

Homework 1 Report - PM2.5 Prediction

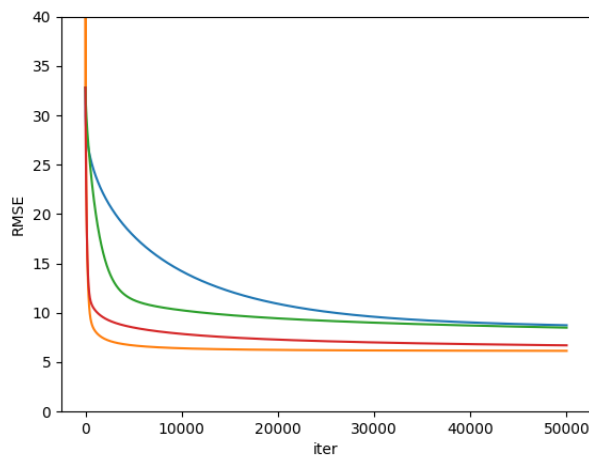
學號：b05902043 系級：資工二 姓名：劉鴻慶

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）。

	Only pm2.5	All feature
kaggle (public/private score)	10.42993 / 9.81910	8.80590 / 8.31444

由 kaggle 的 public/private score 可知 所有 feature 拿去做 gradient descent 的 model 較只有 pm2.5 的做 gradient descent 的 model 效果好（root mean-square error 較小）推測是因為當選用越多 feature，function set 就越大，我們就越可能從較大的 function set 中挑選出合適的 function。

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。



L_rate	0.1 (藍線)	1 (綠線)	10 (紅線)	100 (黃線)
RMSE after 50000 iteration	8.76	8.51	6.71	6.15

由圖片可得知使用越大的 learning rate 來 training 其相對的 RMSE 收斂的速度較快，50000 步後的 RMSE 也越小（越貼近 minimum）。比較特別的是，初始的 RMSE 約為 32.8，但 learning rate = 100(黃線) 在開始的前幾步 RMSE 卻會增

加到 100 左右才開始下降，若將 learning rate 條更大，前期增加 RMSE 的情況將會更明顯，所以 learning rate 也不是一味的越大越好，需要找到一個合適的 learning rate。

3. (1%) 請分別使用至少四種不同數值的 regularization parameter λ 進行 training（其他參數需一至），討論其 root mean-square error（根據 kaggle 上的 public/private score）。

λ	10	1	0.1	0
kaggle Public/private	10.03845/9.40919	7.42469/7.37681	6.43579/6.59165	6.30485/6.50341

由圖可知越大的 λ 對 kaggle 分數並沒有幫助，反而 $\lambda = 0$ 的表現最好。由於這次的 model 採用的是較簡單的線性模型，所以比較不會產生 overfitting 的問題，若是增加 λ ，反而限制了某些 weight 的選用，造成 RMSE 會上升。

4. (1%) 請這次作業你的 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？）

Preprocessing:

- 1.在開始 train 時有將資料切成 50 筆一段然後將每段的 RMSE 印出來發現 1200~1250 和 1400~1450 者兩個區段的誤差異常的大所以將這兩段切掉。
2. PM2.5 和 PM10 有異常的負值或等於零，將異常的值等於他的前一小時和後一小時的平均，如果沒有前一小時或後一小時直接將他等於有的哪一邊

Feature 是用全部 162 feature 的一次項加 bias 做 linear regression 的 close form solution

最後 best_hw1.sh 的 kaggle 分數：6.36024 / 6.21744 (public / private)