

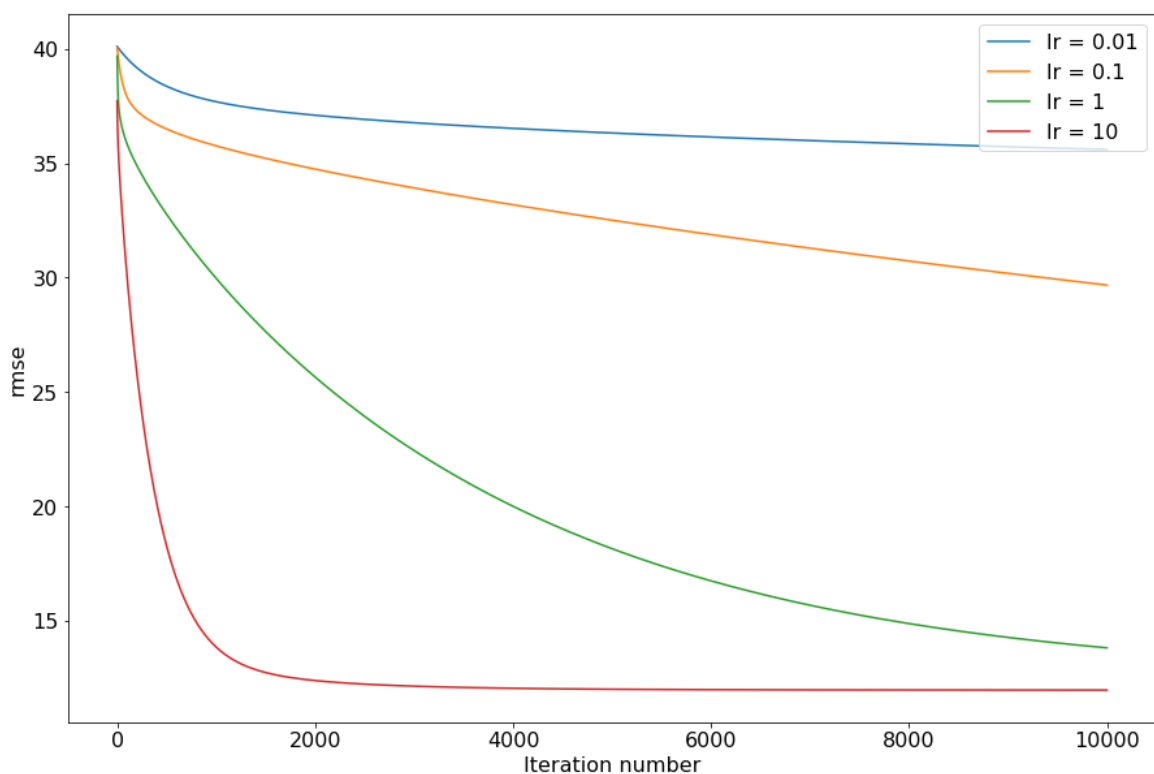
Homework 1 Report - PM2.5 Prediction

學號：r06922095 系級：資工碩一 姓名：陳代穎

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）。

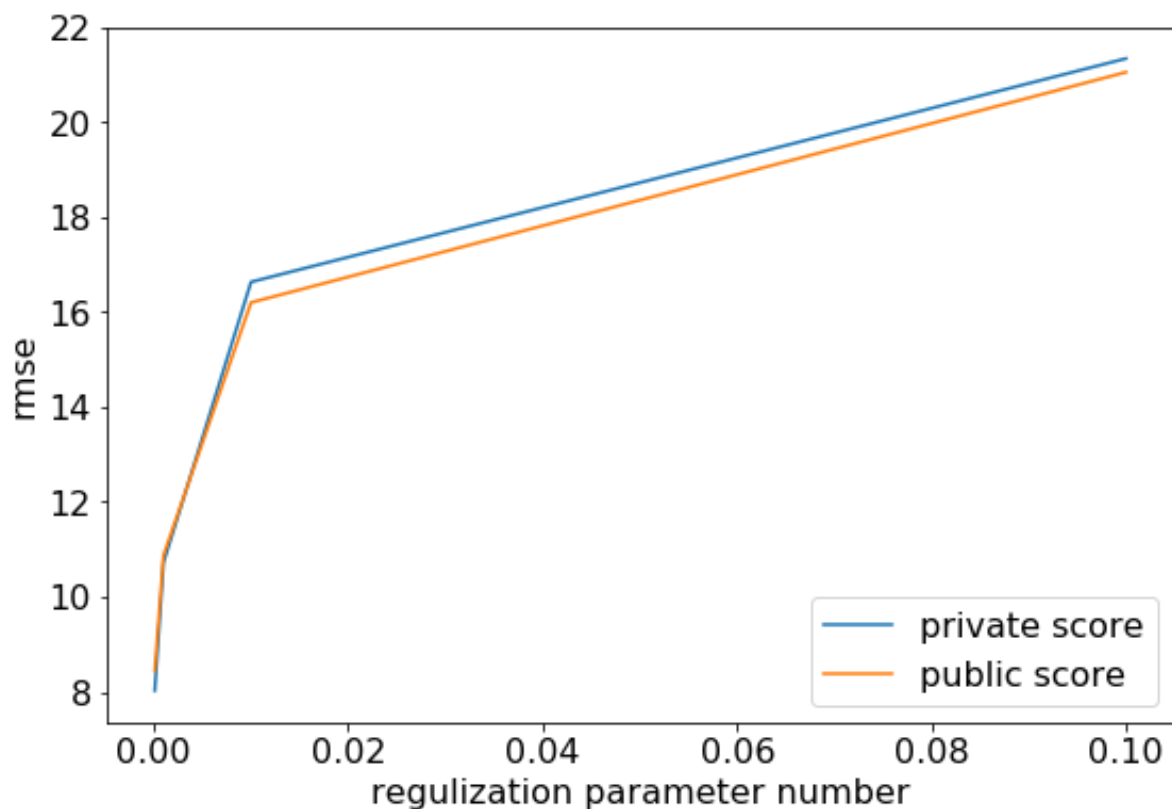
取同樣training set、learning rate及epochs的情況下，使用9小時內所有features的public/private score為8.03843/8.02530，使用9小時內PM2.5的public/private score為8.30004/8.65210。可以看出使用所有features進行training有小的root mean-square error。原因為取所有features的model有163個variables，而只取PM2.5的model僅有10個variables。因此前者的model相比之下較complex，因此bias相對較小；後者則為underfitting，具有較大的bias。

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。



learning rate = 10時，一開始的step大，能快速減少error，但隨著iteration的增加，adagrad使得step越來越小，最終找到local minima，收斂在11.96左右。learning rate太小時，如0.01，雖然step小，一開始error還是有些許下降，但最終由於adagrad造成step實在太小，收斂在35.60左右。

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）



Iteration: 20000 learning rate: 10 model: linear regression

分別用 $\lambda = 0.0001, 0.001, 0.01, 0.1$ 進行training，error隨著 λ 增加而增加，且private score與public score差異不大。因為regularization是用在overfitting的狀況，但因為linear regression model對於這個task偏向underfitting，在training data上就已經得不到好的效果，若再加上regularization只會使得training error更大，造成testing error的增加。 λ 越大regularization的影響就越大，從圖中便能看出此現象。

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？）

Data preprocessing：觀察training data後可以看到會有部分區域全為0的情況、PM2.5小於等於0的情況，必須把這些當作noisy移除。

Features：若只選用9個小時的PM2.5當作features會underfitting，因此最後選擇所有feature進行training。

訓練參數：在進行training時有試過直接用closed form solution，但可能是因為linear regression model的bias太大，因此反而效果沒有gradient descent來得好。最後選擇gradient descent，learning rate較大比較快，且因為有adagrad不必擔心不會收斂的問題。iteration的次數試過後發現太多圈反而error會變大，所以最後選擇15000次iteration。