

# 2019 ECT 作業六

一、 請用 python 依照步驟對 BreastCancer.csv 進行 KNN 及 KMeans 分

析，過程中對所有重要程式步驟進行截圖並加以說明，越詳盡越好。(80%)

(a). 將 radius\_mean 及 area\_mean 切為 feature, diagnosis 切為 target

(b). 切分資料及與訓練集, 設 test\_size=0.34, random\_state=5

(c). 用 KNeighborsClassifier 進行 KNN 分類, n\_neighbors 設為 6

(d). 用 metrics 算出此模型對於測試集的準確度

(e). 運用 matplotlib 中的 scatter 圖, x 軸設為 radius\_mean, y 軸設為

area\_mean, c 設為 label, 印出測試分類圖形。

(f). 用 cluster.Mean 設 n\_clusters=2

(g). 用 fit\_predict 對切分的 feature 進行預測

(h). 運用 matplotlib 中的 scatter 圖, x 軸設為 radius\_mean, y 軸設為

area\_mean, c 設為分群結果, 印出分類圖形。

(i). 移除 area\_mean 中大於 2000 的資料

(j). 重複上述 a~e 的動作, 同時回答問題: 在此案例中移除與分布較遠的

資料是否有達到更好的效果?

二、 請用 weka 對 BreastCancer.csv, 進行 IBK(knn) k 設為 6 及

simplekMeans 進行分析, Percentage split 設為 66%, 截圖並附上過程及準

確率。(20%)

- 繳交期限： 4/10（三）中午 12:00
- 請轉檔為 PDF 格式, 檔名為：ECT\_HW6\_學號\_版本.pdf,  
並同時附上 python 的 ipynb.檔,命名格式同上
- 上傳至 LMS 作業區,遲交一天扣該次作業成績 5%
- 補交請上傳至 HW6 補交區, 僅開放到 4/17(三)中午 12:00