

1. 台北的蔡先生接手了家中出租套房的事業，他決定透過自己資料分析的專業來將過去出租的 7 間套房來建立 LinearRegression model，藉由該模型來幫助他決定未來新套房的出租價格。請利用 house. arff 建立 LinearRegression，選擇 “Use training set”，在過程中對重要步驟截圖並加以說明，並回答以下問題：

house. arff 資料格式說明：

houseSize: 房屋大小; lotSize: 佔地大小; bedrooms: 房間數;

granite: 是否使用花崗岩; bathroom: 是否有廁所;

sellingPrice: 租金價格/月

新套房的資訊如下：

房屋大小3873、佔地大小9527、五間房間、有使用花崗岩、無廁所

(a) 請說明套房的租金價格分別與哪兩個因素最要緊及最無關？(15%)

Linear Regression Model

| | |
|--------------------------------|---|
| sellingPrice = | 最要緊的是 bedrooms，係數 43494.8401，在取其絕對值後，為方程式中最大，也就是影響最巨。 |
| -26.4701 * houseSize + | |
| 7.8362 * lotSize + | 最無關的是 granite，係數 0，在取其絕對值後，為方程式中最小，也就是影響最微。 |
| 43494.8401 * bedrooms + | |
| 42583.3146 * bathroom + | |
| -31877.7173 | |

(b) 請說明套房的租金價格和哪一個因素呈現負相關？(15%)

houseSize 係數為負。

(c) 以此模型預測新套房的租金金額應該為多少？(15%)

$-26.4701 * 3873 + 7.8362 * 9527 + 43494.8401 * 5 + 42583.3146 * 0 - 31877.7173 = 157733.2633$ 。

(d) 蔡先生意識到加裝廁所能讓租金上升，而打造一間廁所的金額為 11 萬。請問若在新套房裡加裝廁所，需租出去幾個月能夠回本？(15%)

$-26.4701 * 3873 + 7.8362 * 9527 + 43494.8401 * 5 + 42583.3146 * 1 - 31877.7173 = 200316.5779$ 。

$11000 / (200316.5779 - 157733.2633) = 2.5821$... 需要三個月才可回本。

2. 用 Weka 軟體對 diabetes. arff 建立 Logistic regression，選擇 “Use training set”，設定 Attribute: class 為 Output，在過程中對重要步驟截圖並加以說明，並回答以下問題：

(a) 請說明哪一個因素與 output 結果呈現正相關？(10%)

Pres & insu 的係數為正，呈現正相關。

Intercept 為正，但是它是 not attribute but constant。

Logistic Regression with ridge parameter of 1.0E-8
Coefficients...

| | Class |
|-----------|-----------------|
| Variable | tested_negative |
| ===== | |
| preg | -0.1232 |
| plas | -0.0352 |
| pres | 0.0133 |
| skin | -0.0006 |
| insu | 0.0012 |
| mass | -0.0897 |
| pedi | -0.9452 |
| age | -0.0149 |
| Intercept | 8.4047 |

(b) 請將該資料集第一筆資料帶入 $ax+c$ 公式，並計算出結果。(需詳細列出計算過程，記錄在 Word 上或手算拍照附圖皆可) (15%)

| No. | 1: preg | 2: plas | 3: pres | 4: skin | 5: insu | 6: mass | 7: pedi | 8: age | 9: class |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| | Numeric | Numeric | Numeric | Numeric | Numeric | Numeric | Numeric | Numeric | Nominal |
| 1 | 6.0 | 148.0 | 72.0 | 35.0 | 0.0 | 33.6 | 0.627 | 50.0 | teste... |
| 2 | 1.0 | 85.0 | 66.0 | 29.0 | 0.0 | 26.6 | 0.351 | 31.0 | teste... |
| 3 | 8.0 | 183.0 | 64.0 | 0.0 | 0.0 | 23.3 | 0.672 | 32.0 | teste... |
| 4 | 1.0 | 89.0 | 66.0 | 23.0 | 94.0 | 28.1 | 0.167 | 21.0 | teste... |
| 5 | 0.0 | 137.0 | 40.0 | 35.0 | 168.0 | 43.1 | 2.288 | 33.0 | teste... |

$$ax+c = (-0.1232*6) + (-0.0352*148) + 0.0133*72 + (-0.0006*35) + 0.0012*0 + (0.0897*33.6) + (-0.9452*0.627) + (-0.0149*50) + 8.4047 = -0.9590604$$

(c) 請計算出 class = tested_negative 的機率。(需詳細列出計算過程，記錄在 Word 上或手算拍照附圖皆可) (15%)

$$P(y = 1 | x) = 1/(1+e^{(-0.9590604)}) = 0.27706635788。$$