**Lab03 วิชา การเรียนรู้แบบไม่มีผู้สอนของเครื่อง (Unsupervised Machine Learning)**

หัวข้อ

1. k-Mean Clustering

คำสั่งส่วนที่ 1

1. ให้นักศึกษาเขียนโปรแกรมด้วยภาษาไพธอน เพื่อหาคำตอบของการจัดกลุ่มข้อมูลที่กำหนดให้ในใบงานด้วยวิธี k-Mean ใน**การทำงานรอบที่ 1 และ 2** ตามลำดับ โดยใช้ centroid ที่กำหนดให้เป็นค่าเริ่มต้น และใช้ Euclidean Distance เป็นเมตริกซ์สำหรับการวัดค่า proximity (จะเขียนฟังก์ชันเองหรือใช้ไลบรารี่ก็ได้)
2. บันทึกผลลัพธ์ของ Distance ระหว่างข้อมูลแต่ละจุดและ centroid ทุกจุด รวมถึงระบุด้วยว่าจุดข้อมูลแต่ละจุดจัดอยู่ในคลัสเตอร์เดียวกับจุด centroid ใด ลงในส่วนบันทึกผลการทดลอง (ตาราง) ส่วนที่ 1

\* โจทย์กำหนดให้เป็นผลจาก iteration ที่ 1 และ 2 เท่านั้น ถ้าใส่ค่าผลลัพธ์จาก iteration สุดท้ายมาจะถือว่าผิดคำสั่ง

คำสั่งส่วนที่ 2

1. ใช้ชุดข้อมูล load\_iris()

iris = datasets.load\_iris()  
X = iris.data  
y = iris.target

1. ตรวจสอบหมายเลข Label ของข้อมูลแต่ละตัว บันทึกผลลงในช่อง target
2. ใช้ kmean ในการวิเคราะห์คลัสเตอร์ของ load\_iris() โดยใช้ข้อมูลทั้ง 100% ในการเทรน โดยไม่กำหนด target ให้ model (เทรนโดยใช้แค่ data) บันทึกผลลัพธ์ของการทำคลัสเตอริ่ง
3. รายละเอียดการบันทึกผล
   1. บันทึกผลลงในใบงาน (หน้า 2) ไม่ต้องกังวลเกี่ยวกับแบบฟอร์มหรือจำนวนหน้า แม้ว่า ตารางจะตกหน้า แยกออกจากกัน หรือทำให้มีพื้นที่ว่างในบางหน้ามากเกินไป ก็ไม่มีปัญหา
   2. เซฟไฟล์นี้ ตั้งชื่อว่า Lab03\_result\_รหัสนักศึกษา.docx ส่งพร้อมโค้ด ตั้งชื่อว่า Lab03\_code\_รหัสนักศึกษา.py (แยกไฟล์กันมา)
   3. ส่งไฟล์โดยอัปโหลดไปยัง <https://silpakorn-my.sharepoint.com/:f:/g/personal/aonpong_p_su_ac_th/EmRn8pOWnXJOvxtOcoRs98IBI6ex0H6-AwN_m-BxJYG3gg?e=XSs2Zt>

\* พยายามควบคุมพารามิเตอร์ โค้ดที่ส่งจะต้องไม่มี Warning (เช่น Convergence Warning หรือ Warning อื่น ๆ)

**630710440 พีท อ่อนทอง**

**บันทึกผลการทดลอง ส่วนที่ 1**

**รอบที่ 1**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Point** | **A1** | **A2** | **d1** | **d2** | **d3** | **Cluster** |
| c1=1 | 0 | 0 | 0.00 | 9.90 | 19.42 | 1 |
| 2 | 2 | 5 | 5.39 | 5.39 | 17.03 | 1 |
| 3 | 3 | 1 | 3.16 | 7.21 | 16.28 | 1 |
| 4 | 5 | 4 | 6.40 | 3.61 | 14.00 | 2 |
| c2=5 | 7 | 7 | 9.90 | 0.00 | 12.37 | 2 |
| 6 | 11 | 1 | 11.05 | 7.21 | 8.54 | 2 |
| 7 | 13 | 5 | 13.93 | 6.32 | 6.08 | 3 |
| 8 | 15 | 2 | 15.13 | 9.43 | 4.47 | 3 |
| 9 | 16 | 6 | 17.09 | 9.06 | 3.61 | 3 |
| c3=10 | 19 | 4 | 19.42 | 12.37 | 0.00 | 3 |

**รอบที่ 2**

|  |  |  |
| --- | --- | --- |
| **Iter 2** | **A1** | **A2** |
| c1 | **1.66666667** | **2.** |
| c2 | **7.66666667** | **4.** |
| c3 | **15.75** | **4.25** |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Point** | **A1** | **A2** | **d1** | **d2** | **d3** | **Cluster** |
| 1 | 0 | 0 | 2.60 | 8.65 | 16.31 | 1 |
| 2 | 2 | 5 | 3.02 | 5.75 | 13.77 | 1 |
| 3 | 3 | 1 | 1.67 | 5.55 | 13.16 | 1 |
| 4 | 5 | 4 | 3.89 | 2.67 | 10.75 | 2 |
| 5 | 7 | 7 | 7.31 | 3.07 | 9.17 | 2 |
| 6 | 11 | 1 | 9.39 | 4.48 | 5.76 | 2 |
| 7 | 13 | 5 | 11.72 | 5.43 | 2.85 | 3 |
| 8 | 15 | 2 | 13.33 | 7.60 | 2.37 | 3 |
| 9 | 16 | 6 | 14.88 | 8.57 | 1.77 | 3 |
| 10 | 19 | 4 | 17.45 | 11.33 | 3.26 | 3 |

บันทึกผลการทดลอง ส่วนที่ 2

**ผลการทดลอง k-mean**

|  |  |
| --- | --- |
| K-mean result | |
| k-Mean | Target |
| 2 | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0  0 0 |
| 3 | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  1 1 1 1 1 1 1 1 1 1 1 1 1 0 2 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  2 2 2 0 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 0 2 0 0 0 0 2 0 0 0 0  0 0 2 2 0 0 0 0 2 0 2 0 2 0 0 2 2 0 0 0 0 0 2 0 0 0 0 2 0 0 0 2 0 0 0 2 0  0 2 |
| 4 | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 2 0 2 0 2 0 2 2 2 2 0 2 0 2 2 0 2 0 2 0 0  0 0 0 0 0 2 2 2 2 0 2 0 0 0 2 2 2 0 2 2 2 2 2 0 2 2 3 0 3 0 3 3 2 3 3 3 0  0 3 0 0 3 3 3 3 0 3 0 3 0 3 3 0 0 3 3 3 3 3 0 0 3 3 0 0 3 3 3 0 3 3 3 0 0  3 0 |
| 5 | 4 1 1 1 4 4 1 4 1 1 4 1 1 1 4 4 4 4 4 4 4 4 1 4 1 1 4 4 4 1 1 4 4 4 1 1 4  4 1 4 4 1 1 4 4 1 4 1 4 1 0 0 0 2 0 2 0 2 0 2 2 2 2 0 2 0 2 2 0 2 0 2 0 0  0 0 0 0 0 2 2 2 2 0 2 0 0 0 2 2 2 0 2 2 2 2 2 0 2 2 3 0 3 0 3 3 2 3 3 3 0  0 3 0 0 3 3 3 3 0 3 0 3 0 3 3 0 0 3 3 3 3 3 0 0 3 3 0 0 3 3 3 0 3 3 3 0 0  3 0 |
| 6 | 4 1 1 1 4 4 1 4 1 1 4 1 1 1 4 4 4 4 4 4 4 4 1 4 1 1 4 4 4 1 1 4 4 4 1 1 4  4 1 4 4 1 1 4 4 1 4 1 4 1 5 5 5 2 5 2 5 2 5 2 2 5 2 5 2 5 2 2 5 2 0 2 5 5  5 5 5 5 5 2 2 2 2 0 2 5 5 5 2 2 2 5 2 2 2 2 2 5 2 2 3 0 3 0 3 3 2 3 3 3 0  0 3 0 0 0 0 3 3 0 3 0 3 0 3 3 0 0 0 3 3 3 0 0 0 3 0 0 0 3 3 0 0 3 3 0 0 0  0 0 |
| 7 | 4 1 1 1 4 4 1 4 1 1 4 1 1 1 4 4 4 4 4 4 4 4 1 4 1 1 4 4 4 1 1 4 4 4 1 1 4  4 1 4 4 1 1 4 4 1 4 1 4 1 5 5 5 2 5 0 5 2 5 2 2 5 2 5 2 5 0 2 5 2 0 2 0 5  5 5 5 5 5 2 2 2 2 0 0 5 5 5 2 2 2 5 2 2 2 2 2 5 2 2 6 0 3 6 6 3 2 3 6 3 6  6 6 0 0 6 6 3 3 0 6 0 3 0 6 3 0 0 6 3 3 3 6 0 0 3 6 6 0 6 6 6 0 6 6 6 0 6  6 0 |

\*เพราะคลัสเตอริ่ง เป็นวิธีการที่ไม่สนใจว่า label ของกลุ่มคืออะไร รู้เพียงแค่ว่าเป็นกลุ่มเดียวกันหรือไม่ ดังนั้น เราอาจพบเหตุการณ์ที่หมายเลข label ไม่ตรงกันได้บ่อยครั้ง “จากโมเดลที่เทรนมาด้วยกัน ถ้า label ที่ควรเป็นกลุ่มเดียวกันมีหมายเลขเดียวกัน ถือว่าใช้ได้”

\*\*ผลการทดลองในตารางนี้ เมื่อวางข้อมูลทั้งหมดลงมาแล้วอาจยาวหลายหน้า ไม่ต้องสนใจ วางได้เลย