

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

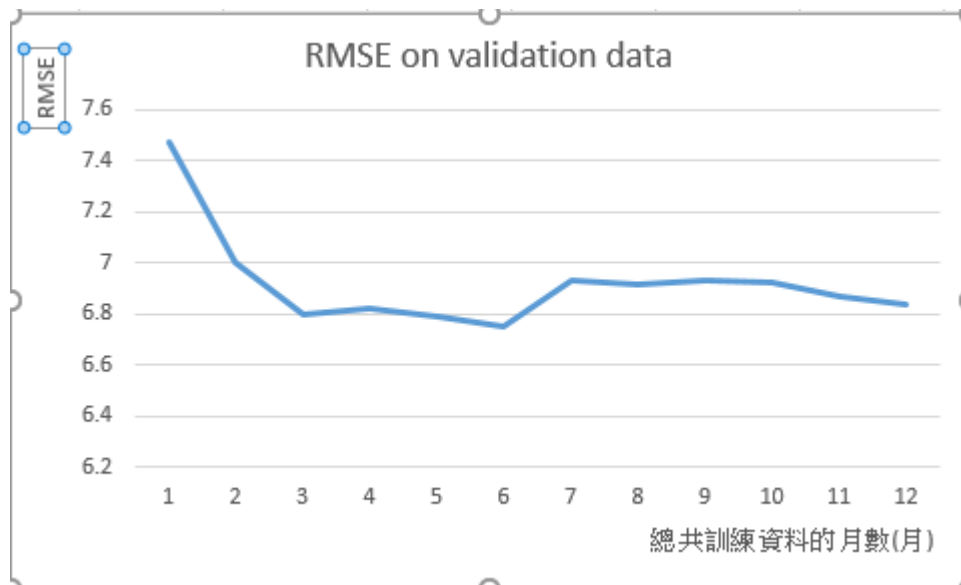
答：

在第一次寫的時候，因為不知道什麼特徵跟 PM2.5 的相關性較大，所以將 18 項各九小時都當作輸入的特徵，之後利用相關係數去計算每項特徵跟 PM2.5 間的相關係數，進而取出 O_3 、PM10、PM2.5、RAINFALL 這幾項跟 PM2.5 較相關的特徵當作輸入特徵。但我後來想到 PM2.5 本身對於自己之後的值影響應該是會最大，於是加入了 PM2.5 的第 6~第 9 小時的二次方當作輸入特徵，而我最後模型所輸入的特徵是 PM2.5 一次方的前九小時和 PM2.5 二次方的第 6~第 9 小時總共 12 個特徵。

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：

因為我取的 feature 都沒有超過 9 個小時，因此我將 'test_X.csv' 這份 test data 的 0~8 小時的數據當作我的 validation data，總共有 240 筆資料，所得到的 Root Mean Square Error。



由上圖可知，訓練資料量與 RMSE 並不是成正比關係，但從大致上的趨勢可以看出訓練資料量越多，model 就會越好，但不是成正比的關係有可能是因為一些月份的數據離每個月的平均值比較遠，因此對 weight 的影響較大，導致效果下降。

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：

這題的 validation data 與第二題相同，接為 "test_X.csv" 的 240 筆。

Models of different features	Loss of Validation data
PM2.5 * 8hours	6.8741044
PM2.5*7hours	6.8623072
(O ₃ +PM10+PM2.5+RAINFALL)*8hours	6.5104017
PM2.5*8hours+(PM2.5) ² *4hours	6.7493204
PM2.5*8hours+(PM2.5) ² *8hours	6.6960231

對於同樣是一次方的模型，取到與 PM2.5 前四相關的特徵的表現是最好，再來是只取 PM2.5 這項特徵，至於其他一次方的模型表現都沒有來的這麼好。

加入二次方的模型，若取 PM2.5 二次方的八個小時，在 validation data 上面表現的也很好，但若取 PM2.5 之外的特徵的二次方模型，表現就會下降許多。

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

答：

正規化主要是避免對其中幾項 feature 的 weight 過大，而容易造成 overfitting，而我總共在一次方和二次方 linear regression 的 model 上時做了正規化，得到的結果是在一次方的 linear regression model 裡正規化的效果不是那麼顯著，我想的原因對於一次方的 model，本來每個 weight 的貢獻就差不多，即使做了正規化效果也不會太明顯。至於在二次方的 model，如果先做了 feature scaling，正規化的效果就會好一點點，而如果沒有先做 feature scaling，正規化就不會有太大的影響。而這是因為對於沒有做 feature scaling 二次方的 feature 而言，weight 理論上是要比一次方的 feature 小，但這與正規化的精神就有點互相違背，所以這是我想沒有做 feature scaling 不會讓正規化效果明顯的主要原因。

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：

$$\begin{aligned}
 \text{Loss function} &= (y - X \cdot w)^T \cdot (y - X \cdot w) \\
 &= (y^T - w^T \cdot X^T) \cdot (y - X \cdot w) \\
 &= (y^T \cdot y - y^T \cdot X \cdot w - w^T \cdot X^T \cdot y + w^T \cdot X^T \cdot X \cdot w) \\
 &= (y^T \cdot y - 2 * y^T \cdot X \cdot w + w^T \cdot X^T \cdot X \cdot w)
 \end{aligned}$$

要最小化 loss function，代表其對 w 的一次微分為 0:

$$\frac{dL}{dw} = -2 * X^T \cdot y + 2 * X^T \cdot X \cdot w = 0$$

$$\rightarrow w = (X^T \cdot X)^{-1} \cdot X^T \cdot y$$