

DataScience HW2

資料集說明：

目的，透過 8 個 attributes 去分析各個 data 是否為同一類。

data.csv：

id	feature1	feature2	feature3	feature4	feature5	feature6	feature7	feature8
0	0.58	0.61	0.47	0.13	0.5	0	0.48	0.22
1	0.43	0.67	0.48	0.27	0.5	0	0.53	0.22
2	0.64	0.62	0.49	0.15	0.5	0	0.53	0.22
3	0.58	0.44	0.57	0.13	0.5	0	0.54	0.22
4	0.42	0.44	0.48	0.54	0.5	0	0.48	0.22

Id 為每項 data 的編號，每項 data 總共有 8 個 attributes。全部有 1484 筆資料。

test.csv：

index	0	1
0	766	405
1	387	1020
2	565	14
3	549	272
4	18	583

在此文件內，顯示的是我們需要比較是否為相同 cluster 的 data 編號，index 0 要比較的資料為 id 766 以及 id 405。

Submission.csv：

index	ans
0	
1	
2	
3	
4	

需要將預測結果(0：不為同一類, 1：同一類) 寫進 Submission.csv 內，也就是在 index 0 的 ans 需要把 id 766, id 405 是否為同一類的結果寫入，同理 index 1 的 ans 需要寫入 id 387, id 1020 是否為同一類的結果，以此類推(總共需要預測 300 筆資料(0~299))。