

2020 ECT 作業三

1. 用 Weka 軟體對 BreastCancerNUM.arff, BreastCancerNOM.arff 分別建立 Linear Regression model 及 Logistic model, 選擇 “Use training set”, 設定 Attribute: type 為 diagnosis, 在過程中對重要步驟截圖加以說明, 並回答以下問題：
 - (a) LinearRegression model 中, 多少因素與乳癌為正相關?多少為負相關? (10%)
 - (b) 舉出三個 LinearRegression model 認為與乳癌無關的因素(10%)
 - (c) 比較 Logistic model 與 Linear Regression model 的均方根誤差, 兩者差距多少? 哪個模型表現較好? (30%)
 2. 用 python 對 BreastCancer.csv 建立 LinearRegression model 及 Logistic model(`solver='liblinear'`), 在過程中對重要步驟截圖加以說明, 並回答以下問題：

`y.values.ravel()`

 - (a) 以 `coef_` function 印出各 Attribute 的相關係數, 請問 LinearRegression model 中, Attribute ‘area_mean’ 的係數是多少?(10%)
 - (b) LinearRegression model 中, 影響最大的因素為何?(20%)
 - (c) 以 `score()` function 印出兩模型的正確率, 請問何者較高?(20%)
- 繳交期限: 4/8(三) 中午 12:00

- 請轉檔為 **PDF** 格式, 檔名為:**ECT_HW3_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 **HW3 補交區**, 僅開放到 **4/15(三) 中午 12:00**