

學號：R05942107 系級：電信所碩一 姓名：吳易霖

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

feature	Private Score	Public Score	Private+ Public
所有污染源	5.50413	7.83378	13.33791
PM2.5	5.62719	7.44013	13.06732

可以看出只選擇 PM2.5 當 feature 時的 Public error 比較小，但是 Private 錯誤反而就變高了，但是兩個加總起來看應該還是只選擇 PM2.5 的錯誤少一點，而可能的原因應該是選所有污染源會造成 feature 太多的問題。

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

feature	Private Score	Public Score	Private+ Public
所有污染源	5.37642	7.74074	13.11716
PM2.5	5.79427	7.57651	13.37078

在把時間縮減到只抽前 5 小時後，feature 選擇所有污染的 error 反而變得比只取 PM2.5 好了，而只取 PM2.5 反而變得比之前更差。

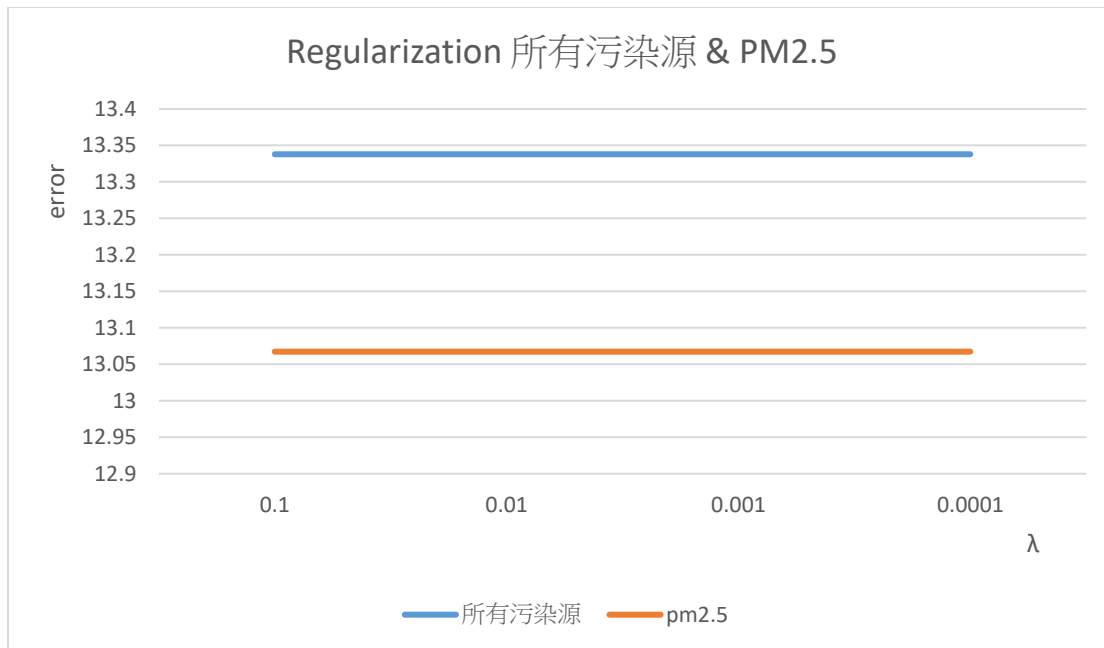
原因大概是總 feature 數量太多或太少了，原本所有汙染源的 9 個時間會有 18X9 個 feature 非常多，只取前 5 個時間就減少到 18X5，因此 error 變少了；而原本只取 PM2.5 的 9 個時間會有 1X9 個 feature 已經夠少，只取前 5 個時間就變成了 1X5 更少了，因此 error 會變大。

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

λ	Private+ Public
0.1	5.50414+ 7.83378=13.33792
0.01	5.50413+ 7.83378=13.33791
0.001	5.50413+ 7.83378=13.33791
0.0001	5.50413+ 7.83378=13.33791

pm2.5：

λ	Private+ Public
0.1	5.62720+7.44012=13.06732
0.01	5.62719+7.44013=13.06732
0.001	5.62719+7.44013=13.06732
0.0001	5.62719+7.44013=13.06732



可以發現改變 λ 完全不會影響 error 的大小，可能的原因是 λ 的數值太小，Regularization 對 cost 造成的影響幾乎可以忽略不計。

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) $(X^T X)^{-2} X^T y$

要取得最小 loss function 之下的的 w ，就是 loss function 對 w 偏微分為 0 時的 w 值

$$\frac{\partial \text{Loss}}{\partial w} = \sum_{n=1}^N 2(y^n - x^n \cdot w) \cdot (-x^n) = -2 \sum_{n=1}^N (y^n - x^n \cdot w) \cdot x^n = 0$$

$$\text{所以 } X^T (y - Xw) = 0$$

$$\Rightarrow X^T y - X^T X w = 0$$

$$\Rightarrow X^T y = X^T X w$$

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

答案為(c)