

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

兩種 model：learning rate = 10、repeat = 1000000、 $\lambda = 0.001$

- (1) 9 小時 18 features model：

Public test cost = 6.59424、Private test cost = 5.54816、Train cost = 5.90

RMSE = 6.093

- (2) 9 小時 pm25：

Public test cost = 6.90934、Private test cost = 5.50136、Train cost = 6.13

RMSE = 6.245

18 個 feature 的 RMSE 分數較低，因此透過全部 feature 做 model 的方式較佳，可能是因為所有特徵考慮的話，能夠選到那些對 pm25 影響最大的 feature。

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

所有 model：learning rate = 10、repeat = 1000000、 $\lambda = 0.001$

- (1) 9 小時 18 features model：

Public test cost = 6.59424、Private test cost = 5.54816、Train cost = 5.90

RMSE = 6.093

- (2) 5 小時 18 features model：

Public test cost = 6.95377、Private test cost = 5.42063、Train cost = 5.882

RMSE = 6.234

- (3) 9 小時 pm25：

Public test cost = 6.90934、Private test cost = 5.50136、Train cost = 6.13

RMSE = 6.245

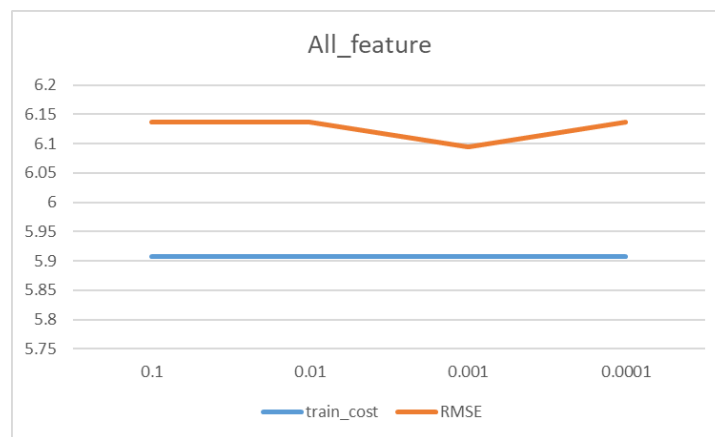
- (4) 5 小時 pm25：

Public test cost = 7.23504、Private test cost = 5.72630、Train cost = 5.19

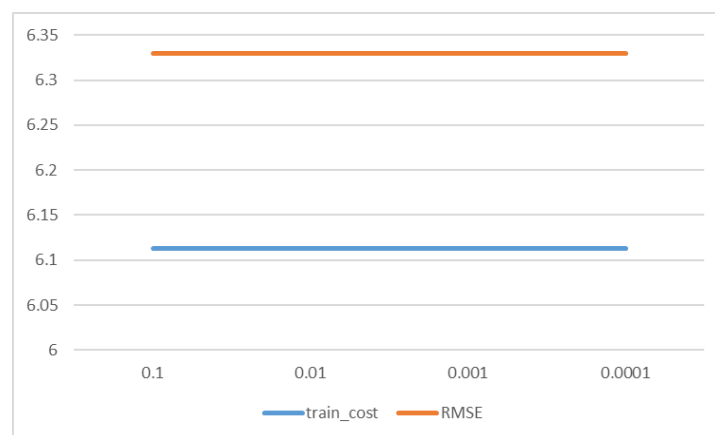
RMSE = 6.524

兩種 model 的 5 小時 RMSE 都較 9 小時的 RMSE 來得高，因此 9 小時的 model 會較佳，可能的原因是因為 9 小時考慮的 feature 較多，有較多的變數去調整及預測最終值。

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



上圖為全部 feature



上圖只考慮 pm25

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 $\sum_{n=1}^N (x^n - w \cdot x^n)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1 \ x^2 \ \dots \ x^N]^T$ 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) $(X^T X)^{-2} X^T y$

$$Ax \sim b$$

For normal equation : $A^T A x = A^T b$

If $A^T A$ can be inverse:

So the Best close form solution : $x = (A^T A)^{-1} A^T b$

Ans : C