

ML HW1 R05922105 資工碩二 陳俞安

1.

Feature	Training Loss	Public Score	Private Score
PM2.5 + 1 bias	5.780805551323174	7.47123	5.64609
18 features + 1 bias	5.780805551323174	7.67618	5.39794

在這題中，我統一採用參數為， $ITERATION = 35000$, $LEARNING RATE = 0.5$, $HOURS = 9$ (預測前幾小)，而學習率有使用 Adagrad。在測 Training Loss 時，將 train.csv 切成 9:1 為 training data 與 testing data。

在訓練過程中，如果取比較多 feature 看似效果比較好，但是上傳至 kaggle 反而分數比較低，可以看出 feature 並不是越多越好，有可能造成 overfitting。

我在kaggle經由training得到最好的成績是6.97030，與上不同的是我取了不同的 feature，分別是 CO, NO2, O3, PM10, PM2.5, WD_HR, WIND_DIREC, WIND_SPEED, WS_HR，總共8個feature，並且有對 feature 做 scaling。

2.

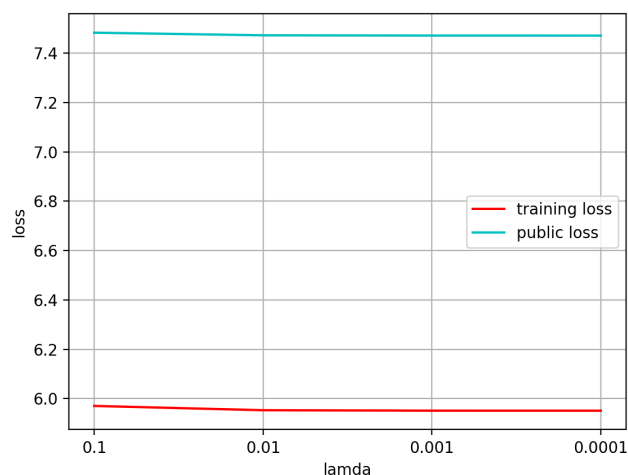
Feature	Training Loss: 5 hour	Public Score: 5 hour	Training Loss: 9 hour	Public Score: 9 hour
PM2.5 + 1 bias	5.9374307865841685	22.56996	5.95137989494136	7.47123
18 features + 1 bias	5.692415983732592	21.84138	5.780805551323174	7.47123

在測training loss時，可以看出似乎用較少的時數去預測PM2.5會得到比較好的結果，但上傳至kaggle後RMSE卻是異常的大，有訓練不足的可能，也可以看出要預測PM2.5是需要事前更多小時來進行判斷。

3.

PM2.5 + 1 bias

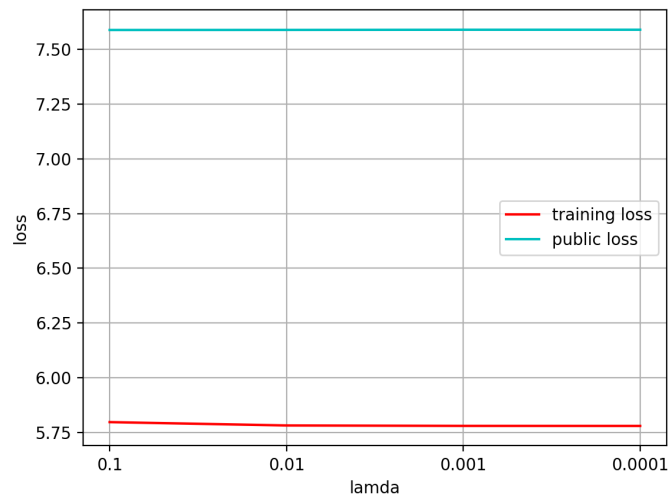
Lamda	0.1	0.01	0.001	0.0001
Training Loss	5.9707130982599175	5.9533296263963	5.951575034449084	5.951399410558054
Public Score	7.48290	7.47240	7.47135	7.47124



加上 regularization 後，public score 有下降的趨勢，regularization 有處理到 overfitting。

18 features + 1 bias

Lamda	0.1	0.01	0.001	0.0001
Training Loss	5.798192191221287	5.782941898148087	5.781027242922321	5.780827810433557
Public Score	7.58756	7.58791	7.58866	7.58875



用了全部feature後，regularization對於public score並沒有太大的幫助，反而 RMSE 還上升，我想是因為18 features 在 training 過程並沒有達到overfitting (training set還太少)， regularization 才會沒什麼功用。

4. (C)

第一步： $y = X \cdot w$ ，要解 w

第二步：在等式左邊同乘上X轉置矩陣， $X^T \cdot y = X^T \cdot X \cdot w$

第三步：在等式左邊在同乘 $\text{inv}(X^T \cdot X)$ ， $\text{inv}(X^T \cdot X) \cdot X^T \cdot y = (\text{inv}(X^T \cdot X)) \cdot (X^T \cdot X) \cdot w$
得到 $\text{inv}(X^T \cdot X) \cdot X^T \cdot y = w$