學號：R05921016 系級： 電機碩二 姓名：傅鈞笙

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

※以下簡稱：

model A --使用9(或5)小時內所有feature 的一次項進行linear regression

model B -- 只使用9(或5)小時內pm2.5 的一次項進行linear regression

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

| | training RMSE | testing public RMSE | testing private RMSE |
|---|---|---|---|
| 9小時內所有feature | 5.701669 | 7.46631 | 5.30105 |
| 9小時內pm2.5 | 6.123022 | 7.44013 | 5.62719 |

Model B在training data set 上的cost 大概會比model A 高出0.4，而在testing data set (public + private) 中，則大概高出0.3。 在training data 與testing data 中，model B 的預測誤差都比較大。

由此可以判斷，model B 應該是一個過於簡單的model。因此，只取9小時內pm2.5 的資料當作input 應該是不夠的，相比之下取全部9小時內的污染源feature的一次項當作feature ，才有足夠的資訊，以進行更準確的預測。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

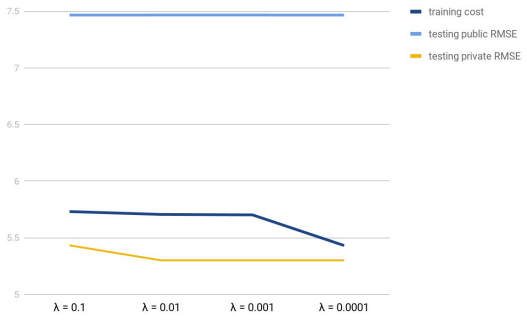
| | training RMSE | testing public RMSE | testing private RMSE |
|---|---|---|---|
| 9小時內所有feature | 5.701669 | 7.46631 | 5.30105 |
| 5小時內所有feature | 5.816742 | 7.66477 | 5.32990 |
| 9小時內pm2.5 | 6.123022 | 7.44013 | 5.62719 |
| 5小時內pm2.5 | 6.207004 | 7.57904 | 5.79187 |

改取前5個小時後，不論是model A 或model B，training 和testing error 都上升了，並且兩者training error 都已經收斂。

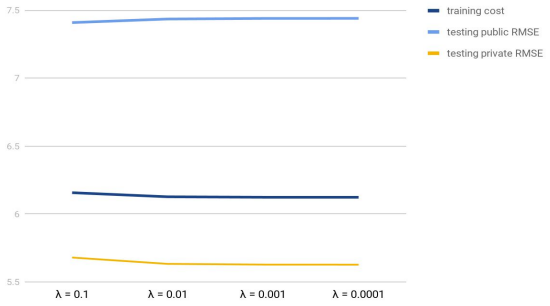
由此可知只取5 小時的資料只會增加bias error，而變成under fitting 的model。

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

Model A:




Model B:




Regularization 的效果並不明顯。

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 $X = [x^1\ x^2 \ldots x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2 \ldots y^N]^T$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$X^TX$為 invertible)

(a) $(X^TX)X^Ty$

(b) $(X^TX)^{-0}X^Ty$

(c) $(X^TX)^{-1}X^Ty$

(d) $(X^TX)^{-2}X^Ty$

若 w1使得 $y = Xw$ 有最小的loss function 值，則 $\frac{d}{dw}L\,|_{w1} = 0$

$$\Rightarrow L = (y - Xw)^T(y - Xw) = y^Ty - y^TXw - w^TX^TY + w^TX^TXw$$

$$= y^Ty - 2y^TXw + w^T(X^TX)w$$

$$\Rightarrow \frac{d}{dw}L = -2y^TX + 2w^T(X^TX) = 0 \quad since\ (X^TX)\ is\ a\ symmetric\ matrix$$

$$\Rightarrow X^Ty = (X^TX)w$$

$$\Rightarrow w = (X^TX)^{-1}X^Ty$$