進階資料探勘第一次程式作業

繳交期限: 到 2023/11/30

繳交方式: 寄 email 給助教，信件主旨請打上:[資料探勘第一次程式作業繳交]，信件內容要有學號、姓名、一份報告(PDF 檔)、2 份程式碼(python 檔)

作業主題: 利用 python 練習 KMeans 演算法

作業說明:

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
2. 請寫一份自己的 KMeans 演算法(不能用任何現有的 KMeans 函數)，演算法分 4 群
3. 請用 sklearn.cluster 函式庫裡 KMeans 函數寫一份程式:

A. KMeans 函數參數 n_clusters 固定為 4 (分 4 群)

B. KMeans 函數參數 init 固定為'random' (隨機初始化群中心點)

C. 其餘參數可自行調整 (參數說明請參考附件 PPT)

4. 須繳交兩份 python 檔程式碼(一份自己寫的版本，一份利用現成的 KMeans 函數寫的版本)以及一份程式報告 PDF 檔。自己寫的版本切勿抄襲!

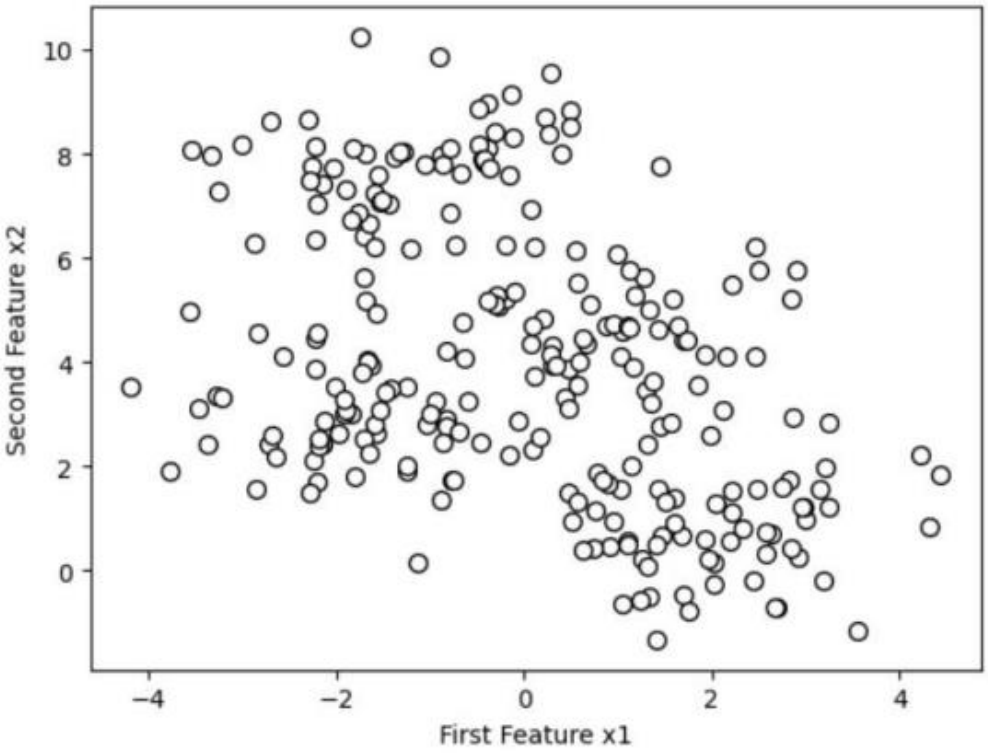
5. 報告需要有以下內容:

A. 自己寫的 KMeans 演算法、用現成的 KMeans 函數寫的皆要做程式碼說明，說明得愈完整、愈清楚，分數愈高

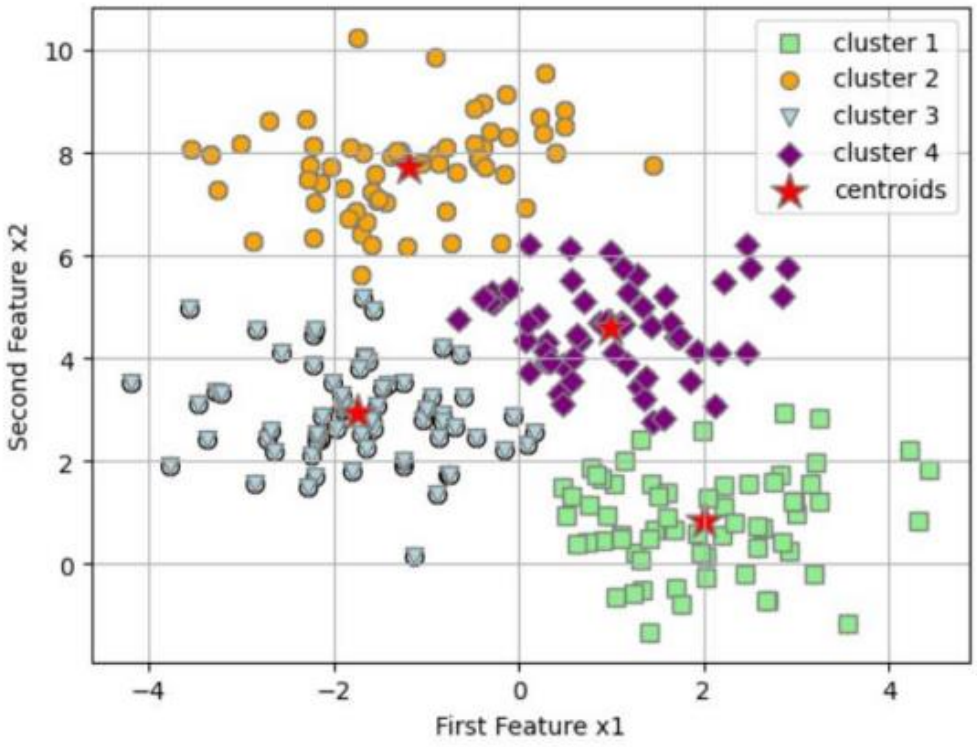
B. 現成的 KMeans 函數請針對參數調整的原因或結果比較做說明即可，例: 為何 max_iter 設為 400，說明愈清楚分數愈高

C. 自己寫的 KMeans 演算法、用現成的 KMeans 函數寫的皆需要附上分群前的資料點散圖(如 Fig. 1.)以及分群後的資料點散圖(如 Fig. 2.)，總共要附上 4 張圖。請標明清楚群中心點、每個點屬於哪一群(不同群請以不同顏色標示)，圖旁邊需附上圖示，標明特徵空間座標，

且須標明清楚圖片是屬於哪一個程式版本的圖:



- Fig. 1.上圖為分群前的圖



- Fig. 2.上圖為分群後的圖