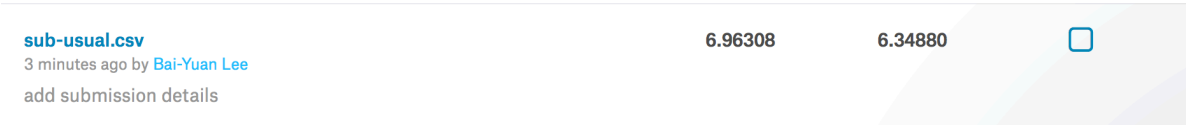
Homework 1 Report - PM2.5 Prediction

學號：b05902002 系級：資工二 姓名：李栢淵

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）。

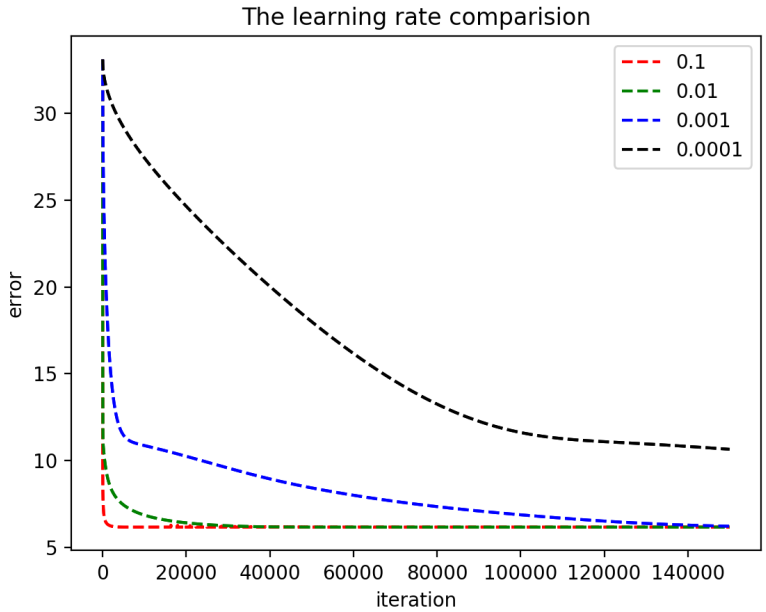
所有feature的一次項（含bias項）

PM2.5的一次項（含bias項）



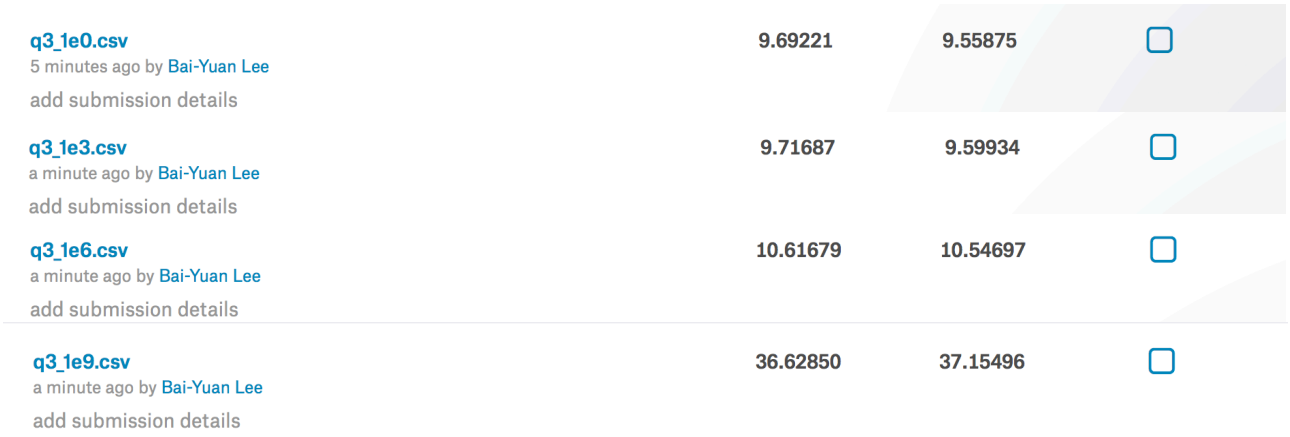
我用的是有清過PM2.5的training data，w從0向量開始train，發現只含一次項的結果，比全部feature餵下去好滿多的，可能是因為我有清理PM2.5，其他的feature沒有清理，所以讓全部feature的一次項的error高很多。

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。



我用的是有清過PM2.5的training data，w從0向量開始train，發現0.1跟0.01收斂的很快，0.001看起來比較適中，0.0001就收斂的太慢了。所以learning rate 越大，收斂越快。

3. (1%) 請分別使用至少四種不同數值的 regulization 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）。



由上到下的 parameter 分別是 1、10^3、10^6、10^9

我用的是「沒有」清過PM2.5的training data，w從0向量開始train，當parameter越大，public和private的error都會越來越大，我認為這滿合理的，畢竟當parameter到達太大，就等同於過度regulization，讓 $|w|$ 太小了，error 就會變大。

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？）

我有對 train data 跟 test data 裡面的PM2.5資料進行處理，而且只有拿PM2.5出來train，用了PM2.5的一次方到五次方（含bias項）。在train data裡，只要PM2.5的值 ≥ 130 或是 ≤ 0，我就把那一筆資料拔掉，而因此濾掉了500多筆資料。在 test data 裡，對於PM2.5的值 ≥ 130 或是 ≤ 0進行修改，因為有連續九項PM2.5，如果是第一項出問題，就用第二項，最後一項出問題就用倒數第二項，其他前後加起來除以2。

此外我先用解析解尋找 w_0，再進行gradient descent 和 regulization，Learning rate 是 0.0001，epoch次數為1000000，regulization parameter 為 1，會需要epoch那麼多次可能是因為一開始解析解找到的 $|w_0|$ 太大了。