

學號：R05943129 系級： 電子碩一 姓名：侯人文

1. 請簡明扼要地闡述你如何抽取模型的輸入特徵 (feature)

答：取前 9 小時的 pm2.5 指標，取 6 到 9 小時的 SO2，取 6 到 9 小時的 CO，取第 8~9 小時的 WIND_SPEED，取第 8~9 小時的 O3。

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：

0~8 小 PM2.5，4~8 小 SO2，4~8 小 CO， 第 8 小 WIND_SPEED，O3 小 O3	5.661，
0~8 小 PM2.5，0~8 小 SO2，0~8 小 CO， 第 8 小 WIND_SPEED，第 8 小 O3	5.76585
0~8 小 PM2.5，5~8 小 SO2，5~8 小 CO， 7~8 小 WIND_SPEED，7~8 小 O3	5.6252
0~8 小 PM2.5，6~8 小 SO2，6~8 小 CO， 6~8 小 WIND_SPEED，6~8 小 O3	5.6995

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：我有試過多取一些參數來 train，例如我有試過 O3、WIND_DIRECT，但不是越多參數準確率就越準，也有試過多取一些 SO2、CO，可以看上面我所做得表，不是越多參數，準確率越好，而是要適當的參數才會比較好。

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

答：我認為並沒有影響多少。

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：

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m \left(h_{\theta}(x^{(i)}) - y^{(i)} \right)^2.$$

$$J(\theta) = \frac{1}{2m} (X\theta - \vec{y})^T (X\theta - \vec{y})$$

平方 取反矩陣相乘

$$\vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} \quad X = \begin{bmatrix} -(x^{(1)})^T & - \\ -(x^{(2)})^T & - \\ \vdots & \\ -(x^{(m)})^T & - \end{bmatrix}$$

$$J(\theta) = ((X\theta)^T - y^T)(X\theta - y)$$

相乘化簡

$$J(\theta) = (X\theta)^T X\theta - (X\theta)^T y - y^T (X\theta) + y^T y$$

$$J(\theta) = \theta^T X^T X\theta - 2(X\theta)^T y + y^T y$$

$$\frac{\partial J}{\partial \theta} = 2X^T X\theta - 2X^T y = 0$$

微分=0 取最小化損失向量

$$X^T X\theta = X^T y$$

$$\theta = (X^T X)^{-1} X^T y$$

得到 w