學號:R06725005 系級: 資管碩一 姓名:郝思喬

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

Model	RMSE(Private+Public)
(1)	5.50427 + 7.83403 = 13.3383
(2)	5.62719 + 7.44013 = 13.06732

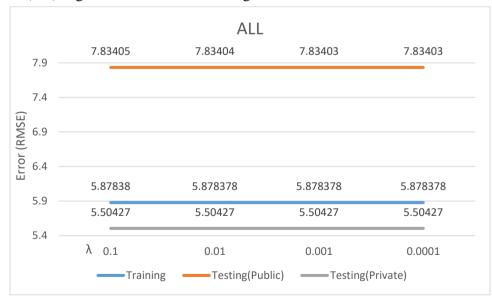
Model(1)的誤差較 Model(2)的大,有可能的原因有某些 feature 在這 Model 中並 非重要的因素,或是某些 feature 的資料是誤差較大的。

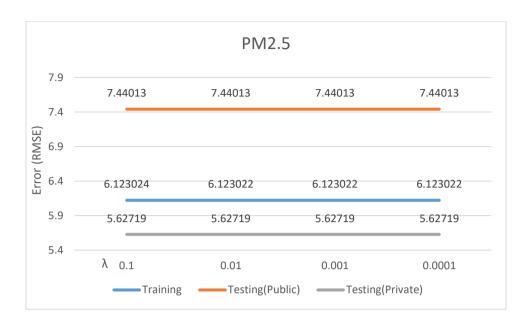
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時,討論其變化

Model	RMSE(Private+Public)
(1)	5.37824+ 7.73769 = 13.11593
(2)	5.79187 + 7.57904 = 13.37091

Model(1)的誤差反而較 Model(2)的小,有可能的原因有某些污染源的資料在連續 5 小時內是較可以預測的,相對於只有用 PM2.5 的可以預測的較準確,也有可能是因為 Testing Data 本身資料誤差的一些關係導致結果相反。

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 \mathbf{x}^n ,其標註(label)為一存量 \mathbf{y}^n ,模型參數為一向量 \mathbf{w} (此處忽略偏權值 \mathbf{b}),則線性回歸的損失函數(loss function)為 $\sum_{n=1}^N (\mathbf{y}^n - \mathbf{x}^n \cdot \mathbf{w})^2$ 。若將所有訓練資料的特徵值以矩陣 $\mathbf{X} = [\mathbf{x}^1 \ \mathbf{x}^2 \ \dots \ \mathbf{x}^N]^T$ 表示,所有訓練資料的標註以向量 $\mathbf{y} = [\mathbf{y}^1 \ \mathbf{y}^2 \ \dots \ \mathbf{y}^N]^T$ 表示,請問如何以 \mathbf{X} 和 \mathbf{y} 表示可以最小化損失函數的向量 \mathbf{w} ?請寫下算式並選出正確答案。(其中 $\mathbf{X}^T\mathbf{X}$ 為 invertible)

- (a) $(X^TX)X^Ty$
- (b) $(X^TX)^{-0}X^Ty$
- (c) $(X^TX)^{-1}X^Ty$
- (d) $(X^TX)^{-2}X^Ty$

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

- $\rightarrow (X^TX)w = X^Ty$
- $\rightarrow \mathbf{w} = (\mathbf{X}^T\mathbf{X})^{\text{-}1}\mathbf{X}^T\mathbf{y}$

Ans: (c)