

## การทดลองจัดกลุ่มข้อมูล 2 มิติ ด้วยวิธี k-Means และ Hierarchical Clustering

กำหนดชุดข้อมูล 4 ชุด ชุดละ 100 จุด ดังไฟล์ต่อไปนี้

- data2Dset1.csv
- data2Dset2.csv
- data2Dset3.csv
- data2Dset4.csv

### คำสั่ง

1. เขียนโปรแกรมภาษาไพทอน อ่านข้อมูลจากไฟล์ชุดข้อมูลที่กำหนดให้ ส่วนของโค้ดที่แนะนำให้ใช้เพื่ออ่านข้อมูล

```
data = pd.read_csv('data2Dset1.csv', header=None)
```
2. เขียนโปรแกรม plot จุดข้อมูล
3. เขียนโปรแกรมจัดกลุ่มชุดข้อมูลที่อ่านเข้ามา โดยใช้วิธี k-Means
  - 1) ให้เลือกใช้ k ที่ต่างกัน 3 ค่า
  - 2) plot ผลการจัดกลุ่มที่ได้ สำหรับแต่ละค่า k
  - 3) วิเคราะห์ผลการทดลองที่ได้ (ควรมี k 1 ค่าที่นักศึกษาคิดว่าดีที่สุดสำหรับชุดข้อมูลดังกล่าว)
4. เขียนโปรแกรมจัดกลุ่มชุดข้อมูลที่อ่านเข้ามา โดยใช้วิธี Hierarchical Clustering
  - 1) ให้เลือกใช้ method ที่ต่างกัน 3 แบบ แสดง dendrogram ที่ได้แต่ละแบบ
  - 2) เลือก cut-off โดยกำหนด criterion='distance' และให้นักศึกษาเลือกระบุค่า t ที่คิดว่าเหมาะสม สำหรับแต่ละ dendrogram ที่ได้ในข้อ 1)
  - 3) Plot ผลการจัดกลุ่ม ที่ได้แต่ละแบบในข้อ 2)
5. ทำการทดลองตามข้อ 1 ถึง ข้อ 4 ให้ครบทั้ง 4 ชุดข้อมูล
6. เขียนบรรยายสรุปผลการทดลอง แสดงความคิดเห็น วิธีใด เหมาะกับ ชุดข้อมูลแบบไหน แต่ละวิธีมีข้อดี/ข้อเสีย อย่างไร
7. ทำรายงาน ที่ประกอบด้วยส่วนของโค้ด และส่วนของผล แล้วแปลงเป็น pdf ชื่อไฟล์ 6xxxxxxxxxx.pdf เมื่อ 6xxxxxxxxxx แทนรหัสนักศึกษา ส่งผ่านลิงค์ฟอร์มที่กำหนด

