

Homework 1 Report - PM2.5 Prediction

學號：b05611033 系級：生機二 姓名：杜杰翰

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）。

使用 adagrad 並做 $\text{Lambda} = 1$ 的 L2 正規化，iteration = 50000

使用 9 小時內所有 feature 的一次項在 kaggle 上所得到的 rmse 是

public: 20.21031

private: 11.91653

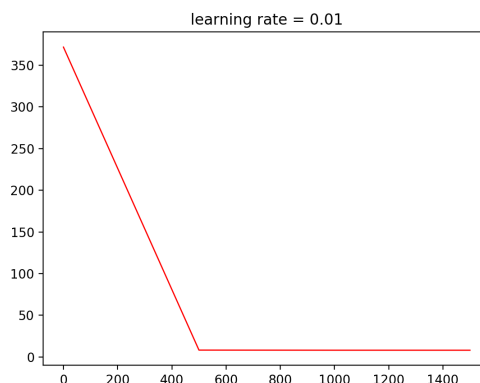
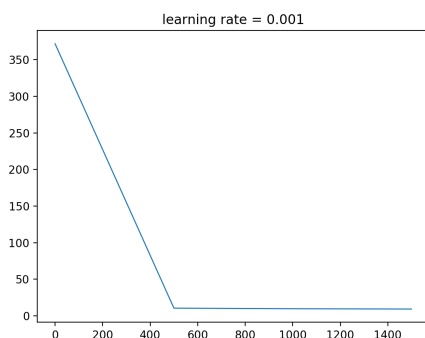
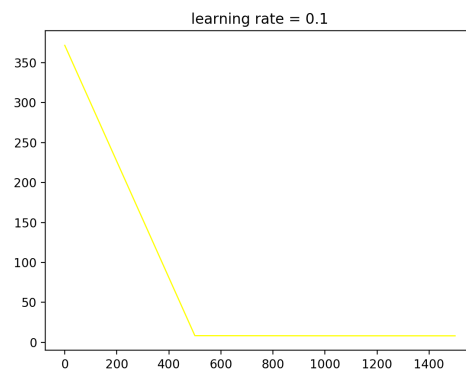
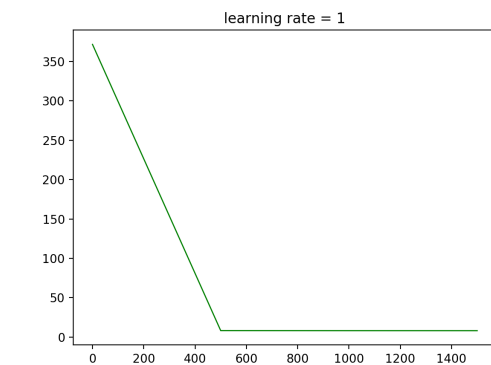
而使用 9 小時內 PM2.5 的一次項得到的 rmse 是

public: 11.69226

private: 10.32100

可能是因為有很多 feature 與 PM2.5 並無關聯，所以在 training 上無法會產生很大的 error。

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。



從圖中可以看到，因為使用 adagrad，所以初始 learning rate 為多少其實並不重要，最後都會往好的方向收線，且收斂速度都差不多，若將 4 張圖放在一起將嚴重重疊看不出有 4 條線，因此把他分成 4 張圖。

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）。

Lambda = 1

Private: 10.37753

Public: 11.79350

Lambda = 10

Private: 10.37971

Public: 11.79536

Lambda = 100

Private: 10.40048

Public: 11.81193

Lambda = 1000

Private: 10.53165

Public: 11.84876

這時使用

當 lambda 很小時，對 error 的變化並不大，

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？）

我使用前兩個小時來預測第三個小時的模式。在 feature 中我使用以下東西：

bias

PM2.5 一次項

PM2.5 二次項

PM10 一次項

PM10 二次項

SO2 一次項

並將 train.csv 中幾筆不合常理大的 PM2.5 縮小，一些過大的 SO2 縮小，將以上 feature 中小於 0 的資料轉正，最後並將各項 feature 百分比化，進行 training 後再將結果的 PM2.5 等比例還原。

在 training 的過程中，我使用 adagrad, learning rate 為 1，並做了 L2 正規化， $\text{Lambda} = 100000$ ，最後作出了 7.多的 error。