.87	1.9	2.8	19.4	107	2.95	2.97	0.37	1.76	4.5	1.25	3.4	915
14.02	1.68	2.21	16	96	2.65	2.33	0.26	1.98	4.7	1.04	3.59	1035
13.73	1.5	2.7	22.5	101	3	3.25	0.29	2.38	5.7	1.19	2.71	1285
13.58	1.66	2.36	19.1	106	2.86	3.19	0.22	1.95	6.9	1.09	2.88	1515
13.68	1.83	2.36	17.2	104	2.42	2.69	0.42	1.97	3.84	1.23	2.87	990
13.76	1.53	2.7	19.5	132	2.95	2.74	0.5	1.35	5.4	1.25	3	1235
13.51	1.8	2.65	19	110	2.35	2.53	0.29	1.54	4.2	1.1	2.87	1095
13.48	1.81	2.41	20.5	100	2.7	2.98	0.26	1.86	5.1	1.04	3.47	920
13.28	1.64	2.84	15.5	110	2.6	2.68	0.34	1.36	4.6	1.09	2.78	880
13.05	1.65	2.55	18	98	2.45	2.43	0.29	1.44	4.25	1.12	2.51	1105
13.07	1.5	2.1	15.5	98	2.4	2.64	0.28	1.37	3.7	1.18	2.69	1020
14.22	3.99	2.51	13.2	128	3	3.04	0.2	2.08	5.1	0.89	3.53	760
13.56	1.71	2.31	16.2	117	3.15	3.29	0.34	2.34	6.13	0.95	3.38	795
13.41	3.84	2.12	18.8	90	2.45	2.68	0.27	1.48	4.28	0.91	3	1035
13.88	1.89	2.59	15	101	3.25	3.56	0.17	1.7	5.43	0.88	3.56	1095
13.24	3.98	2.29	17.5	103	2.64	2.63	0.32	1.66	4.36	0.82	3	680
13.05	1.77	2.1	17	107	3	3	0.28	2.03	5.04	0.88	3.35	885
14.21	4.04	2.44	18.9	111	2.85	2.65	0.3	1.25	5.24	0.87	3.33	1080
14.38	3.59	2.28	16	102	3.25	3.17	0.27	2.19	4.9	1.04	3.44	1065
13.9	1.68	2.12	16	101	3.1	3.39	0.21	2.14	6.1	0.91	3.33	985
14.1	2.02	2.4	18.8	103	2.75	2.92	0.32	2.38	6.2	1.07	2.75	1060
13.94	1.73	2.27	17.4	108	2.88	3.54	0.32	2.08	8.9	1.12	3.1	1260
13.05	1.73	2.04	12.4	92	2.72	3.27	0.17	2.91	7.2	1.12	2.91	1150
13.83	1.65	2.6	17.2	94	2.45	2.99	0.22	2.29	5.6	1.24	3.37	1265
13.82	1.75	2.42	14	111	3.88	3.74	0.32	1.87	7.05	1.01	3.26	1190
13.77	1.9	2.68	17.1	115	3	2.79	0.39	1.68	6.3	1.13	2.93	1375
13.74	1.67	2.25	16.4	118	2.6	2.9	0.21	1.62	5.85	0.92	3.2	1060
13.56	1.73	2.46	20.5	116	2.96	2.78	0.2	2.45	6.25	0.98	3.03	1120
14.22	1.7	2.3	16.3	118	3.2	3	0.26	2.03	6.38	0.94	3.31	970
13.29	1.97	2.68	16.8	102	3	3.23	0.31	1.66	6	1.07	2.84	1270
13.72	1.43	2.5	16.7	108	3.4	3.67	0.19	2.04	6.8	0.89	2.87	1285
12.37	0.94	1.36	10.6	88	1.98	0.57	0.28	0.42	1.95	1.05	1.82	520
12.33	1.1	2.28	16	101	2.05	1.09	0.63	0.41	3.27	1.25	1.67	680
12.64	1.36	2.02	16.8	100	2.02	1.41	0.53	0.62	5.75	0.98	1.59	450
13.67	1.25	1.92	18	94	2.1	1.79	0.32	0.73	3.8	1.23	2.46	630
12.37	1.13	2.16	19	87	3.5	3.1	0.19	1.87	4.45	1.22	2.87	420
12.17	1.45	2.53	19	104	1.89	1.75	0.45	1.03	2.95	1.45	2.23	355
12.37	1.21	2.56	18.1	98	2.42	2.65	0.37	2.08	4.6	1.19	2.3	678
13.11	1.01	1.7	15	78	2.98	3.18	0.26	2.28	5.3	1.12	3.18	502
12.37	1.17	1.92	19.6	78	2.11	2	0.27	1.04	4.68	1.12	3.48	510

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Alcohol	Malic acid	Ash	Alcalinity of ash	Magne-sium	Total phenols	Flava-noids	Nonflava-noid-phenols	Proantho-cyanins	Color intensity	Hue	OD280 /OD315	Pro-line
13.34	0.94	2.36	17	110	2.53	1.3	0.55	0.42	3.17	1.02	1.93	750
12.21	1.19	1.75	16.8	151	1.85	1.28	0.14	2.5	2.85	1.28	3.07	718
12.29	1.61	2.21	20.4	103	1.1	1.02	0.37	1.46	3.05	0.91	1.82	870
13.86	1.51	2.67	25	86	2.95	2.86	0.21	1.87	3.38	1.36	3.16	410
13.49	1.66	2.24	24	87	1.88	1.84	0.27	1.03	3.74	0.98	2.78	472
12.99	1.67	2.6	30	139	3.3	2.89	0.21	1.96	3.35	1.31	3.5	985
11.96	1.09	2.3	21	101	3.38	2.14	0.13	1.65	3.21	0.99	3.13	886
11.66	1.88	1.92	16	97	1.61	1.57	0.34	1.15	3.8	1.23	2.14	428
13.03	0.9	1.71	16	86	1.95	2.03	0.24	1.46	4.6	1.19	2.48	392
11.84	2.89	2.23	18	112	1.72	1.32	0.43	0.95	2.65	0.96	2.52	500
12.33	0.99	1.95	14.8	136	1.9	1.85	0.35	2.76	3.4	1.06	2.31	750
12.7	3.87	2.4	23	101	2.83	2.55	0.43	1.95	2.57	1.19	3.13	463
12	0.92	2	19	86	2.42	2.26	0.3	1.43	2.5	1.38	3.12	278
12.72	1.81	2.2	18.8	86	2.2	2.53	0.26	1.77	3.9	1.16	3.14	714
12.08	1.13	2.51	24	78	2	1.58	0.4	1.4	2.2	1.31	2.72	630
13.05	3.86	2.32	22.5	85	1.65	1.59	0.61	1.62	4.8	0.84	2.01	515
11.84	0.89	2.58	18	94	2.2	2.21	0.22	2.35	3.05	0.79	3.08	520
12.67	0.98	2.24	18	99	2.2	1.94	0.3	1.46	2.62	1.23	3.16	450
12.16	1.61	2.31	22.8	90	1.78	1.69	0.43	1.56	2.45	1.33	2.26	495
11.65	1.67	2.62	26	88	1.92	1.61	0.4	1.34	2.6	1.36	3.21	562
11.64	2.06	2.46	21.6	84	1.95	1.69	0.48	1.35	2.8	1	2.75	680
12.08	1.33	2.3	23.6	70	2.2	1.59	0.42	1.38	1.74	1.07	3.21	625
12.08	1.83	2.32	18.5	81	1.6	1.5	0.52	1.64	2.4	1.08	2.27	480
12	1.51	2.42	22	86	1.45	1.25	0.5	1.63	3.6	1.05	2.65	450
12.69	1.53	2.26	20.7	80	1.38	1.46	0.58	1.62	3.05	0.96	2.06	495
12.29	2.83	2.22	18	88	2.45	2.25	0.25	1.99	2.15	1.15	3.3	290
11.62	1.99	2.28	18	98	3.02	2.26	0.17	1.35	3.25	1.16	2.96	345
12.47	1.52	2.2	19	162	2.5	2.27	0.32	3.28	2.6	1.16	2.63	937
11.81	2.12	2.74	21.5	134	1.6	0.99	0.14	1.56	2.5	0.95	2.26	625
12.29	1.41	1.98	16	85	2.55	2.5	0.29	1.77	2.9	1.23	2.74	428
12.37	1.07	2.1	18.5	88	3.52	3.75	0.24	1.95	4.5	1.04	2.77	660
12.29	3.17	2.21	18	88	2.85	2.99	0.45	2.81	2.3	1.42	2.83	406
12.08	2.08	1.7	17.5	97	2.23	2.17	0.26	1.4	3.3	1.27	2.96	710
12.6	1.34	1.9	18.5	88	1.45	1.36	0.29	1.35	2.45	1.04	2.77	562
12.34	2.45	2.46	21	98	2.56	2.11	0.34	1.31	2.8	0.8	3.38	438
11.82	1.72	1.88	19.5	86	2.5	1.64	0.37	1.42	2.06	0.94	2.44	415
12.51	1.73	1.98	20.5	85	2.2	1.92	0.32	1.48	2.94	1.04	3.57	672
12.42	2.55	2.27	22	90	1.68	1.84	0.66	1.42	2.7	0.86	3.3	315
12.25	1.73	2.12	19	80	1.65	2.03	0.37	1.63	3.4	1	3.17	510
12.72	1.75	2.28	22.5	84	1.38	1.76	0.48	1.63	3.3	0.88	2.42	488
12.22	1.29	1.94	19	92	2.36	2.04	0.39	2.08	2.7	0.86	3.02	312
11.61	1.35	2.7	20	94	2.74	2.92	0.29	2.49	2.65	0.96	3.26	680
11.46	3.74	1.82	19.5	107	3.18	2.58	0.24	3.58	2.9	0.75	2.81	562
12.52	2.43	2.17	21	88	2.55	2.27	0.26	1.22	2	0.9	2.78	325
11.76	2.68	2.92	20	103	1.75	2.03	0.6	1.05	3.8	1.23	2.5	607
11.41	0.74	2.5	21	88	2.48	2.01	0.42	1.44	3.08	1.1	2.31	434
12.08	1.39	2.5	22.5	84	2.56	2.29	0.43	1.04	2.9	0.93	3.19	385

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Alcohol	Malic acid	Ash	Alcalinity of ash	Magne-sium	Total phenols	Flava-noids	Nonflava-noid-phenols	Proantho-cyanins	Color intensity	Hue	OD280 /OD315	Pro-line
11.03	1.51	2.2	21.5	85	2.46	2.17	0.52	2.01	1.9	1.71	2.87	407
11.82	1.47	1.99	20.8	86	1.98	1.6	0.3	1.53	1.95	0.95	3.33	495
12.42	1.61	2.19	22.5	108	2	2.09	0.34	1.61	2.06	1.06	2.96	345
12.77	3.43	1.98	16	80	1.63	1.25	0.43	0.83	3.4	0.7	2.12	372
12	3.43	2	19	87	2	1.64	0.37	1.87	1.28	0.93	3.05	564
11.45	2.4	2.42	20	96	2.9	2.79	0.32	1.83	3.25	0.8	3.39	625
11.56	2.05	3.23	28.5	119	3.18	5.08	0.47	1.87	6	0.93	3.69	465
12.42	4.43	2.73	26.5	102	2.2	2.13	0.43	1.71	2.08	0.92	3.12	365
13.05	5.8	2.13	21.5	86	2.62	2.65	0.3	2.01	2.6	0.73	3.1	380
11.87	4.31	2.39	21	82	2.86	3.03	0.21	2.91	2.8	0.75	3.64	380
12.07	2.16	2.17	21	85	2.6	2.65	0.37	1.35	2.76	0.86	3.28	378
12.43	1.53	2.29	21.5	86	2.74	3.15	0.39	1.77	3.94	0.69	2.84	352
11.79	2.13	2.78	28.5	92	2.13	2.24	0.58	1.76	3	0.97	2.44	466
12.37	1.63	2.3	24.5	88	2.22	2.45	0.4	1.9	2.12	0.89	2.78	342
12.04	4.3	2.38	22	80	2.1	1.75	0.42	1.35	2.6	0.79	2.57	580
12.86	1.35	2.32	18	122	1.51	1.25	0.21	0.94	4.1	0.76	1.29	630
12.88	2.99	2.4	20	104	1.3	1.22	0.24	0.83	5.4	0.74	1.42	530
12.81	2.31	2.4	24	98	1.15	1.09	0.27	0.83	5.7	0.66	1.36	560
12.7	3.55	2.36	21.5	106	1.7	1.2	0.17	0.84	5	0.78	1.29	600
12.51	1.24	2.25	17.5	85	2	0.58	0.6	1.25	5.45	0.75	1.51	650
12.6	2.46	2.2	18.5	94	1.62	0.66	0.63	0.94	7.1	0.73	1.58	695
12.25	4.72	2.54	21	89	1.38	0.47	0.53	0.8	3.85	0.75	1.27	720
12.53	5.51	2.64	25	96	1.79	0.6	0.63	1.1	5	0.82	1.69	515
13.49	3.59	2.19	19.5	88	1.62	0.48	0.58	0.88	5.7	0.81	1.82	580
12.84	2.96	2.61	24	101	2.32	0.6	0.53	0.81	4.92	0.89	2.15	590
12.93	2.81	2.7	21	96	1.54	0.5	0.53	0.75	4.6	0.77	2.31	600
13.36	2.56	2.35	20	89	1.4	0.5	0.37	0.64	5.6	0.7	2.47	780
13.52	3.17	2.72	23.5	97	1.55	0.52	0.5	0.55	4.35	0.89	2.06	520
13.62	4.95	2.35	20	92	2	0.8	0.47	1.02	4.4	0.91	2.05	550
12.25	3.88	2.2	18.5	112	1.38	0.78	0.29	1.14	8.21	0.65	2	855
13.16	3.57	2.15	21	102	1.5	0.55	0.43	1.3	4	0.6	1.68	830
13.88	5.04	2.23	20	80	0.98	0.34	0.4	0.68	4.9	0.58	1.33	415
12.87	4.61	2.48	21.5	86	1.7	0.65	0.47	0.86	7.65	0.54	1.86	625
13.32	3.24	2.38	21.5	92	1.93	0.76	0.45	1.25	8.42	0.55	1.62	650
13.08	3.9	2.36	21.5	113	1.41	1.39	0.34	1.14	9.4	0.57	1.33	550
13.5	3.12	2.62	24	123	1.4	1.57	0.22	1.25	8.6	0.59	1.3	500
12.79	2.67	2.48	22	112	1.48	1.36	0.24	1.26	10.8	0.48	1.47	480
13.11	1.9	2.75	25.5	116	2.2	1.28	0.26	1.56	7.1	0.61	1.33	425
13.23	3.3	2.28	18.5	98	1.8	0.83	0.61	1.87	10.52	0.56	1.51	675
12.58	1.29	2.1	20	103	1.48	0.58	0.53	1.4	7.6	0.58	1.55	640
13.17	5.19	2.32	22	93	1.74	0.63	0.61	1.55	7.9	0.6	1.48	725
13.84	4.12	2.38	19.5	89	1.8	0.83	0.48	1.56	9.01	0.57	1.64	480
12.45	3.03	2.64	27	97	1.9	0.58	0.63	1.14	7.5	0.67	1.73	880
14.34	1.68	2.7	25	98	2.8	1.31	0.53	2.7	13	0.57	1.96	660
13.48	1.67	2.64	22.5	89	2.6	1.1	0.52	2.29	11.75	0.57	1.78	620
12.36	3.83	2.38	21	88	2.3	0.92	0.5	1.04	7.65	0.56	1.58	520
13.69	3.26	2.54	20	107	1.83	0.56	0.5	0.8	5.88	0.96	1.82	680

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Alcohol	Malic acid	Ash	Alcalinity of ash	Magne-sium	Total phenols	Flava-noids	Nonflava-noid-phenols	Proantho-cyanins	Color intensity	Hue	OD280 /OD315	Pro-line
12.85	3.27	2.58	22	106	1.65	0.6	0.6	0.96	5.58	0.87	2.11	570
12.96	3.45	2.35	18.5	106	1.39	0.7	0.4	0.94	5.28	0.68	1.75	675
13.78	2.76	2.3	22	90	1.35	0.68	0.41	1.03	9.58	0.7	1.68	615
13.73	4.36	2.26	22.5	88	1.28	0.47	0.52	1.15	6.62	0.78	1.75	520
13.45	3.7	2.6	23	111	1.7	0.92	0.43	1.46	10.68	0.85	1.56	695
12.82	3.37	2.3	19.5	88	1.48	0.66	0.4	0.97	10.26	0.72	1.75	685
13.58	2.58	2.69	24.5	105	1.55	0.84	0.39	1.54	8.66	0.74	1.8	750
13.4	4.6	2.86	25	112	1.98	0.96	0.27	1.11	8.5	0.67	1.92	630
12.2	3.03	2.32	19	96	1.25	0.49	0.4	0.73	5.5	0.66	1.83	510
12.77	2.39	2.28	19.5	86	1.39	0.51	0.48	0.64	9.9	0.57	1.63	470
14.16	2.51	2.48	20	91	1.68	0.7	0.44	1.24	9.7	0.62	1.71	660
13.71	5.65	2.45	20.5	95	1.68	0.61	0.52	1.06	7.7	0.64	1.74	740
13.4	3.91	2.48	23	102	1.8	0.75	0.43	1.41	7.3	0.7	1.56	750
13.27	4.28	2.26	20	120	1.59	0.69	0.43	1.35	10.2	0.59	1.56	835
13.17	2.59	2.37	20	120	1.65	0.68	0.53	1.46	9.3	0.6	1.62	840
14.13	4.1	2.74	24.5	96	2.05	0.76	0.56	1.35	9.2	0.61	1.6	560

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค

เซตข้อมูล Forest Fires

เซตข้อมูล Forest Fires จำนวน 514 ข้อมูลประกอบด้วย

1) X คือ แกนพิกัดเชิงพื้นที่ภายในแผนที่ Montesinho Park: 1 ถึง 9

2) y คือ แกนแกนภายในแผนที่ Montesinho Park: 2 ถึง 9

3) เดือน คือ เดือนของปี: 'Jan' ถึง 'Dec'

4) วัน คือ วันของสัปดาห์: 'mon' to 'sun'

5) FFMC คือ Fuel Moisture Code ค่าความชื้นของเชื้อเพลิงชั้นบนสุด

6) DMC คือ Duff Moisture Code ค่าเฉลี่ยความชื้นของชั้นลึกปานกลางของเชื้อเพลิง

7) ดัชนี DC - DC จากระบบ FWI: 7.9 ถึง 860.6

8) ดัชนี ISI คือ Initial Spread Index ค่าการลุกลามของไฟ

9) อุณหภูมิ คือ มีหน่วยเป็นองศาเซลเซียส: 2.2 ถึง 33.30

10) RH คือ ความชื้นสัมพัทธ์ใน%: 15.0 ถึง 100

11) ความเร็วลม คือมีหน่วยเป็นความเร็วลมเป็นกิโลเมตร / ชม.: 0.40 ถึง 9.40

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
7	5	mar	fri	86.2	26.2	94.3	5.1	8.2	51	6.7
7	4	oct	tue	90.6	35.4	669	6.7	18	33	0.9
7	4	oct	sat	90.6	43.7	687	6.7	14.6	33	1.3
8	6	mar	fri	91.7	33.3	77.5	9	8.3	97	4
8	6	mar	sun	89.3	51.3	102	9.6	11.4	99	1.8
8	6	aug	sun	92.3	85.3	488	14.7	22.2	29	5.4
8	6	aug	mon	92.3	88.9	496	8.5	24.1	27	3.1
8	6	aug	mon	91.5	145	608	10.7	8	86	2.2
8	6	sep	tue	91	130	693	7	13.1	63	5.4
7	5	sep	sat	92.5	88	699	7.1	22.8	40	4
7	5	sep	sat	92.5	88	699	7.1	17.8	51	7.2
7	5	sep	sat	92.8	73.2	713	22.6	19.3	38	4
6	5	aug	fri	63.5	70.8	665	0.8	17	72	6.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
6	5	sep	mon	90.9	127	687	7	21.3	42	2.2
6	5	sep	wed	92.9	133	700	9.2	26.4	21	4.5
6	5	sep	fri	93.3	141	714	13.9	22.9	44	5.4
5	5	mar	sat	91.7	35.8	80.8	7.8	15.1	27	5.4
8	5	oct	mon	84.9	32.8	664	3	16.7	47	4.9
6	4	mar	wed	89.2	27.9	70.8	6.3	15.9	35	4
6	4	apr	sat	86.3	27.4	97.1	5.1	9.3	44	4.5
6	4	sep	tue	91	130	693	7	18.3	40	2.7
5	4	sep	mon	91.8	78.5	724	9.2	19.1	38	2.7
7	4	jun	sun	94.3	96.3	200	56.1	21	44	4.5
7	4	aug	sat	90.2	111	537	6.2	19.5	43	5.8
7	4	aug	sat	93.5	139	594	20.3	23.7	32	5.8
7	4	aug	sun	91.4	142	601	10.6	16.3	60	5.4
7	4	sep	fri	92.4	118	668	12.2	19	34	5.8
7	4	sep	mon	90.9	127	687	7	19.4	48	1.3
6	3	sep	sat	93.4	145	721	8.1	30.2	24	2.7
6	3	sep	sun	93.5	149	729	8.1	22.8	39	3.6
6	3	sep	fri	94.3	85.1	692	15.9	25.4	24	3.6
6	3	sep	mon	88.6	91.8	710	7.1	11.2	78	7.6
6	3	sep	fri	88.6	69.7	707	5.8	20.6	37	1.8
6	3	sep	sun	91.7	75.6	718	7.8	17.7	39	3.6
6	3	sep	mon	91.8	78.5	724	9.2	21.2	32	2.7
6	3	sep	tue	90.3	80.7	730	6.3	18.2	62	4.5
6	3	oct	tue	90.6	35.4	669	6.7	21.7	24	4.5
7	4	oct	fri	90	41.5	683	8.7	11.3	60	5.4
7	3	oct	sat	90.6	43.7	687	6.7	17.8	27	4
4	4	mar	tue	88.1	25.7	67.6	3.8	14.1	43	2.7
4	4	jul	tue	79.5	60.6	367	1.5	23.3	37	3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
4	4	aug	sat	90.2	96.9	624	8.9	18.4	42	6.7
4	4	aug	tue	94.8	108	647	17	16.6	54	5.4
4	4	sep	sat	92.5	88	699	7.1	19.6	48	2.7
4	4	sep	wed	90.1	82.9	736	6.2	12.9	74	4.9
5	6	sep	wed	94.3	85.1	692	15.9	25.9	24	4
5	6	sep	mon	90.9	127	687	7	14.7	70	3.6
6	6	jul	mon	94.2	62.3	443	11	23	36	3.1
4	4	mar	mon	87.2	23.9	64.7	4.1	11.8	35	1.8
4	4	mar	mon	87.6	52.2	104	5	11	46	5.8
4	4	sep	thu	92.9	137	706	9.2	20.8	17	1.3
4	3	aug	sun	90.2	99.6	631	6.3	21.5	34	2.2
4	3	aug	wed	92.1	111	654	9.6	20.4	42	4.9
4	3	aug	wed	92.1	111	654	9.6	20.4	42	4.9
4	3	aug	thu	91.7	114	661	6.3	17.6	45	3.6
4	3	sep	thu	92.9	137	706	9.2	27.7	24	2.2
4	3	sep	tue	90.3	80.7	730	6.3	17.8	63	4.9
4	3	oct	sun	92.6	46.5	692	8.8	13.8	50	2.7
2	2	feb	mon	84	9.3	34	2.1	13.9	40	5.4
2	2	feb	fri	86.6	13.2	43	5.3	12.3	51	0.9
2	2	mar	sun	89.3	51.3	102	9.6	11.5	39	5.8
2	2	mar	sun	89.3	51.3	102	9.6	5.5	59	6.3
2	2	aug	thu	93	75.3	467	7.7	18.8	35	4.9
2	2	aug	sun	90.2	99.6	631	6.3	20.8	33	2.7
2	2	aug	mon	91.1	103	639	5.8	23.1	31	3.1
2	2	aug	thu	91.7	114	661	6.3	18.6	44	4.5
2	2	sep	fri	92.4	118	668	12.2	23	37	4.5
2	2	sep	fri	92.4	118	668	12.2	19.6	33	5.4
2	2	sep	fri	92.4	118	668	12.2	19.6	33	6.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
4	5	mar	fri	91.7	33.3	77.5	9	17.2	26	4.5
4	5	mar	fri	91.2	48.3	97.8	12.5	15.8	27	7.6
4	5	sep	fri	94.3	85.1	692	15.9	17.7	37	3.6
5	4	mar	fri	91.7	33.3	77.5	9	15.6	25	6.3
5	4	aug	tue	88.8	147	615	9	17.3	43	4.5
5	4	sep	fri	93.3	141	714	13.9	27.6	30	1.3
9	9	feb	thu	84.2	6.8	26.6	7.7	6.7	79	3.1
9	9	feb	fri	86.6	13.2	43	5.3	15.7	43	3.1
1	3	mar	mon	87.6	52.2	104	5	8.3	72	3.1
1	2	aug	fri	90.1	108	530	12.5	14.7	66	2.7
1	2	aug	tue	91	121	562	7	21.6	19	6.7
1	2	aug	sun	91.4	142	601	10.6	19.5	39	6.3
1	2	aug	sun	90.2	99.6	631	6.3	17.9	44	2.2
1	2	aug	tue	94.8	108	647	17	18.6	51	4.5
1	2	aug	wed	92.1	111	654	9.6	16.6	47	0.9
1	2	aug	thu	91.7	114	661	6.3	20.2	45	3.6
1	2	sep	thu	92.9	137	706	9.2	21.5	15	0.9
1	2	sep	thu	92.9	137	706	9.2	25.4	27	2.2
1	2	sep	thu	92.9	137	706	9.2	22.4	34	2.2
1	2	sep	sun	93.5	149	729	8.1	25.3	36	3.6
6	5	mar	sat	91.7	35.8	80.8	7.8	17.4	25	4.9
6	5	aug	sat	90.2	96.9	624	8.9	14.7	59	5.8
8	6	mar	fri	91.7	35.8	80.8	7.8	17.4	24	5.4
8	6	aug	sun	92.3	85.3	488	14.7	20.8	32	6.3
8	6	aug	sun	91.4	142	601	10.6	18.2	43	4.9
8	6	aug	mon	91.1	103	639	5.8	23.4	22	2.7
4	4	sep	sun	89.7	90	704	4.8	17.8	64	1.3
3	4	feb	sat	83.9	8	30.2	2.6	12.7	48	1.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
3	4	mar	sat	69	2.4	15.5	0.7	17.4	24	5.4
3	4	aug	sun	91.4	142	601	10.6	11.6	87	4.5
3	4	aug	sun	91.4	142	601	10.6	19.8	39	5.4
3	4	aug	sun	91.4	142	601	10.6	19.8	39	5.4
3	4	aug	tue	88.8	147	615	9	14.4	66	5.4
2	4	aug	tue	94.8	108	647	17	20.1	40	4
2	4	sep	sat	92.5	121	674	8.6	24.1	29	4.5
2	4	jan	sat	82.1	3.7	9.3	2.9	5.3	78	3.1
4	5	mar	fri	85.9	19.5	57.3	2.8	12.7	52	6.3
4	5	mar	thu	91.4	30.7	74.3	7.5	18.2	29	3.1
4	5	aug	sun	90.2	99.6	631	6.3	21.4	33	3.1
4	5	sep	sat	92.5	88	699	7.1	20.3	45	3.1
4	5	sep	mon	88.6	91.8	710	7.1	17.4	56	5.4
4	4	mar	fri	85.9	19.5	57.3	2.8	13.7	43	5.8
3	4	mar	fri	91.7	33.3	77.5	9	18.8	18	4.5
3	4	sep	sun	89.7	90	704	4.8	22.8	39	3.6
3	4	sep	mon	91.8	78.5	724	9.2	18.9	35	2.7
3	4	mar	tue	88.1	25.7	67.6	3.8	15.8	27	7.6
3	5	mar	tue	88.1	25.7	67.6	3.8	15.5	27	6.3
3	4	mar	sat	91.7	35.8	80.8	7.8	11.6	30	6.3
3	4	mar	sat	91.7	35.8	80.8	7.8	15.2	27	4.9
3	4	mar	mon	90.1	39.7	86.6	6.2	10.6	30	4
3	4	aug	thu	93	75.3	467	7.7	19.6	36	3.1
3	4	aug	mon	91.5	145	608	10.7	10.3	74	2.2
3	4	aug	mon	91.5	145	608	10.7	17.1	43	5.4
3	4	sep	sun	92.4	124	681	8.5	22.5	42	5.4
3	4	sep	tue	84.4	73.4	672	3.2	17.9	45	3.1
3	4	sep	fri	94.3	85.1	692	15.9	19.8	50	5.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
3	4	oct	sun	92.6	46.5	692	8.8	20.6	24	5.4
3	5	mar	mon	87.6	52.2	104	5	9	49	2.2
3	5	sep	fri	93.5	149	729	8.1	17.2	43	3.1
3	5	oct	wed	91.4	37.9	674	5.2	15.9	46	3.6
2	5	oct	sun	92.6	46.5	692	8.8	15.4	35	0.9
4	6	feb	sat	68.2	21.5	87.2	0.8	15.4	40	2.7
4	6	mar	mon	87.2	23.9	64.7	4.1	14	39	3.1
4	6	mar	sun	89.3	51.3	102	9.6	10.6	46	4.9
4	6	sep	thu	93.7	80.9	685	17.9	17.6	42	3.1
3	5	mar	tue	88.1	25.7	67.6	3.8	14.9	38	2.7
3	5	aug	sat	93.5	139	594	20.3	17.6	52	5.8
3	6	sep	sun	92.4	124	681	8.5	17.2	58	1.3
3	6	sep	mon	90.9	127	687	7	15.6	66	3.1
9	9	jul	tue	85.8	48.3	313	3.9	18	42	2.7
1	4	sep	tue	91	130	693	7	21.7	38	2.2
2	5	sep	mon	90.9	127	687	7	21.9	39	1.8
1	2	aug	wed	95.5	99.9	513	13.2	23.3	31	4.5
8	6	aug	fri	90.1	108	530	12.5	21.2	51	8.9
1	2	jul	sat	90	51.3	296	8.7	16.6	53	5.4
2	5	aug	wed	95.5	99.9	513	13.2	23.8	32	5.4
6	5	aug	thu	95.2	132	579	10.4	27.4	22	4
5	4	mar	mon	90.1	39.7	86.6	6.2	13.2	40	5.4
8	3	sep	tue	84.4	73.4	672	3.2	24.2	28	3.6
2	2	aug	tue	94.8	108	647	17	17.4	43	6.7
8	6	sep	thu	93.7	80.9	685	17.9	23.7	25	4.5
6	5	jun	fri	92.5	56.4	433	7.1	23.2	39	5.4
9	9	jul	sun	90.1	68.6	355	7.2	24.8	29	2.2
3	4	jul	sat	90.1	51.2	424	6.2	24.6	43	1.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
5	4	sep	fri	94.3	85.1	692	15.9	20.1	47	4.9
1	5	sep	sat	93.4	145	721	8.1	29.6	27	2.7
7	4	aug	sun	94.8	108	647	17	16.4	47	1.3
2	4	sep	sat	93.4	145	721	8.1	28.6	27	2.2
2	2	aug	wed	92.1	111	654	9.6	18.4	45	3.6
2	4	aug	wed	92.1	111	654	9.6	20.5	35	4
7	4	sep	fri	92.4	118	668	12.2	19	34	5.8
7	4	mar	mon	90.1	39.7	86.6	6.2	16.1	29	3.1
6	4	aug	thu	95.2	132	579	10.4	20.3	41	4
6	3	mar	sat	90.6	50.1	100	7.8	15.2	31	8.5
8	6	sep	sat	92.5	121	674	8.6	17.8	56	1.8
8	5	sep	sun	89.7	90	704	4.8	17.8	67	2.2
6	5	mar	thu	84.9	18.2	55	3	5.3	70	4.5
6	5	aug	wed	92.1	111	654	9.6	16.6	47	0.9
6	5	aug	wed	96	127	571	16.5	23.4	33	4.5
6	5	mar	fri	91.2	48.3	97.8	12.5	14.6	26	9.4
8	6	aug	thu	95.2	132	579	10.4	20.7	45	2.2
5	4	sep	wed	92.9	133	700	9.2	21.9	35	1.8
8	6	aug	wed	85.6	90.4	610	6.6	17.4	50	4
7	4	aug	sun	91.4	142	601	10.6	20.1	39	5.4
4	4	sep	mon	90.9	127	687	7	17.7	39	2.2
1	4	aug	sat	90.2	96.9	624	8.9	14.2	53	1.8
1	4	aug	sat	90.2	96.9	624	8.9	20.3	39	4.9
6	5	apr	thu	81.5	9.1	55.2	2.7	5.8	54	5.8
2	5	aug	sun	90.2	99.6	631	6.3	19.2	44	2.7
2	5	sep	wed	90.1	82.9	736	6.2	18.3	45	2.2
8	6	aug	tue	88.8	147	615	9	14.4	66	5.4
1	3	sep	sun	92.4	124	681	8.5	23.9	32	6.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
8	6	oct	mon	84.9	32.8	664	3	19.1	32	4
5	4	feb	sun	86.8	15.6	48.3	3.9	12.4	53	2.2
7	4	oct	mon	91.7	48.5	696	11.1	16.8	45	4.5
8	6	aug	fri	93.9	136	587	15.1	20.8	34	4.9
2	5	sep	tue	91	130	693	7	17.6	46	3.1
8	6	mar	sun	89.3	51.3	102	9.6	11.5	39	5.8
1	5	sep	mon	90.9	127	687	7	21	42	2.2
6	4	mar	sat	90.8	41.9	89.4	7.9	13.3	42	0.9
7	4	mar	sun	90.7	44	92.4	5.5	11.5	60	4
6	5	mar	fri	91.2	48.3	97.8	12.5	11.7	33	4
2	5	aug	thu	95.2	132	579	10.4	24.2	28	2.7
2	2	aug	tue	94.8	108	647	17	24.6	22	4.5
4	5	sep	wed	92.9	133	700	9.2	24.3	25	4
2	2	aug	tue	94.8	108	647	17	24.6	22	4.5
2	5	aug	fri	93.9	136	587	15.1	23.5	36	5.4
6	5	apr	thu	81.5	9.1	55.2	2.7	5.8	54	5.8
4	5	sep	thu	92.9	137	706	9.2	21.5	15	0.9
3	4	sep	tue	91	130	693	7	13.9	59	6.3
2	4	sep	mon	63.5	70.8	665	0.8	22.6	38	3.6
1	5	sep	tue	91	130	693	7	21.6	33	2.2
6	5	mar	sun	90.1	37.6	83.7	7.2	12.4	54	3.6
7	4	feb	sun	83.9	8.7	32.1	2.1	8.8	68	2.2
8	6	oct	wed	91.4	37.9	674	5.2	20.2	37	2.7
5	6	mar	sat	90.6	50.1	100	7.8	15.1	64	4
4	5	sep	thu	92.9	137	706	9.2	22.1	34	1.8
2	2	aug	sat	93.5	139	594	20.3	22.9	31	7.2
7	5	sep	tue	91	130	693	7	20.7	37	2.2
6	5	sep	fri	92.4	118	668	12.2	19.6	33	6.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
8	3	sep	thu	93.7	80.9	685	17.9	23.2	26	4.9
4	4	oct	sat	90.6	43.7	687	6.7	18.4	25	3.1
7	4	aug	sat	93.5	139	594	20.3	5.1	96	5.8
7	4	sep	fri	94.3	85.1	692	15.9	20.1	47	4.9
7	3	mar	mon	87.6	52.2	104	5	11	46	5.8
4	4	mar	sat	91.7	35.8	80.8	7.8	17	27	4.9
4	4	mar	sat	91.7	35.8	80.8	7.8	17	27	4.9
4	4	sep	sun	92.4	124	681	8.5	16.9	60	1.3
1	3	sep	mon	88.6	91.8	710	7.1	12.4	73	6.3
4	5	sep	wed	92.9	133	700	9.2	19.4	19	1.3
6	5	mar	mon	90.1	39.7	86.6	6.2	15.2	27	3.1
8	6	aug	sun	90.2	99.6	631	6.3	16.2	59	3.1
3	4	sep	fri	93.3	141	714	13.9	18.6	49	3.6
4	3	mar	mon	87.6	52.2	104	5	11	46	5.8
2	2	jul	fri	88.3	150	310	6.8	13.4	79	3.6
7	4	sep	wed	90.1	82.9	736	6.2	15.4	57	4.5
4	4	sep	sun	93.5	149	729	8.1	22.9	39	4.9
7	5	oct	mon	91.7	48.5	696	11.1	16.1	44	4
8	6	aug	sat	92.2	81.8	481	11.9	20.1	34	4.5
4	6	sep	sun	93.5	149	729	8.1	28.3	26	3.1
8	6	aug	sat	92.2	81.8	481	11.9	16.4	43	4
4	4	sep	wed	92.9	133	700	9.2	26.4	21	4.5
1	5	sep	sun	93.5	149	729	8.1	27.8	27	3.1
6	4	sep	tue	91	130	693	7	18.7	43	2.7
9	4	sep	tue	84.4	73.4	672	3.2	24.3	36	3.1
4	5	sep	sat	92.5	121	674	8.6	17.7	25	3.1
8	6	aug	sun	91.4	142	601	10.6	19.6	41	5.8
2	2	sep	sat	92.5	121	674	8.6	18.2	46	1.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
1	2	sep	tue	91	130	693	7	18.8	40	2.2
6	5	sep	sat	92.5	121	674	8.6	25.1	27	4
7	5	apr	sun	81.9	3	7.9	3.5	13.4	75	1.8
6	3	apr	wed	88	17.2	43.5	3.8	15.2	51	2.7
4	4	apr	fri	83	23.3	85.3	2.3	16.7	20	3.1
2	4	aug	sun	94.2	122	590	12.9	15.4	66	4
7	4	aug	sun	91.8	175	701	13.8	21.9	73	7.6
2	4	aug	sun	91.8	175	701	13.8	22.4	54	7.6
3	4	aug	sun	91.8	175	701	13.8	26.8	38	6.3
5	4	aug	sun	91.8	175	701	13.8	25.7	39	5.4
2	4	aug	wed	92.2	91.6	504	9.6	20.7	70	2.2
8	6	aug	wed	93.1	157	667	13.5	28.7	28	2.7
3	4	aug	wed	93.1	157	667	13.5	21.7	40	0.4
8	5	aug	wed	93.1	157	667	13.5	26.8	25	3.1
8	5	aug	wed	93.1	157	667	13.5	24	36	3.1
6	5	aug	wed	93.1	157	667	13.5	22.1	37	3.6
7	4	aug	thu	91.9	109	566	8	21.4	38	2.7
6	3	aug	thu	91.6	138	622	6.3	18.9	41	3.1
2	5	aug	thu	87.5	77	695	5	22.3	46	4
8	6	aug	sat	94.2	117	581	11	23.9	41	2.2
4	3	aug	sat	94.2	117	581	11	21.4	44	2.7
3	4	aug	sat	91.8	171	692	13.7	20.6	59	0.9
7	4	aug	sat	91.8	171	692	13.7	23.7	40	1.8
2	4	aug	mon	93.6	97.9	542	14.4	28.3	32	4
3	4	aug	fri	91.6	112	573	8.9	11.2	84	7.6
2	4	aug	fri	91.6	112	573	8.9	21.4	42	3.1
6	3	aug	fri	91.1	141	629	7.1	19.3	39	3.6
4	4	aug	fri	94.3	168	684	13	21.8	53	3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
4	4	aug	tue	93.7	102	550	14.6	22.1	54	7.6
6	5	aug	tue	94.3	132	607	22.7	19.4	55	4
2	2	aug	tue	92.1	153	658	14.3	23.7	24	3.1
3	4	aug	tue	92.1	153	658	14.3	21	32	3.1
4	4	aug	tue	92.1	153	658	14.3	19.1	53	2.7
2	2	aug	tue	92.1	153	658	14.3	21.8	56	3.1
8	6	aug	tue	92.1	153	658	14.3	20.1	58	4.5
2	5	aug	tue	92.1	153	658	14.3	20.2	47	4
4	6	dec	sun	84.4	27.2	354	6.8	4.8	57	8.5
8	6	dec	wed	84	27.8	355	5.3	5.1	61	8
4	6	dec	thu	84.6	26.4	352	2	5.1	61	4.9
4	4	dec	mon	85.4	25.4	350	2.6	4.6	21	8.5
3	4	dec	mon	85.4	25.4	350	2.6	4.6	21	8.5
4	4	dec	mon	85.4	25.4	350	2.6	4.6	21	8.5
4	4	dec	mon	85.4	25.4	350	2.6	4.6	21	8.5
4	6	dec	fri	84.7	26.7	353	4.1	2.2	59	4.9
6	5	dec	tue	85.4	25.4	350	2.6	5.1	24	8.5
6	3	feb	sun	84.9	27.5	354	3.4	4.2	51	4
3	4	feb	wed	86.9	6.6	18.7	3.2	8.8	35	3.1
5	4	feb	fri	85.2	4.9	15.8	6.3	7.5	46	8
2	5	jul	sun	93.9	170	412	12.3	23.4	40	6.3
7	6	jul	wed	91.2	183	438	12.5	12.6	90	7.6
7	4	jul	sat	91.6	104	475	9	22.1	49	2.7
7	4	jul	sat	91.6	104	475	9	24.2	32	1.8
7	4	jul	sat	91.6	104	475	9	24.3	30	1.8
2	5	jul	sat	91.6	104	475	9	18.7	53	1.8
9	4	jul	sat	91.6	104	475	9	25.3	39	0.9
4	5	jul	fri	91.6	100	466	6.3	22.9	40	1.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
7	6	jul	tue	93.1	180	431	11	26.9	28	5.4
8	6	jul	tue	92.3	88.8	441	8.5	17.1	67	3.6
7	5	jun	sun	93.1	180	431	11	22.2	48	1.3
6	4	jun	sun	90.4	89.5	291	6.4	14.3	46	1.8
8	6	jun	sun	90.4	89.5	291	6.4	15.4	45	2.2
8	6	jun	wed	91.2	148	377	12.7	19.6	43	4.9
6	5	jun	sat	53.4	71	234	0.4	10.6	90	2.7
6	5	jun	mon	90.4	93.3	298	7.5	20.7	25	4.9
6	5	jun	mon	90.4	93.3	298	7.5	19.1	39	5.4
3	6	jun	fri	91.1	94.1	232	7.1	19.2	38	4.5
3	6	jun	fri	91.1	94.1	232	7.1	19.2	38	4.5
6	5	may	sat	85.1	28	114	3.5	11.3	94	4.9
1	4	sep	sun	89.6	84.1	714	5.7	19	52	2.2
7	4	sep	sun	89.6	84.1	714	5.7	17.1	53	5.4
3	4	sep	sun	89.6	84.1	714	5.7	23.8	35	3.6
2	4	sep	sun	92.4	106	758	9.9	16	45	1.8
2	4	sep	sun	92.4	106	758	9.9	24.9	27	2.2
7	4	sep	sun	92.4	106	758	9.9	25.3	27	2.7
6	3	sep	sun	92.4	106	758	9.9	24.8	28	1.8
2	4	sep	sun	50.4	46.2	707	0.4	12.2	78	6.3
6	5	sep	wed	92.6	115	777	8.8	24.3	27	4.9
4	4	sep	wed	92.6	115	777	8.8	19.7	41	1.8
3	4	sep	wed	91.2	135	818	7.2	18.5	30	2.7
4	5	sep	thu	92.4	96.2	739	8.6	18.6	24	5.8
4	4	sep	thu	92.4	96.2	739	8.6	19.2	24	4.9
6	5	sep	thu	92.8	119	784	7.5	21.6	27	2.2
5	4	sep	thu	92.8	119	784	7.5	21.6	28	6.3
6	3	sep	thu	92.8	119	784	7.5	18.9	34	7.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
1	4	sep	thu	92.8	119	784	7.5	16.8	28	4
6	5	sep	thu	92.8	119	784	7.5	16.8	28	4
3	5	sep	thu	90.7	137	823	6.8	12.9	39	2.7
6	5	sep	thu	88.1	53.3	727	5.4	13.7	56	1.8
1	4	sep	sat	92.2	102	752	8.4	24.2	27	3.1
5	4	sep	sat	92.2	102	752	8.4	24.1	27	3.1
6	5	sep	sat	92.2	102	752	8.4	21.2	32	2.2
6	5	sep	sat	92.2	102	752	8.4	19.7	35	1.8
4	3	sep	sat	92.2	102	752	8.4	23.5	27	4
3	3	sep	sat	92.2	102	752	8.4	24.2	27	3.1
7	4	sep	sat	91.2	124	795	8.5	21.5	28	4.5
4	4	sep	sat	91.2	124	795	8.5	17.1	41	2.2
1	4	sep	mon	92.1	87.7	721	9.5	18.1	54	3.1
2	3	sep	mon	91.6	108	764	6.2	18	51	5.4
4	3	sep	mon	91.6	108	764	6.2	9.8	86	1.8
7	4	sep	mon	91.6	108	764	6.2	19.3	44	2.2
6	3	sep	mon	91.6	108	764	6.2	23	34	2.2
8	6	sep	mon	91.6	108	764	6.2	22.7	35	2.2
2	4	sep	mon	91.6	108	764	6.2	20.4	41	1.8
2	5	sep	mon	91.6	108	764	6.2	19.3	44	2.2
8	6	sep	mon	91.9	112	770	6.5	15.7	51	2.2
6	3	sep	mon	91.5	130	807	7.5	20.6	37	1.8
8	6	sep	mon	91.5	130	807	7.5	15.9	51	4.5
6	3	sep	mon	91.5	130	807	7.5	12.2	66	4.9
2	2	sep	mon	91.5	130	807	7.5	16.8	43	3.1
1	4	sep	mon	91.5	130	807	7.5	21.3	35	2.2
5	4	sep	fri	92.1	99	745	9.6	10.1	75	3.6
3	4	sep	fri	92.1	99	745	9.6	17.4	57	4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
5	4	sep	fri	92.1	99	745	9.6	12.8	64	3.6
5	4	sep	fri	92.1	99	745	9.6	10.1	75	3.6
4	4	sep	fri	92.1	99	745	9.6	15.4	53	6.3
7	4	sep	fri	92.1	99	745	9.6	20.6	43	3.6
7	4	sep	fri	92.1	99	745	9.6	19.8	47	2.7
7	4	sep	fri	92.1	99	745	9.6	18.7	50	2.2
4	4	sep	fri	92.1	99	745	9.6	20.8	35	4.9
4	4	sep	fri	92.1	99	745	9.6	20.8	35	4.9
6	3	sep	fri	92.5	122	790	10.2	15.9	55	3.6
6	3	sep	fri	92.5	122	790	10.2	19.7	39	2.7
1	4	sep	fri	92.5	122	790	10.2	21.1	39	2.2
6	5	sep	fri	92.5	122	790	10.2	18.4	42	2.2
4	3	sep	fri	92.5	122	790	10.2	17.3	45	4
7	4	sep	fri	88.2	55.2	732	11.6	15.2	64	3.1
4	3	sep	tue	91.9	112	770	6.5	15.9	53	2.2
6	5	sep	tue	91.9	112	770	6.5	21.1	35	2.7
6	5	sep	tue	91.9	112	770	6.5	19.6	45	3.1
4	5	sep	tue	91.1	132	812	12.5	15.9	38	5.4
4	5	sep	tue	91.1	132	812	12.5	16.4	27	3.6
6	5	sep	sat	91.2	94.3	744	8.4	16.8	47	4.9
4	5	sep	sun	91	276	825	7.1	13.8	77	7.6
7	4	sep	sun	91	276	825	7.1	13.8	77	7.6
3	4	jul	wed	91.9	134	521	8	14.2	58	4
4	5	aug	sun	92	203	665	8.1	10.4	75	0.9
5	4	aug	thu	94.8	222	699	13.9	20.3	42	2.7
6	5	sep	fri	90.3	290	855	7.4	10.3	78	4
6	5	sep	sat	91.2	94.3	744	8.4	15.4	57	4.9
8	6	aug	mon	92.1	207	673	8.2	21.1	54	2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
2	2	aug	sat	93.7	231	715	8.4	21.9	42	2.2
6	5	mar	thu	90.9	18.9	30.6	8	8.7	51	5.8
4	5	jan	sun	18.7	1.1	171	0	5.2	100	0.9
5	4	jul	wed	93.7	101	459	11.9	19.3	39	7.2
8	6	aug	thu	90.7	194	643	6.8	16.2	63	2.7
8	6	aug	wed	95.2	218	690	18	28.2	29	1.8
9	6	aug	thu	91.6	248	754	6.3	20.5	58	2.7
8	4	aug	sat	91.6	274	819	7.7	21.3	44	4.5
2	4	aug	sun	91.6	181	613	7.6	20.9	50	2.2
3	4	sep	sun	90.5	96.7	751	11.4	20.6	55	5.4
5	5	mar	thu	90.9	18.9	30.6	8	11.6	48	5.4
6	4	aug	fri	94.8	227	707	12	23.3	34	3.1
7	4	aug	fri	94.8	227	707	12	23.3	34	3.1
7	4	feb	mon	84.7	9.5	58.3	4.1	7.5	71	6.3
8	6	sep	fri	91.1	91.3	738	7.2	20.7	46	2.7
1	3	sep	sun	91	276	825	7.1	21.9	43	4
2	4	mar	tue	93.4	15	25.6	11.4	15.2	19	7.6
6	5	feb	mon	84.1	4.6	46.7	2.2	5.3	68	1.8
4	5	feb	sun	85	9	56.9	3.5	10.1	62	1.8
4	3	sep	sun	90.5	96.7	751	11.4	20.4	55	4.9
5	6	aug	sun	91.6	181	613	7.6	24.3	33	3.6
1	2	aug	sat	93.7	231	715	8.4	25.9	32	3.1
9	5	jun	wed	93.3	49.5	298	14	28	34	4.5
9	5	jun	wed	93.3	49.5	298	14	28	34	4.5
3	4	sep	thu	91.1	88.2	732	8.3	22.8	46	4
9	9	aug	fri	94.8	227	707	12	25	36	4
8	6	aug	thu	90.7	194	643	6.8	21.3	41	3.6
2	4	sep	wed	87.9	84.8	725	3.7	21.8	34	2.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
2	2	aug	tue	94.6	212	681	9.5	27.9	27	2.2
6	5	sep	sat	87.1	291	861	4	17	67	4.9
4	5	feb	sat	84.7	8.2	55	2.9	14.2	46	4
4	3	sep	fri	90.3	290	855	7.4	19.9	44	3.1
1	4	jul	tue	92.3	96.2	450	12.1	23.4	31	5.4
6	3	feb	fri	84.1	7.3	52.8	2.7	14.7	42	2.7
7	4	feb	fri	84.6	3.2	43.6	3.3	8.2	53	9.4
9	4	jul	mon	92.3	92.1	442	9.8	22.8	27	4.5
7	5	aug	sat	93.7	231	715	8.4	26.4	33	3.6
5	4	aug	sun	93.6	235	723	10.1	24.1	50	4
8	6	aug	thu	94.8	222	699	13.9	27.5	27	4.9
6	3	jul	tue	92.7	164	576	8.9	26.3	39	3.1
6	5	mar	wed	93.4	17.3	28.3	9.9	13.8	24	5.8
2	4	aug	sun	92	203	665	8.1	24.9	42	5.4
2	5	aug	sun	91.6	181	613	7.6	24.8	36	4
8	8	aug	wed	91.7	191	636	7.8	26.2	36	4.5
2	4	aug	wed	95.2	218	690	18	30.8	19	4.5
8	6	jul	sun	88.9	263	796	5.2	29.3	27	3.6
1	3	sep	sat	91.2	94.3	744	8.4	22.3	48	4
8	6	aug	sat	93.7	231	715	8.4	26.9	31	3.6
2	2	aug	thu	91.6	248	754	6.3	20.4	56	2.2
8	6	aug	thu	91.6	248	754	6.3	20.4	56	2.2
2	4	aug	mon	92.1	207	673	8.2	27.9	33	2.2
1	3	aug	thu	94.8	222	699	13.9	26.2	34	5.8
3	4	aug	sun	91.6	181	613	7.6	24.6	44	4
7	4	sep	thu	89.7	287	849	6.8	19.4	45	3.6
1	3	aug	sat	92.1	178	605	9.6	23.3	40	4
8	6	aug	thu	94.8	222	699	13.9	23.9	38	6.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
2	4	aug	sun	93.6	235	723	10.1	20.9	66	4.9
1	4	aug	fri	90.6	270	811	5.5	22.2	45	3.6
2	5	jul	sat	90.8	84.7	377	5.6	23.8	51	1.8
8	6	aug	mon	92.1	207	673	8.2	26.8	35	1.3
8	6	aug	sat	89.4	254	768	9.7	14.2	73	2.7
2	5	aug	sat	93.7	231	715	8.4	23.6	53	4
1	3	sep	fri	91.1	91.3	738	7.2	19.1	46	2.2
5	4	sep	fri	90.3	290	855	7.4	16.2	58	3.6
8	6	aug	mon	92.1	207	673	8.2	25.5	29	1.8
6	5	apr	mon	87.9	24.9	41.6	3.7	10.9	64	3.1
1	2	jul	fri	90.7	80.9	368	16.8	14.8	78	8
2	5	sep	fri	90.3	290	855	7.4	16.2	58	3.6
5	5	aug	sun	94	47.9	101	10.7	17.3	80	4.5
6	5	aug	sun	92	203	665	8.1	19.1	70	2.2
3	4	mar	wed	93.4	17.3	28.3	9.9	8.9	35	8
7	4	sep	wed	89.7	285	844	10.1	10.5	77	4
7	4	aug	sun	91.6	181	613	7.6	19.3	61	4.9
4	5	aug	wed	95.2	218	690	18	23.4	49	5.4
1	4	aug	fri	90.5	197	650	16.3	11.8	88	4.9
7	4	aug	mon	91.5	238	731	7.5	17.7	65	4
4	5	aug	thu	89.4	266	803	5.6	17.4	54	3.1
3	4	aug	thu	91.6	248	754	6.3	16.8	56	3.1
3	4	jul	mon	94.6	160	567	16.7	17.9	48	2.7
2	4	aug	thu	91.6	248	754	6.3	16.6	59	2.7
1	4	aug	wed	91.7	191	636	7.8	19.9	50	4
8	6	aug	sat	93.7	231	715	8.4	18.9	64	4.9
7	4	aug	sat	91.6	274	819	7.7	15.5	72	8
2	5	aug	sat	93.7	231	715	8.4	18.9	64	4.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
8	6	aug	sat	93.7	231	715	8.4	18.9	64	4.9
1	4	sep	sun	91	276	825	7.1	14.5	76	7.6
6	5	feb	tue	75.1	4.4	16.2	1.9	4.6	82	6.3
6	4	feb	tue	75.1	4.4	16.2	1.9	5.1	77	5.4
2	2	feb	sat	79.5	3.6	15.3	1.8	4.6	59	0.9
6	5	mar	mon	87.2	15.1	36.9	7.1	10.2	45	5.8
3	4	mar	wed	90.2	18.5	41.1	7.3	11.2	41	5.4
6	5	mar	thu	91.3	20.6	43.5	8.5	13.3	27	3.6
6	3	apr	sun	91	14.6	25.6	12.3	13.7	33	9.4
5	4	apr	sun	91	14.6	25.6	12.3	17.6	27	5.8
4	3	may	fri	89.6	25.4	73.7	5.7	18	40	4
8	3	jun	mon	88.2	96.2	229	4.7	14.3	79	4
9	4	jun	sat	90.5	61.1	253	9.4	24.5	50	3.1
4	3	jun	thu	93	104	317	10.8	26.4	35	2.7
2	5	jun	thu	93.7	122	350	18	22.7	40	9.4
4	3	jul	thu	93.5	85.3	395	9.9	27.2	28	1.3
4	3	jul	sun	93.7	101	423	14.7	26.1	45	4
7	4	jul	sun	93.7	101	423	14.7	18.2	82	4.5
7	4	jul	mon	89.2	104	432	6.4	22.6	57	4.9
9	9	jul	thu	93.2	114	560	9.5	30.2	25	4.5
4	3	jul	thu	93.2	114	560	9.5	30.2	22	4.9
3	4	aug	sun	94.9	130	587	14.1	23.4	40	5.8
8	6	aug	sun	94.9	130	587	14.1	31	27	5.4
2	5	aug	sun	94.9	130	587	14.1	33.1	25	4
2	4	aug	mon	95	136	596	21.3	30.6	28	3.6
5	4	aug	tue	95.1	141	606	17.7	24.1	43	6.3
5	4	aug	tue	95.1	141	606	17.7	26.4	34	3.6
4	4	aug	tue	95.1	141	606	17.7	19.4	71	7.6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

X	Y	month	day	FFMC	DMC	DC	ISI	temp	RH	wind
4	4	aug	wed	95.1	141	606	17.7	20.6	58	1.3
4	4	aug	wed	95.1	141	606	17.7	28.7	33	4
4	4	aug	thu	95.8	152	624	13.8	32.4	21	4.5
1	3	aug	fri	95.9	158	634	11.3	32.4	27	2.2
1	3	aug	fri	95.9	158	634	11.3	27.5	29	4.5
6	6	aug	sat	96	164	643	14	30.8	30	4.9
6	6	aug	mon	96.2	176	662	16.8	23.9	42	2.2
4	5	aug	mon	96.2	176	662	16.8	32.6	26	3.1
3	4	aug	tue	96.1	181	671	14.3	32.3	27	2.2
6	5	aug	tue	96.1	181	671	14.3	33.3	26	2.7
7	5	aug	tue	96.1	181	671	14.3	27.3	63	4.9
8	6	aug	tue	96.1	181	671	14.3	21.6	65	4.9
7	5	aug	tue	96.1	181	671	14.3	21.6	65	4.9
4	4	aug	tue	96.1	181	671	14.3	20.7	69	4.9
2	4	aug	wed	94.5	139	689	20	29.2	30	4.9
4	3	aug	wed	94.5	139	689	20	28.9	29	4.9
1	2	aug	thu	91	163	744	10.1	26.7	35	1.8
1	2	aug	fri	91	167	753	7.1	18.5	73	8.5
2	4	aug	fri	91	167	753	7.1	25.9	41	3.6
1	2	aug	fri	91	167	753	7.1	25.9	41	3.6
5	4	aug	fri	91	167	753	7.1	21.1	71	7.6
6	5	aug	fri	91	167	753	7.1	18.2	62	5.4
8	6	aug	sun	81.6	56.7	666	1.9	27.8	35	2.7
4	3	aug	sun	81.6	56.7	666	1.9	27.8	32	2.7
2	4	aug	sun	81.6	56.7	666	1.9	21.9	71	5.8
7	4	aug	sun	81.6	56.7	666	1.9	21.2	70	6.7
1	4	aug	sat	94.4	146	615	11.3	25.6	42	4
6	3	nov	tue	79.5	3	107	1.1	11.8	31	4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้