

ECT_HW3

2020

第一大題

第一大題

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：

第一大題(a)-題目

(a) LinearRegression model中, 多少因素與乳癌為正相關?多少為負相關? (10%)

第一大題(b)-題目

(b) 舉出三個LinearRegression model認為與乳癌無關的因素
(10%)

第一大題(c)-題目

(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%)

第二大題

第二大題

用python對BreastCancer.csv建立LinearRegression model及 Logistic model, 在過程中對重要步驟截圖加以說明,並回答以下問題：

```
logistic_reg.fit(x, y.values.ravel())
```


第二大題(a)-題目

(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%)

第二大題(b)-題目

(b) LinearRegression model中, 影響最大的因素為何?(20%)

第二大題(c)-題目

(c) 以score() function印出兩模型的正確率，請問何者較高?(20%)