學號：B03902016 系級： 資工四 姓名：周聖筌

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

model1: public:7.59727 private: 5.35505 total:12.95232

model2: public:7.44013 private:5.62719 total:13.06732

兩種model雖然在model2在public測資中結果較佳，但出乎我意料之外的，總合來說是取全部污染當feature時的model1好一點。我猜測雖然第一種model考慮的因素太多，有考慮到許多沒有關聯的因素，但第二個model只考慮pm2.5也太少，所以無法有效預測結果，所以兩者皆有很好的預測效果。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

model1: 前9小時: public:7.59727 private:5.35505 total:12.95232

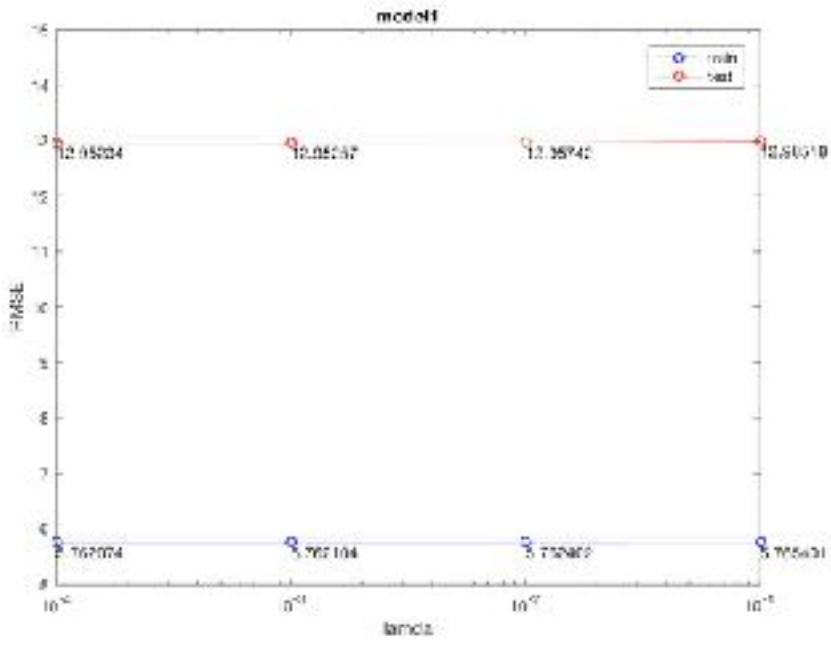
前5小時: public:7.67641 private:5.33896 total:13.01537

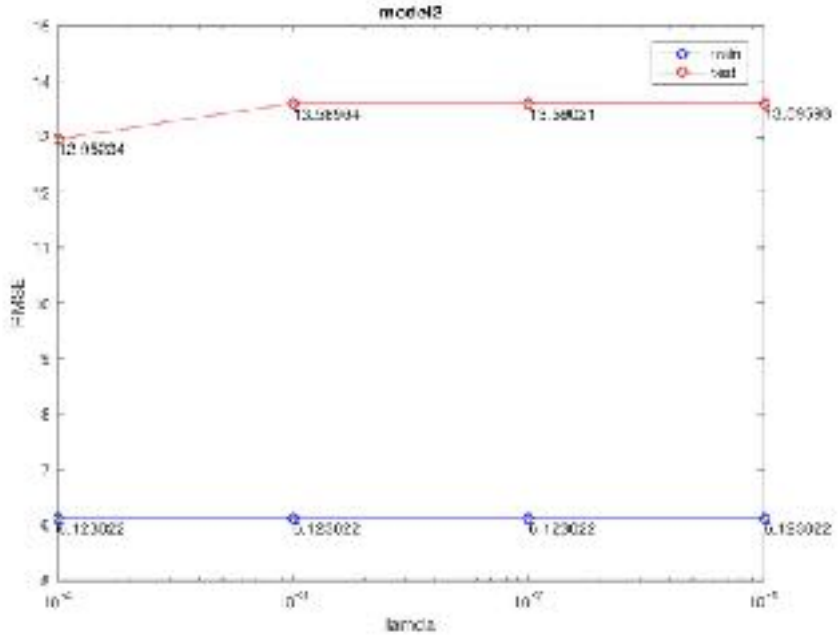
model2: 前9小時: public:7.44013 private:5.62719 total:13.06732

前5小時: public:7.57904 private:5.79187 total:13.37091

兩個model改成抽前五個小時的model後，都得到了比較差的結果。所以知道抽前9個小時是沒有overfitting的情況的，前九個小時的預測情況還是較前五個小時佳一些。

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





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 $X = [x^1\ x^2\ \dots\ x^N]^T$ 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^TX為invertible)

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

$$\sum_{n=1}^{N}\left(y^n - x^n.w\right) = \sum_{n=1}^{N}\left(x^n.w - y^n\right) = \left(X.w - y\right)^T\left(X.w - y\right)$$

其解為此多項式對w微分後等於0時的值。

$$\frac{\partial}{\partial w}\left(w^TX^T - y^T\right)\left(Xw - y\right) = \frac{\partial}{\partial w}\left(w^TX^TXw - w^TX^Ty - y^TXw + y^Ty\right)=0$$

$$X^TXw - X^Ty = 0$$
$$w = \left(X^TX\right)^{-1}X^Ty$$

故得證。