學號：B05203050 系級： 化學三 姓名：陳品翰

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 當作一次項(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

c. 第 1-3 題請都以題目給訂的兩種 model 來回答

d. 同學可以先把 model 訓練好，kaggle 死線之後便可以無限上傳。

e. 根據助教時間的公式表示，(1) 代表 p = 9x18+1 而(2) 代表 p = 9*1+1

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

(1) 抽全部 9 小時內的污染源 feature 當作一次項(加 bias)：

Public RMSE : 5.66201
Private RMSE : 7.15484
Total RMSE : 12.81685
Average RMSE : 6.40843

(2) 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)：

Public RMSE : 5.83391
Private RMSE : 7.13785
Total RMSE : 12.97176
Average RMSE : 6.48588

抽取較多特徵時，雖然其平均 RMSE 較小，但於不同 testing set 之 RMSE 差較大。可見複雜的 model bias 較小，variance 較大 (overfitting)。

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

(1) 抽全部 9 小時內的污染源 feature 當作一次項(加 bias)：

Public RMSE : 5.66201 → 5.95194
Private RMSE : 7.15484 → 7.16301
Total RMSE : 12.81685 → 13.11495
Average RMSE : 6.40843 → 6.55747

(2) 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)：

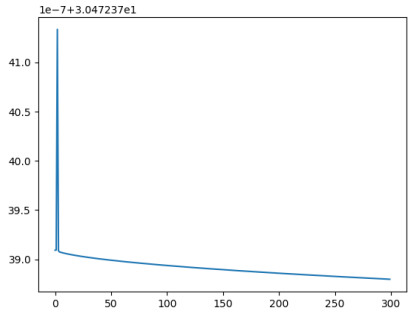
Public RMSE : 5.83391 → 6.18558
Private RMSE : 7.13785 → 7.13906
Total RMSE : 12.97176 → 13.32464
Average RMSE : 6.48588 → 6.66232

改為抽五小時後，model 變簡單，因此皆可見 bias 變大， variance 變小

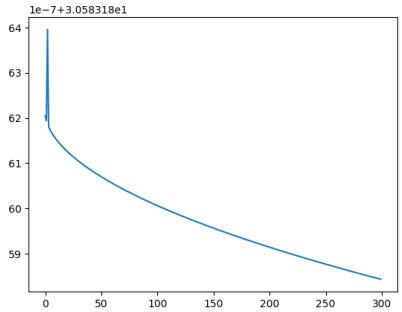
3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖

(1) 抽全部 9 小時內的污染源 feature 當作一次項(加 bias)：

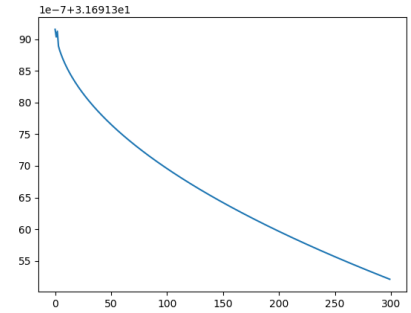
λ=0.0001



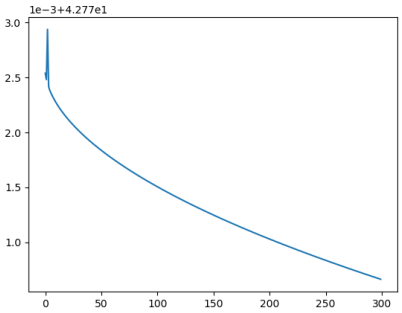
λ=0.001



λ=0.01

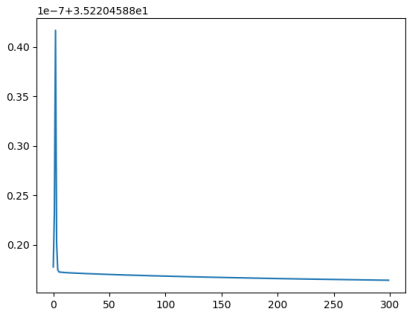


λ=0.1

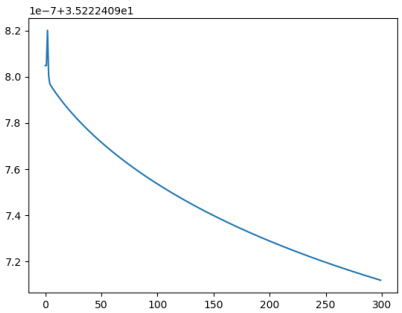


(2) 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)：

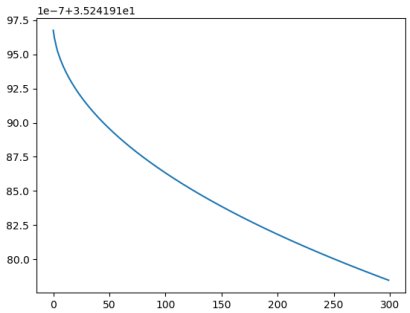
λ=0.0001



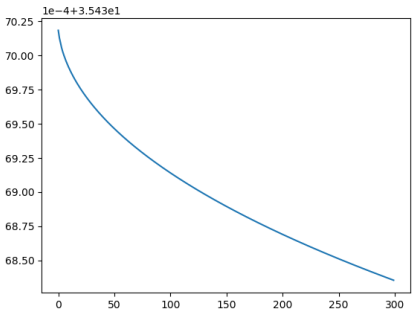
λ=0.001



λ=0.01



λ=0.1



λ 大時 loss 變化較劇烈，因此需用較大之 learning rate 跑。但由於是從最小平方法之解開始 train，因此即使改變 λ， train 完後 kaggle 上的結果仍沒有變化

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)為一純量 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$\sum_{□=1}^{□}$ $(□^{□} - □^{□} \cdot □)^2$ 。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 ？請選出正確答案。(其中 X^TX 為 invertible)

(a) $(X^TX)X^Ty$
(b) $(X^TX)yX^T$
(c) $(X^TX)^{-1}X^Ty$
(d) $(X^TX)^{-1}yX^T$

(c)