Data Science HW2

B10815057 廖聖郝

1. 執行程式:
2. pip install pandas
3. pip install sklearn
4. python clustering.py
5. 程式架構流程:
6. import必要函式庫



1. 讀取檔案，包括資料集與測試資料



1. 用kmeans演算法將資料分成5類



1. 取得分群結果



1. 創建空的array，size預設為測試資料數，用於存放判斷結果



1. for迴圈跑過每筆測試資料判斷是否同一群，將結果輸出到ans



1. 將ans這個array打包成dataframe



1. 設定index欄位名稱為"id"，index從int轉為float (繳交規定)



1. 輸出到csv檔，並設定結果欄位名稱為"ans"



1. Kmeans解釋

kmeans 是一種非監督式學習，且是一種 clustering 常用的方法。

流程:

* 要先決定要分 k 群，接著隨機選 k 個點作為每群中心點。
* 將資料集中的資料，分群到最靠近自己中心點。
* 重新計算每群的中心點。 重複 2.和 3.，直到中心點不再移動時，kmeans 結束。

