學號：R04546031　系級：　工工碩二　姓名：洪唯凱

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	9hr + 18 itmes	9hr + PM2.5
Public+Private	7.48250+5.28983	7.44013+5.62719
RMSE	6.479587	6.596241

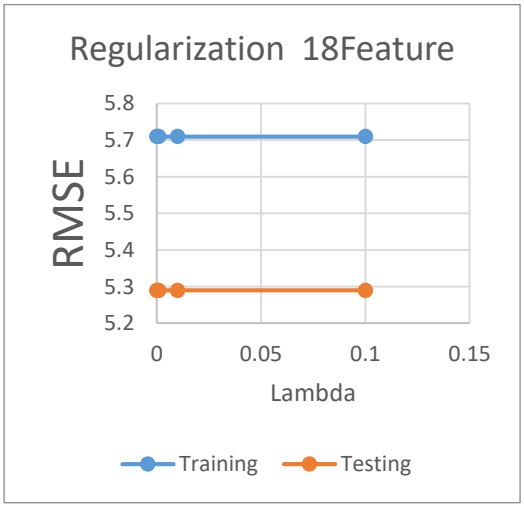
雖然只考慮 PM2.5public 很高，但是到 private 的時候 RMSE 就高了很多，推測仍有些變數還是有用的，整體 RMSE 還是輸 18items

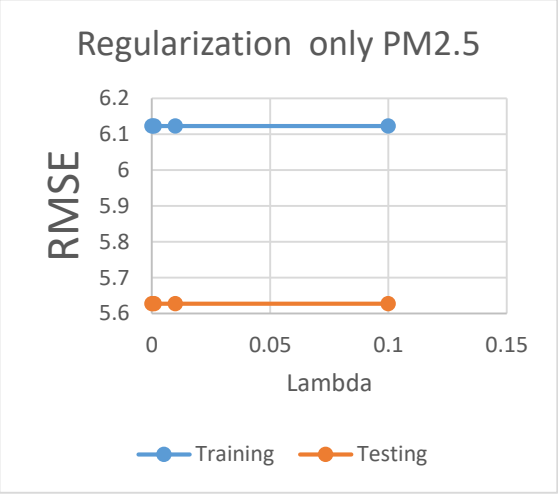
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

Feature	5hr + 18 itmes	5hr + PM2.5
Public+Private	7.66521+5.32875	7.57904+5.79187
RMSE	6.601175	6.744909

跟上述表格比較，Total 都遠輸於 9 小時，表示考量比較多的時間，對於預測 PM2.5 來說比較有解釋力!

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





9hr/testing = private	ALL of 18 Features		PM2.5	
Lambda	Training	Testing	Training	Testing
0.1	5.709472501	5.28983	6.123021522	5.6272
0.01	5.709471489	5.28983	6.123021522	5.62719

0.001	5.709471388	5.28983	6.123021522	5.62719
0.0001	5.709471378	5.28983	6.123021522	5.62719

可能是因為沒有做 Feature Scaling 導致 Regularization 並沒有很明顯

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 n=1Nyn-xnw2 。若將所有訓練資料的特徵值以矩陣 X = [x^1 x^2 … x^N]^T 表示，所有訓練資料的標註以向量 y = [y^1 y^2 … 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)

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

Proof:

$$\sum_{n=1}^{N}(y^n - x^n \cdot w)^2$$

$$= y^{n'}y^n - \boldsymbol{y^{n'}x^nw - w'x^{n'}y^n} + w'x^{n'}x^nw$$

$$= y^{n'}y^n - \boldsymbol{2w'x^{n'}y^n} + w'x^{n'}x^nw$$

$$let \ \frac{\partial(\mathrm{L})}{\partial \mathrm{w}} = -2x^{n'}y^n + 2x^{n'}x^nw = 0$$

$$=> \left(x^{n'}x^n\right)\mathrm{w} = x^{n'}y^n => \mathrm{w} = \left(x^{n'}x^n\right)^{-1}x^{n'}y^n$$