

學號：R05921120 系級：電機碩二 姓名：黃浩恩

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (3) 題：

- (1) 抽全部 9 小時內的污染源 feature 的一次項(加 bias)
- (2) 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)

備註：

- a. NR 請皆設為 0，其他的數值不要做任何更動
- b. 所有 advanced 的 gradient descent 技術(如: adam, adagrad 等) 都是可以用的

1. (2%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

(1) 於 Model 1(全污染源)的分數為： $8.60139+6.27294=\underline{14.87433}$

(2) 於 Model 2(僅 pm2.5 污染源)的分數為： $6.89472+6.47034=\underline{13.36506}$

(3) 討論：Model 1 整體誤差較 Model 2 的原因可能為大部分的污染源與 PM2.5 沒有直接的相關性，而 Model 2 的訓練可以針對 PM2.5 做最直接的參數調整，故 Model 2 之結果優於 Model 1 是可接受的。

2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

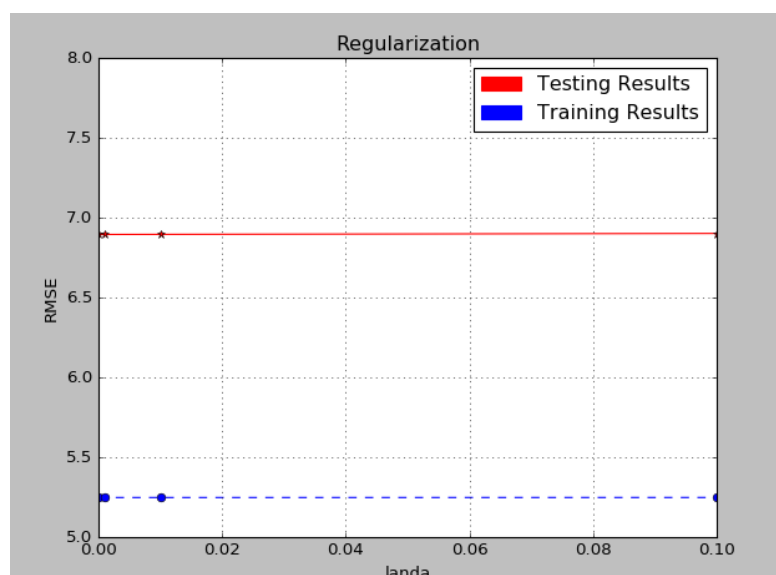
(1) 於 Model 1(全污染源)的分數為： $8.60358+6.52534=\underline{15.12892}$

- i. 討論：於 5 小時與 9 小時之結果進行觀察，可發現 Model 1 因為訓練資料較多，並且可能大部分的污染源與 PM2.5 沒有直接的相關性，故時間的影響權重被稀釋，部分結果變化不大，為合理。

(2) 於 Model 2(僅 pm2.5 污染源)的分數為： $6.85295+6.62478=\underline{13.47773}$

- i. 討論：Model 2 可能因為 PM2.5 受時間過於久的資料影響較小，故 5 小時的預測與訓練可能較 9 小時來的可靠、直觀。

3. (1%)Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖



實際資料(單位為 RMSE)：

Landa(λ)	Training	Testing
0.1	5.24794	6.90109
0.01	5.24769	6.89534
0.001	5.24768	6.89478
0.0001	5.24768	6.89472

其中 Testing 之測試結果來自 kaggle public 的分數。

4. (1%)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 $\sum_{n=1}^N (y^n - x^n \cdot w)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1 \ x^2 \ \dots \ x^N]^T$ 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，請問如何以 X 和 y 表示可以最小化損失函數的向量 w ？請寫下算式並選出正確答案。(其中 $X^T X$ 為 invertible)

- (a) $(X^T X) X^T y$
- (b) $(X^T X)^{-0} X^T y$
- (c) $(X^T X)^{-1} X^T y$
- (d) $(X^T X)^{-2} X^T y$

解：

$$X = \begin{bmatrix} x^1 \\ \vdots \\ x^N \end{bmatrix} \text{ 且 } Y = \begin{bmatrix} y^1 \\ \vdots \\ y^N \end{bmatrix}, \text{ 其中 } X^T y = w X^T X,$$

將 w 右邊的 $X^T X$ 往等號左邊做化簡，為 $w = (X^T X)^{-1} X^T y$ ，故可得知答案為(C)。