

Homework 1 Report – PM2.5 Prediction

學號：b04501127 系級：土木三 姓名：凌于凱

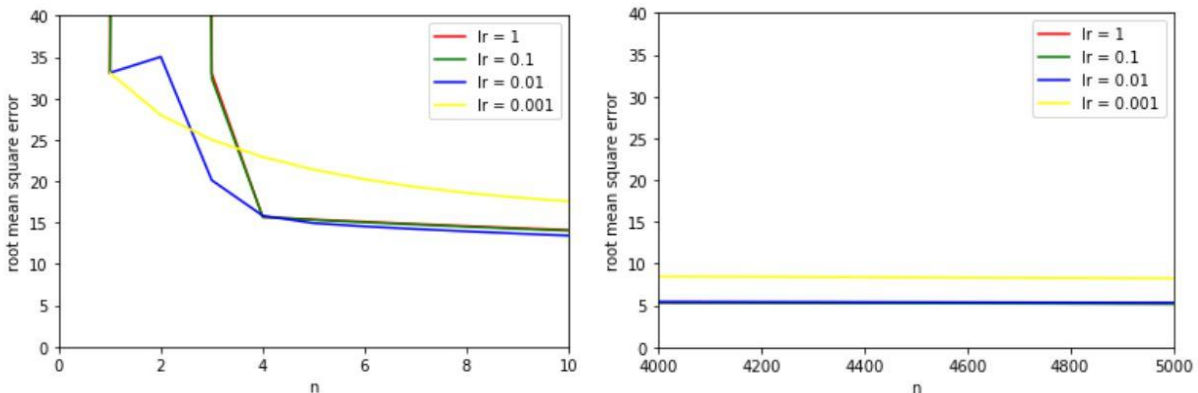
1. (1%) 請分別使用每筆 data9 小時內所有 feature 的一次項（含 bias 項）以及每筆 data9 小時內 PM2.5 的一次項（含 bias 項）進行 training，比較並討論這兩種模型的 root mean-square error（根據 kaggle 上的 public/private score）。

Learning rate = 1.0, iterator = 200000

	Training score	Public score	private score
163 features(all features * 9 + bias)	4.508244	5.65948	6.38972
10 feature(PM2.5 * 9 + bias)	5.116672	6.22288	7.01288

因為沒有 overfitting 的問題，所以所有 feature 的一次項比起只有 PM2.5 的成績來的好。

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。



lr = 1 - lr = 0.1 兩個結果幾乎一致，且在一開始時因為 lr 過大，使得 RMSE 跳到大概 3000 多，但因為是使用 adagrad，所以到最後移動的大小與時間成反比。

lr = 0.01 一開始也因為 lr 較大，而使得誤差上升。

lr = 0.001 收斂較其他三者穩定，但也較慢，並沒有因為 lr 過大使得誤差跳到很大，但到後期收斂速度非常慢，並未與其他相同。

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)。

163 features(all features * 9 + bias), learn rate = 1.0, iterator = 500000

	$\lambda = 0$	$\lambda = 1$	$\lambda = 10$	$\lambda = 100$	$\lambda = 1000$	$\lambda = 10000$
training score	4.50058	4.50072	4.50211	4.50995	4.53139	4.69656
public score	5.66537	5.66037	5.63145	5.59870	5.64885	5.96951
private score	6.40610	6.40622	6.40767	6.41327	6.40444	6.45760

172 features(all features * 9 + PM2.5**2 + bias), learn rate = 1.0, iterator = 500000

	$\lambda = 0$	$\lambda = 1$	$\lambda = 10$	$\lambda = 100$	$\lambda = 1000$	$\lambda = 10000$
training score	4.48833	4.48840	4.48908	4.49463	4.51819	4.64679
public score	5.58177	5.58132	5.57777	5.56621	5.58308	5.66737
private score	6.44517	6.44523	6.44583	6.45137	6.46880	6.60839

從表格可以看出 regularization 對於這兩組的成績幫助並沒有那麼大，且在 private score 上，反而比沒有 regularization 的還要大，我認為可能是因為 feature 最高次方項並沒有大到會發生 overfitting 的程度。

4. (1%) 請這次作業你的 best_hwl.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？)

一樣是使用 linear regression 的 adagrad 實作，但對於 dataset 有做預處理，我是假設所有 PM2.5 為零時是資料遺失，還有一些異常的點，像是數值過大，或是與前一小時的數值相差過大，將發生三種情況的 data 刪除，而在 testing data 上，我將發生這三種情況的直以平均值取代，在 feature 選用上我是以所有 feature 的一次項 + PM2.5 的平方項 + bias；還有做 regularization，將 λ 值取 100，但最後發現有加 regularization 的 public score 雖然比沒有的高，但最後的 private 成績卻比較低，且比起沒有加上 PM2.5 的平方項的 model 的 private score 低。