

## Homework 1 Report – PM2.5 Prediction

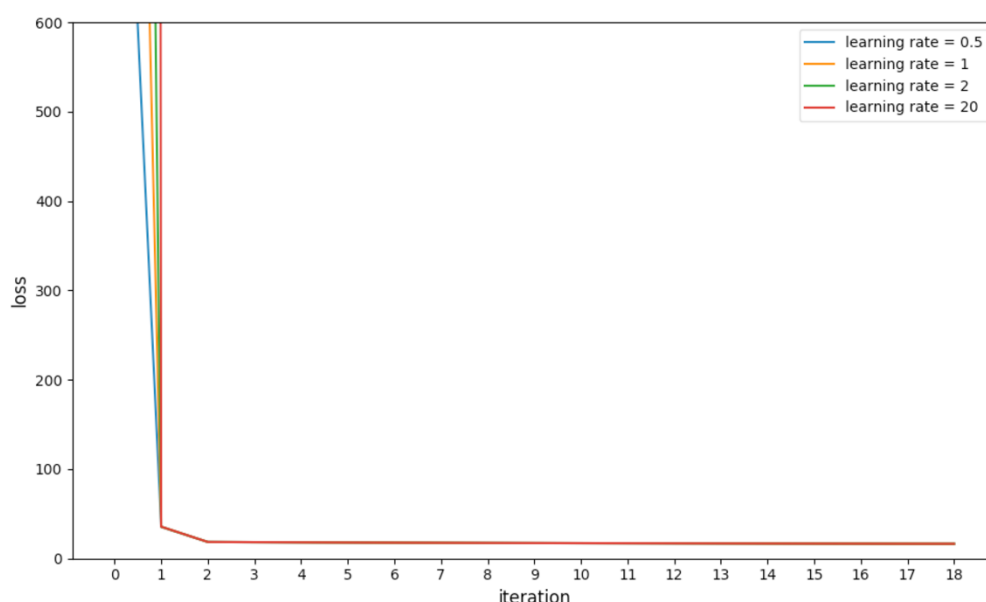
學號: b04901009 系級: 電機三 姓名: 林孟瑾

1. (1%) 請分別使用每筆 data9 小時內所有 feature 的一次項 (含 bias 項) 以及每筆 data9 小時內 PM2.5 的一次項 (含 bias 項) 進行 training, 比較並討論這兩種模型的 root mean-square error (根據 kaggle 上的 public/private score)。

	Private score	Public score
All features	8.72216	8.95816
Only PM2.5	9.69186	9.55813

使用所有 feature 進行 training 所得到的 private/public score 都比只使用 PM2.5 時還要小, 表示預測 PM2.5 時, 其他 feature 也會影響 PM2.5。如果單純只用 PM2.5 來預測, 準確率會較低。

2. (2%) 請分別使用至少四種不同數值的 learning rate 進行 training (其他參數需一致), 作圖並且討論其收斂過程。



隨著 iteration 次數的增加, loss function 逐漸減少。剛開始的 iteration 時, loss function 還很大, 而且減少的幅度較大。iteration 較多次時, loss 的數值較小, 減少的幅度也變小。而不同的 learning rate 所呈現出的 loss 變化, 隨著 learning rate 的增加, loss function 的減少幅度變大。因為 learning rate 增加時, “走向 loss function minimum 的 step” 變大, loss 就較快速的變小。

3. (1%) 請分別使用至少四種不同數值的 regularization parameter  $\lambda$  進行 training (其他參數需一致), 討論其 root mean-square error (根據 kaggle 上的 public/private score)。

	Private score	Public score
$\lambda = 10^5$	7.07758	7.51911
$\lambda = 10^6$	7.07494	7.51836
$\lambda = 10^7$	7.05087	7.47809
$\lambda = 10^8$	6.93972	7.31341
$\lambda = 5 \times 10^8$	7.09162	7.64812
$\lambda = 10^9$	7.46494	8.14540

隨著 $\lambda$ 值的增加，private score 和 public score 逐漸減少，但到了 $\lambda = 10^8$ 之後，score 不減反增，應證了上課所說的，regularization 時， $\lambda$  增加到一定的程度會使 error 為最小值，但大於此  $\lambda$  值，error 反而會增加。

4. (1%) 請問這次作業你的 best\_hw1.sh 是如何實作的？（e.g. 有無對 Data 做任何 Preprocessing? Features 的選用有無任何考量？訓練相關參數的選用有無任何依據？）

整體的架構是 gradient decent。

(1) feature 的選取中，經由觀察後發現，wind direction 的部分，由於是方向，與 PM2.5 的濃度較不會有 linear 的關係，所以我將 "WIND\_DIREC" 和 "WD\_HR" 去除，不列入 training data 裡面。而 "RAINFALL" 的數值幾乎都是相同的，猜測對於 PM2.5 影響不大，故也加以去除。挑去這幾個 feature 之後，訓練的結果通過 strong baseline，表示 feature 的選取對 model 的準確度有極大的關係。

(2) training 時，weight vector 的各個數值一開始是 random initialize，再將 PM2.5 前一小時與前兩小時的位置的 weight 增加為原本的 500 倍跟 10 倍，希望讓此兩個數值的 weight 較大。（因為前一小時的數值和前兩小時的 PM2.5 的數值對預測 PM2.5 值的影響頗大）