學號：R05229014 系級： 大氣碩二 姓名：鄒適文

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) RMSE ≒ 6.475 (public:7.46631 private:5.30105)

(2) RMSE ≒ 6.596 (public:7.44013 private:5.62719)

可以發現兩次的public分數都差不多，對於public的那份data，大概抽取多少feature跟只抽，PM2.5是差不多的，但對於private的那份data，可以發現只抽取pm2.5的表現下降不少，因此以這份資料而言，或許資料抽取全部data的表現會優於只抽pm2.5的表現。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

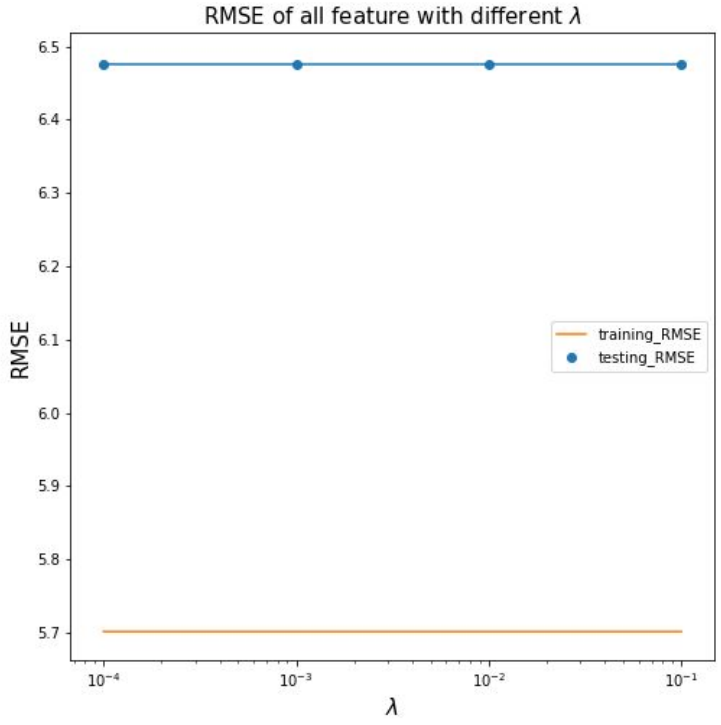
(1) RMSE ≒ 6.601(public:7.66477 private:5.32990)

(2) RMSE ≒ 6.775(public:7.57904 private:5.79187)

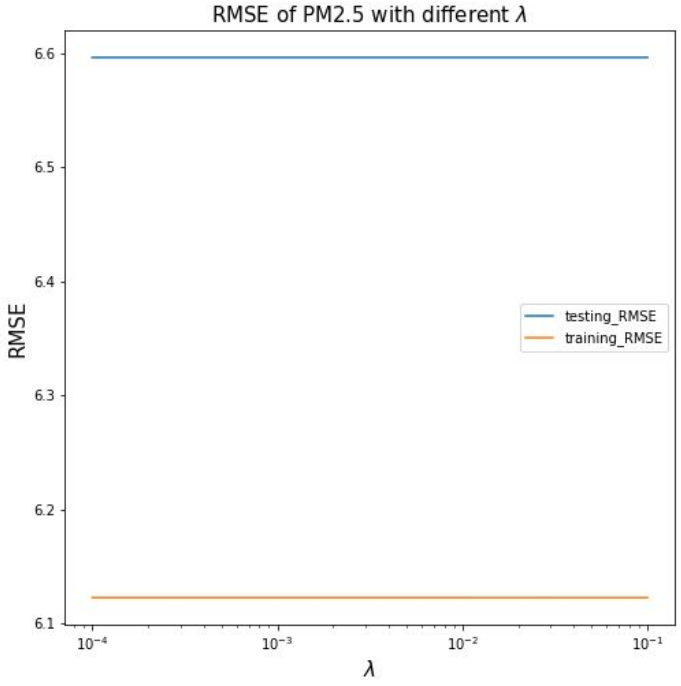
與上面比較，發現不管public，或是private的表現都差了不少，顯然以這份資料而言，預測未來的pm2.5我們抽取9小時前的資料，會比只抽取5小時來的好。

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

(1) 做了Regularization之後發現，用這四個大小的λ，RMSE都沒有差多少，表示這個函數可能對於這組資料，並沒有Overfit的情況發生。



(2) 同上，四個大小的λ幾乎都沒有讓RMSE發生變化，表示這函數對於這組資料可能沒有Overfitting的狀況發生。



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

(C)