

進階資料探勘第二次程式作業

繳交期限: 到 2023/12/14

繳交方式: 寄 email 給助教，信件主旨請打上:[資料探勘第二次程式作業繳交]，信件內容要有學號、姓名、2 張點散圖、2 份 python 程式碼(DBSCAN 一份，Apriori 和 FP-Tree 合為一份)

作業主題:

1. 利用 python 練習使用 DBSCAN 工具
2. 利用 python 練習使用 Apriori 和 FP-Tree 工具

作業說明:

1. 利用 python 練習使用 DBSCAN 工具

1.1 請使用 sklearn.datasets 函式庫裡面的 `make_blobs` 函數來

建立資料，以下為參數設定:

- A. `n_samples=250`: 固定建立 250 個資料
- B. `n_features=2`: 固定每個資料都有兩個特徵值
- C. `centers=4`: 固定先以初始的 4 群來建立資料 (非真實分群後的 4 群)
- D. `shuffle=True`: 固定設為 True
- E. `cluster_std`: 可自行調整，但至少要 ≥ 1 (才夠混亂)
- F. `random_state`: 可自行調整

1.2 請使用 `sklearn.cluster` 函式庫裡面的 **DBSCAN** 函數練習

DBSCAN 程式:

A. 參數 `eps` 和參數 `min_samples` 皆自行設定

B. 分群後的結果**至少要有不同的 2 群**

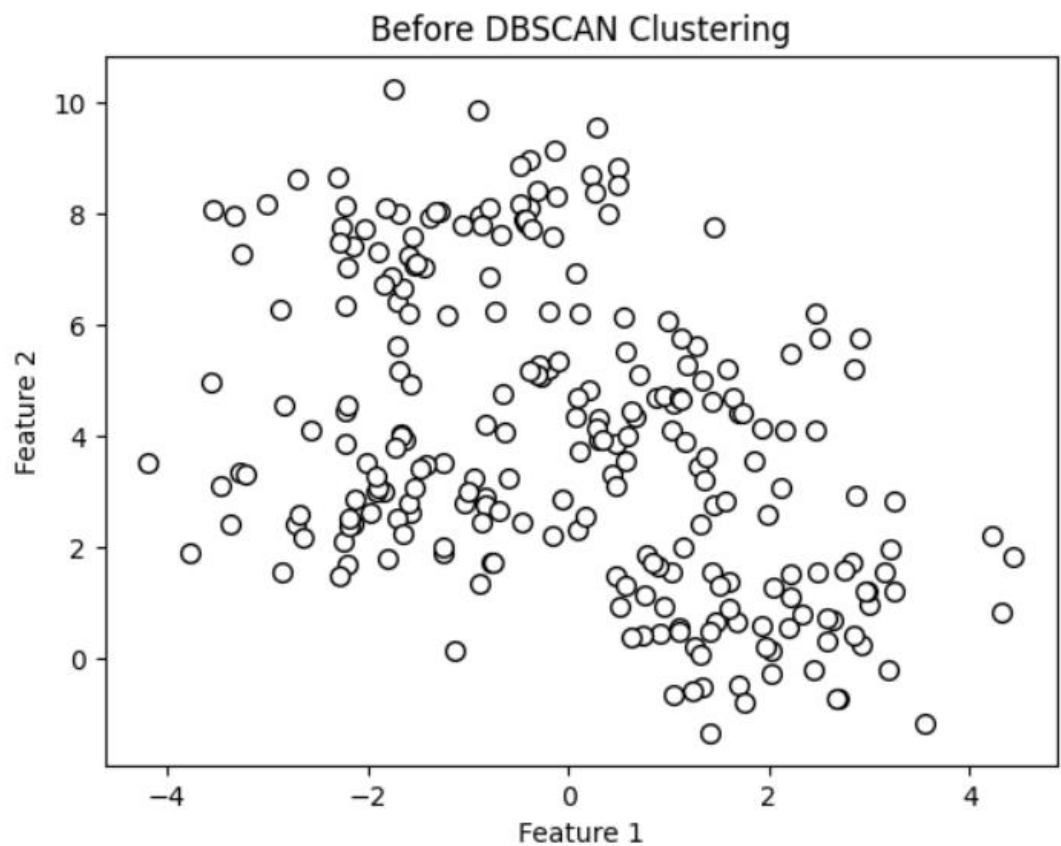
1.3 分群前的資料點散圖以及分群後的資料點散圖皆須附上

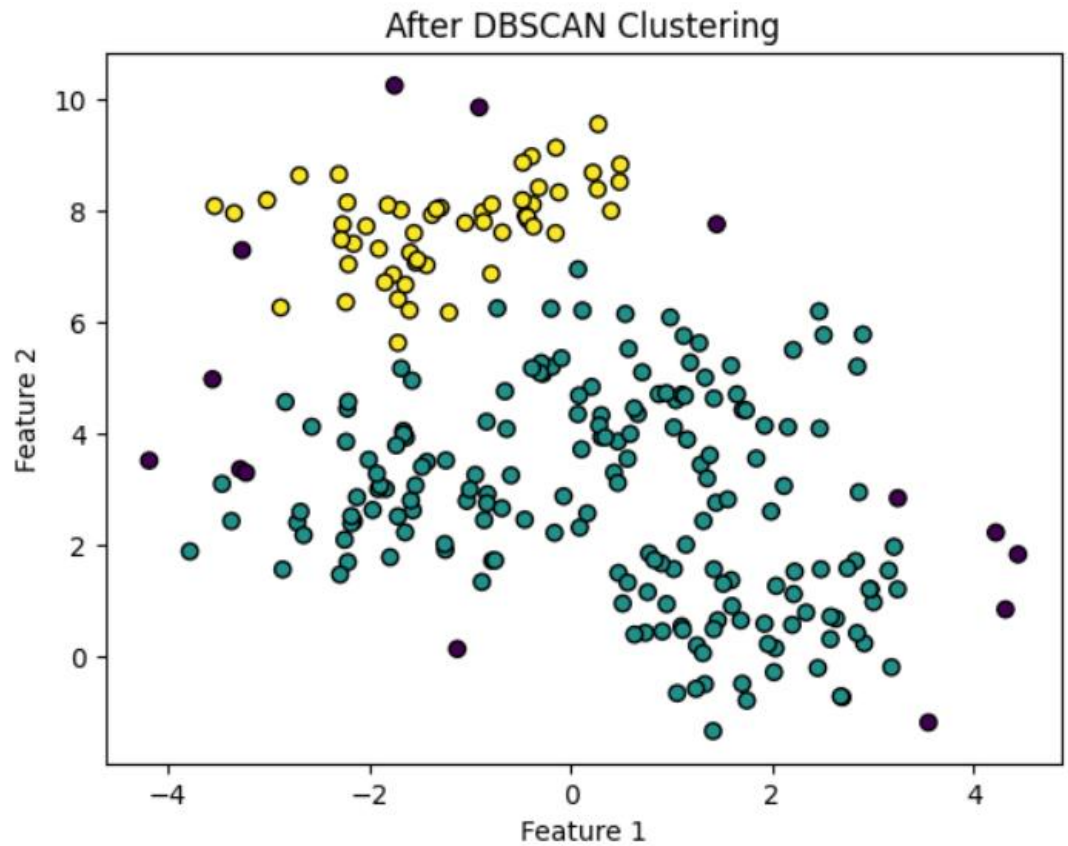
(.jpg 或.png 檔):

A. 圖的 title、特徵空間座標皆須標明清楚

B. 分群後的圖，不同群要用不同的顏色表示，並且

noise points 也要以不同於其他群的顏色表示





2. 利用 python 練習使用 Apriori 和 FP-Tree 工具

2.1 請根據以下交易紀錄表格建立資料 (Hint: python 2 維 list):

TID	Items
1	A, B
2	B, C, D
3	A, C, D, E
4	A, D, E
5	A, B, C
6	A, B, C, D
7	B, C
8	A, B, C
9	A, B, D
10	B, C, E

2.2 請使用 mlxtend.frequent_patterns 函式庫裡的 apriori 和

association_rules 函數練習 Apriori 程式:

A. 請使用 mlxtend.preprocessing 函式庫裡的

TransactionEncoder 函數以及 pandas 函式庫，將交

易資料做轉換，轉換成適用於 apriori 函數的 pandas

的 DataFrame 類型的資料(如下圖 code)

```
# 資料預處理
te = TransactionEncoder()
te_ary = te.fit(transactions).transform(transactions)
df = pd.DataFrame(te_ary, columns=te.columns_)
```

B. 使用 apriori 函數得到 frequent patterns，apriori 函數

的參數要有 DataFrame 類型的交易資料(df)、

min_support(自行調整)、use_colnames=True(才會是

items 的真正名稱)

C. 使用 association_rules 函數得到 association_rules，

參數要有之前得到的 frequent patterns、

metric="confidence"(以 confidence 定義作為基準)、

min_threshold(自行調整)

D. 程式碼要可以 print 出 frequent patterns 和

association_rules 的結果

2.3 請使用 pyfpgrowth 函式庫的 find_frequent_patterns 函數

和 generate_association_rules 函數練習 FP-Tree 程式:

- A. 必須先 pip install pyfpgrowth
- B. find_frequent_patterns 函數詳情使用方法請自行查閱，最低頻數可自行調整(但要 ≥ 2)
- C. generate_association_rules 函數詳情使用方法請自行查閱，最低 confidence threshold 可自行調整
- D. 程式碼要可以 print 出 frequent patterns 和 association_rules 的結果