

# Machine Learning HW1

學號：R05943005 系級：電子所碩一 姓名：呂丞勛

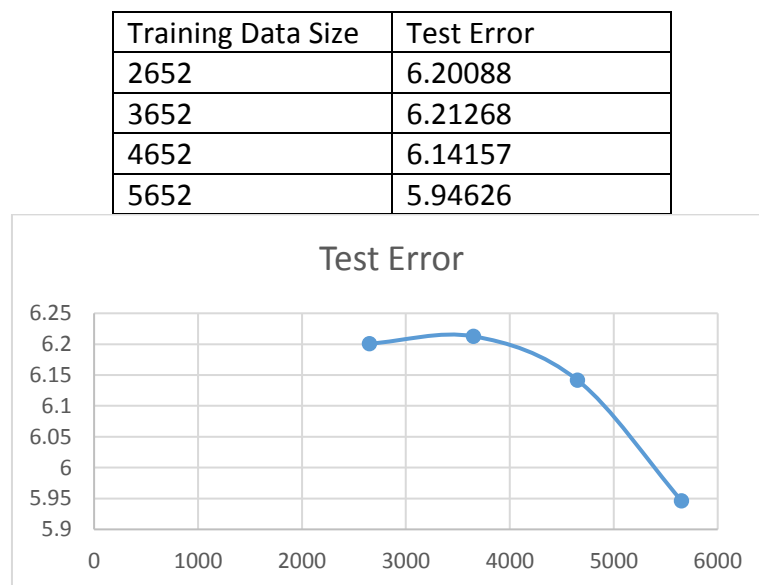
## 1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：

一天的資料中，總共有 18 個觀測值，利用前 9 小時的觀測值來當作一次的 feature 來預測第 10 個小時的 PM2.5，該 feature 有( $9 \times 18 = 162$ )個維度，並且由於一個月有 20 天可以使用，所以一個月會有 480 個小時，共有 471 筆 training data，12 個月便會有 5652 筆來當成 training data。

## 2. 請作圖比較不同訓練資料量對於 PM2.5 預測準確率的影響

答：



原則上，資料越多應該會越好，但是如果搜集過多較差的資料(被雜訊干擾較大)，則太多的資料反而會造成模型的訓練變差，而有較不好的預測結果，由上圖可以看到，越多資料越讓資料預測率降低，表示 training 資料的多寡是很重要的。

## 3. 請比較不同複雜度的模型對於 PM2.5 預測準確率的影響

答：

Feature Size	Test Error
Linear (162)	5.93514
Quadratic (325)	6.24071
Linear (9)	5.75747
Quadratic (18)	5.75745

以直覺來說，越複雜的模型可以較容易的預測出較好的結果，但是實際上複雜的模型反而容易 overfit training data，造成訓練的時候雖然可以達到較低的錯誤率，但在 test 的表現上卻不一定較好，因此選擇適當的模型相當重要。

#### 4. 請討論正規化(regularization)對於 PM2.5 預測準確率的影響

答：

Regularization(labmda)	Training Error	Test Error (kaggle public)
0	5.679518	5.97192
0.0001	5.679519	5.97171
0.001	5.6796	5.97029
0.01	5.683493	5.98084
0.1	5.788625	6.28445

由上表可知，當我們使用 regularization 的時候，通常隨著 $\lambda$ 的值越大，training 的結果會使得誤差越大，然而，在 Test 的情況下則不一定，由於如果沒有任何 regularization，結果可能會使得 model 會 overfitting training data，導致在 test 的情況下變差，但是同樣的，當 $\lambda$ 大超過一定程度的話，則會使得 test error 也隨之上升，因此該參數需要好好調整。

5. 在線性回歸問題中，假設有  $N$  筆訓練資料，每筆訓練資料的特徵 (feature) 為一向量  $x^n$ ，其標註(label)為一存量  $y^n$ ，模型參數為一向量  $w$  (此處忽略偏權值  $b$ )，則線性回歸的損失函數(loss function)為 $\sum_{n=1}^N (y^n - w \cdot x^n)^2$ 。若將所有訓練資料的特徵值以矩陣  $X = [x^1 \ x^2 \ \dots \ x^N]$  表示，所有訓練資料的標註以向量  $y = [y^1 \ y^2 \ \dots \ y^N]^T$  表示，請以  $X$  和  $y$  表示可以最小化損失函數的向量  $w$ 。

答：

依照題目所敘述向量與矩陣，可以將關係列出如下

$$y = Xw$$

若要解出使得題目所定義之 loss function，並使之最小，則必須乘上 $X^T$ 的 pseudo inverse 矩陣，依照 Moore- Penrose pseudo-inverse，我們可以得到該最佳解如下

$$w_{\text{best}} = (X^T X)^{-1} X^T y$$