

Homework 1 Report - PM2.5 Prediction

學號：b05901041 系級：電機二 姓名：蘇家軒

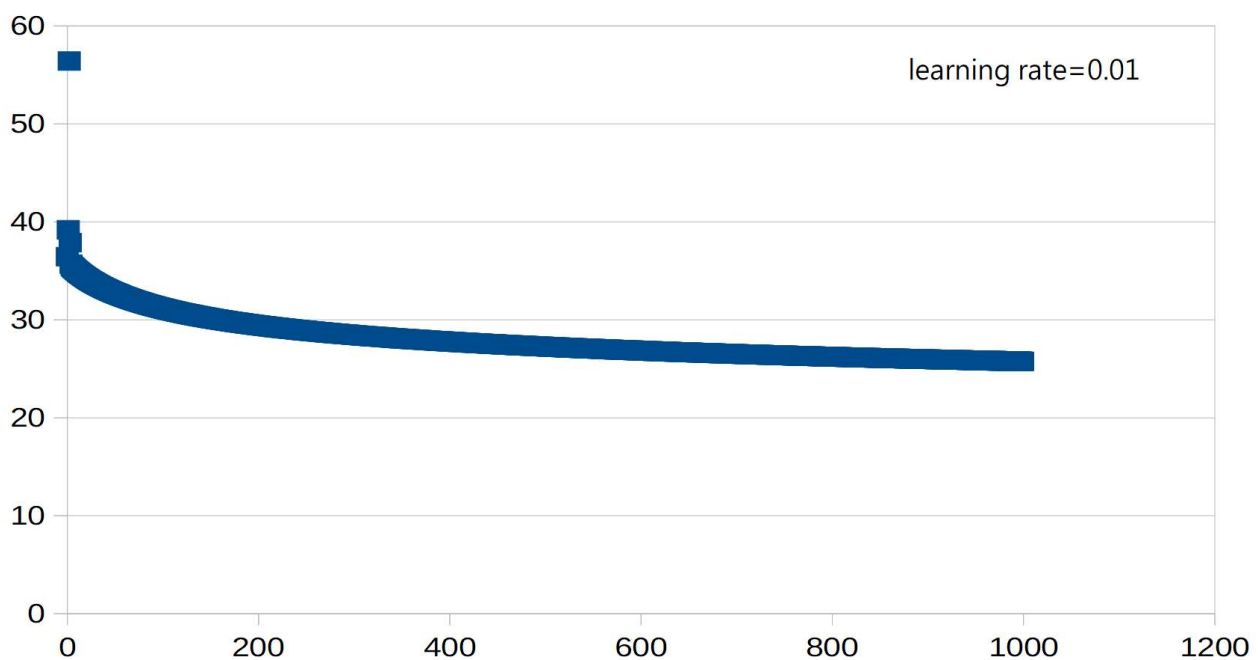
1. (1%) 請分別使用每筆 data9 小時內所有 feature 的一次項（含 bias 項）以及每筆 data9 小時內 PM2.5 的一次項（含 bias 項）進行 training，比較並討論這兩種模型的 root mean-square error（根據 kaggle 上的 public/private score）。

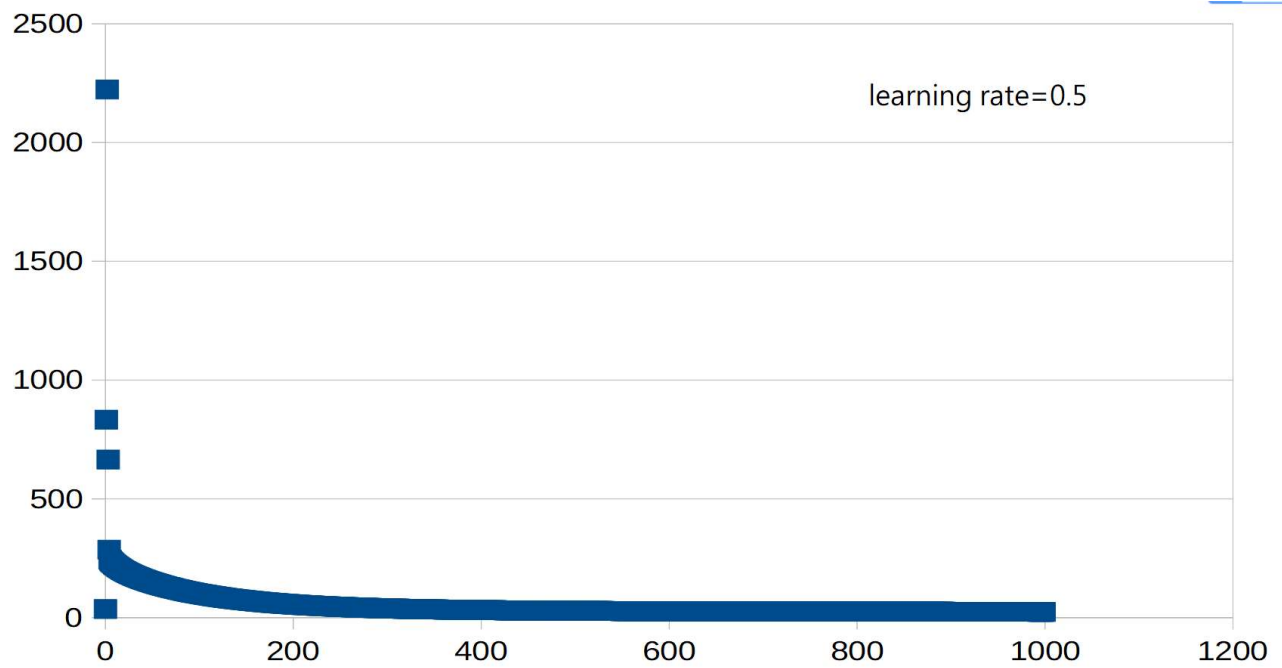
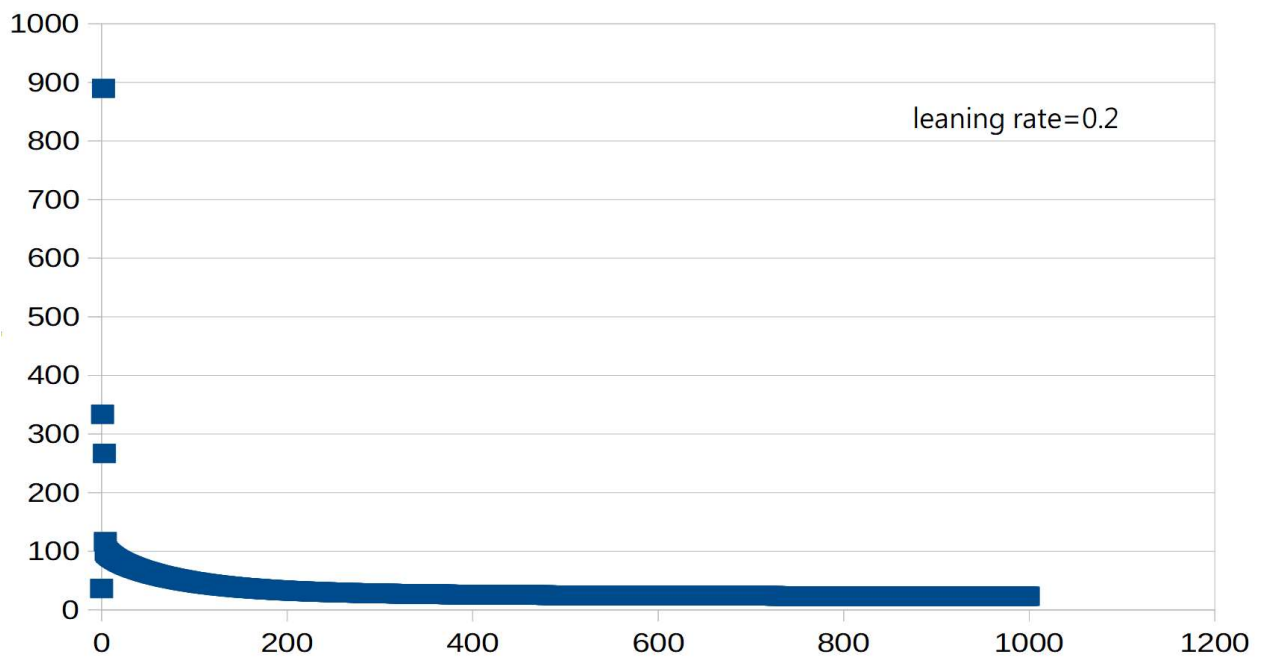
所有 feature 之 public score: 9.10443, private score: 8.91606

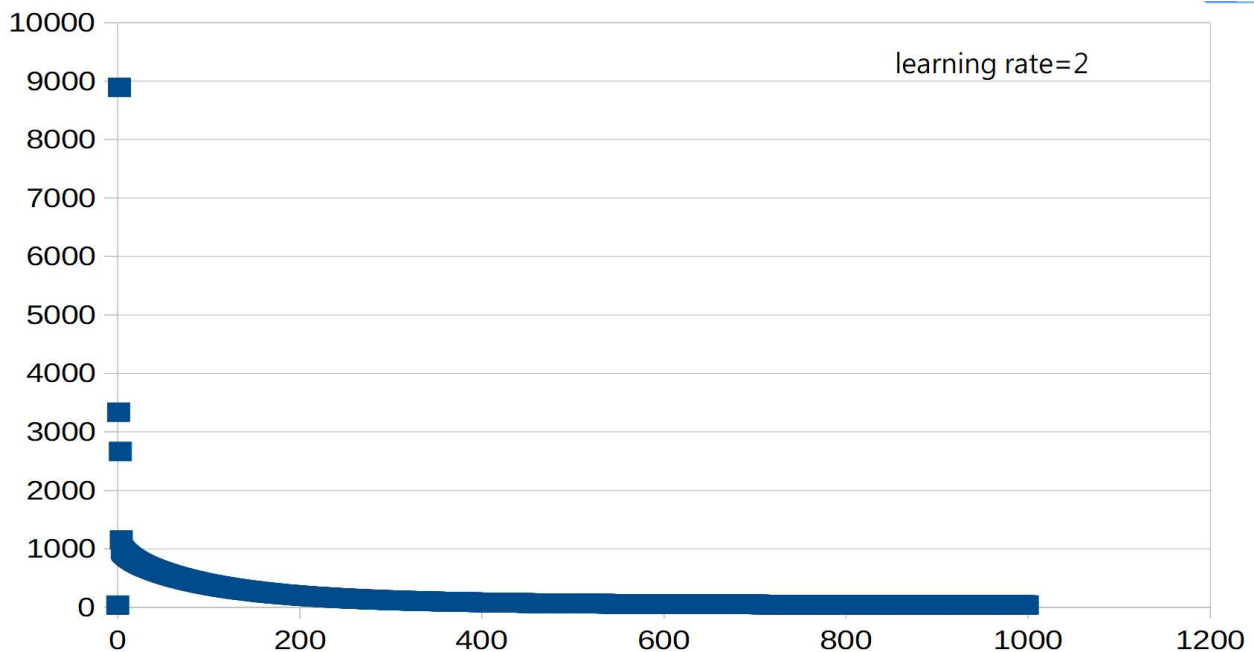
只取 PM2.5 之 public score: 8.24987, private score: 8.43819

理論上應取與下一小時之 PM2.5 相關度較高之 feature 去做 linear regression，而先前之 PM2.5 較其他 feature 相關度高是合理的。

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。







(最後的 cost 分別為 25.75, 23.30, 24.32, 38.38)

由以上的圖表可以看出，當 learning rate 愈大時，在初期所出現的 noise 可能愈大。而在後期因為使用 adagrad，收斂的趨勢看起來差不多。在計算次數相同時，決定最後結果好壞的因素有兩個：一開始開始穩定收斂的起點和 learning rate 本身。愈高的 learning rate 會造成從誤差較大的點開始收斂，但在使用 adagrad 的情況下，後期收斂的過程快速且良好。較小的 learning rate 本身起點誤差小，但收斂速度較慢。在重複 1000 次計算的情況下，learning rate 取 0.2 取得較佳的結果。

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）。

regularization parameter=0, public score:9.10443, private score:8.91606

regularization parameter=1, public score:9.23441, private score:9.15963

regularization parameter=10, public score:9.90547, private score:10.76583

regularization parameter=100, public score:75239.2113, private score:74223.1199

當 $\lambda < 10$ 時都沒有太大的浮動(亦沒有進步)，取到 100 時就爆炸了。

個人認為 regularization 應該用於當某參數浮動便會造成巨大影響時使用，由於我使用線性方程模型，較無此需求。

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？）

我只取 PM2.5 做 training data，先做 data cleaning(把 >200 的數值去掉)，再直接做 linear regression。若 test data 中出現 >200 的 PM2.5 值，以後面的值去取代掉他，再下去套 linear

regression 的式子去求得結果。