

學號：r07942092 系級：電信所碩一 姓名：白佳灝

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  $p = 9 \times 18 + 1$  而(2) 代表  $p = 9 \times 1 + 1$

1. (2%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響  
Iteration 100000 次, learning rate 1

model	Public score	Private score
(2)	5.63779	7.21546
(1)	5.90263	7.22356

模型 2 的表現，不論是在 public score 還是 private score 都比模型 1 來的好一些，因此推測使用多一點的 feature 它的 feature space 比較有可能幫助我們找到最佳的解。而模型 1 可能要增加一些 feature 才有機會讓 training 時找到更好的解，降低 loss。

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

Iteration 100000 次, learning rate 1

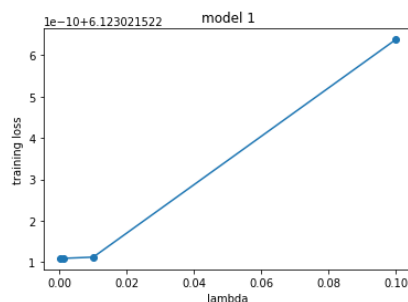
model	Public score	Private score
(2)	5.98345	7.16744
(1)	6.22749	7.22464

觀察 public score 的分數，發現在只使用 5 小時時，兩個模型的 loss 都比使用 9 小時來的高，使用 9hr 時的 feature space 可能比較找到預測 PM2.5 最好的解。

但觀察 private score 時發現卻又發現，模型 2 使用 9hr 的 feature 有可能 overfit training data，因為 9hr 的 private score 比 5hr 的 loss 高出一些。

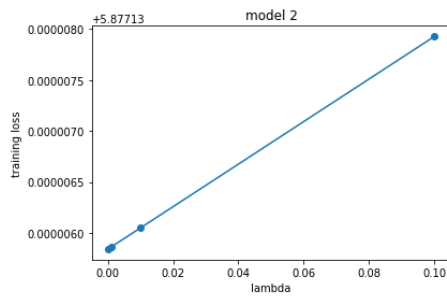
3. (1%)Regularization on all the weight with  $\lambda=0.1$ 、 $0.01$ 、 $0.001$ 、 $0.0001$ ，並作圖

Model (1) Iteration 10000



$\lambda$	training
0.1	6.123021522637468
0.01	6.123021522112521
0.001	6.123021522109559
0.0001	6.123021522109758

## Model (2)



$\lambda$	training
0.1	5.877137928301834
0.01	5.877136047567964
0.001	5.877135859865704
0.0001	5.877135841099204

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)為一純量  $y^n$ ，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  $\sum_{n=1}^N (\mathbf{x}^n - \mathbf{x}^n \cdot \mathbf{w})^2$ 。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  $\mathbf{y} = [y^1 \ y^2 \ \dots \ y^N]^T$  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  $\mathbf{w}$ ？請選出正確答案。(其中  $\mathbf{X}^T \mathbf{X}$  為 invertible)

- (a)  $(\mathbf{X}^T \mathbf{X}) \mathbf{X}^T \mathbf{y}$
- (b)  $(\mathbf{X}^T \mathbf{X}) \mathbf{y} \mathbf{X}^T$
- (c)  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$
- (d)  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{y} \mathbf{X}^T$

Ans: (c)