

Homework 1 Report - PM2.5 Prediction

學號: b05902052 系級: 資工二 姓名: 劉家維

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)。

比較結果:

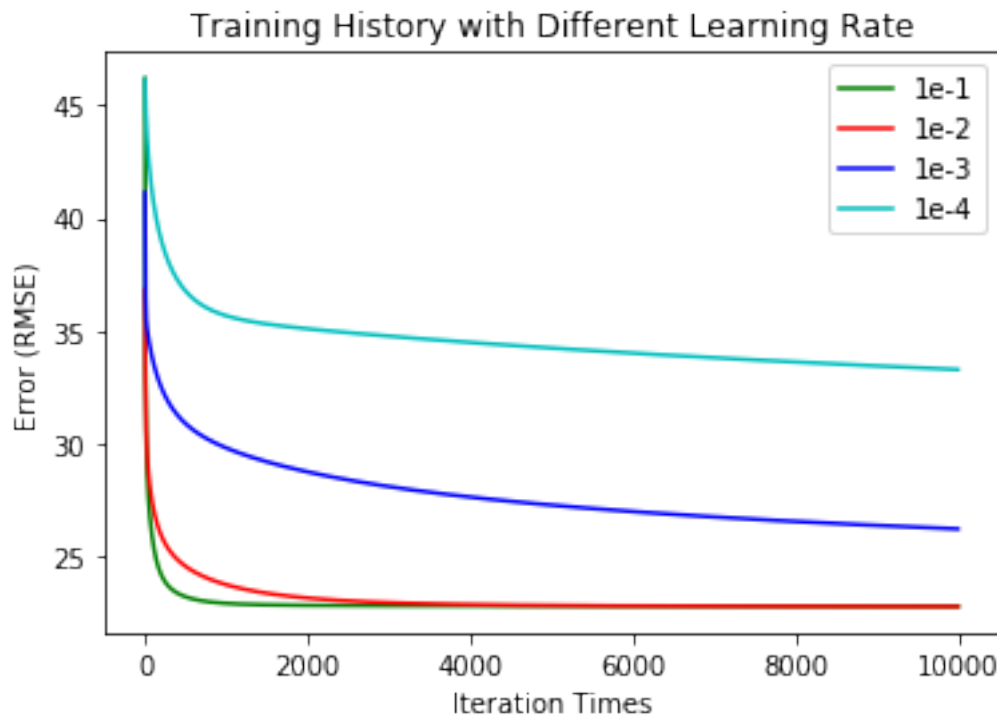
所有 feature 的一次項 (含 bias 項): 8.99587 / 8.71551

PM2.5 的一次項 (含 bias 項): 9.55812 / 9.69186

原本以為只看 PM2.5 的 feature 效果會較好一些, 因為我的 best model 是只看 PM2.5 和 PM10 做預測, 有刪掉其他一些較無用的 feature。但是結果看起來只看 PM2.5 效果並沒有比較好, 推測可能原因如下:

- (1) 我是用 gradient descent 比較這兩者的性能, 也許適合兩者的超參數並不相同。我沒有為兩者調整最好的超參數, 也許這是造成兩者差異的原因。
- (2) 砍到只剩下 PM2.5 的 feature 之後造成資料量太少, 無法充分表達資料的性質。雖然適度的刪減資料的維度可以避免維度災難, 但是砍到只剩下 PM 2.5 這項 feature 可能刪減的太多了。

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。



如想像中的相同，若 **learning rate** 太小，會造成收斂速度較慢。但是值得一提的是，若將 **learning rate** 調的大一點，發現學習曲線會重疊在綠色線上（跟 **learning rate** = 0.1 時相差不多），這點還不太確定原因。

3. (1%) 請分別使用至少四種不同數值的 **regularization parameter** λ 進行 **training**（其他參數需一致），討論其 **root mean-square error**（根據 **kaggle** 上的 **public/private score**）。

$\lambda = 0$: 8.99587 / 8.71551

$\lambda = 1e2$: 8.98922 / 8.71591

$\lambda = 1e4$: 8.87833 / 8.69151

$\lambda = 1e5$: 8.86511 / 8.68410

$\lambda = 1e6$: 9.44021 / 9.06345

$\lambda = 1e8$: 16.20646 / 16.88785

若 λ 太低，會造成模型過於強調特定 **feature** 在訓練集的特性，若以較不具有普遍性、資料集太小、或者是有噪聲的訓練集訓練會喪失泛化能力。

若 λ 太高，會造成模型沒有能力對已知資料集做很好的擬合，自然無法對問題提供良好的預測。

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？）

(1) 有發現某幾筆的 **PM 2.5** 異常的飆高或甚至是負值，因此把訓練集的異常值都刪除；若測試集發現異常值，則用前一天或後一天的值代替。

(2) 僅選用 **PM2.5** 和 **PM10** 這兩項 **feature** 進行訓練及測試。預設從所有 **feature** 中選取兩個 **feature** 進行訓練，在暴力搜尋所有 **feature** 組合並進行 **validation test** 之後，發現 **PM2.5** 和 **PM10** 的效果最好。

(3) 因為後來是直接使用 **scikit-learn** 的線性回歸工具，就沒有再調整 **learning rate** 之類的超參數了。