

# Homework 1 Report - PM2.5 Prediction

學號：b05902019 系級：資工二 姓名：蔡青邑

**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）。**

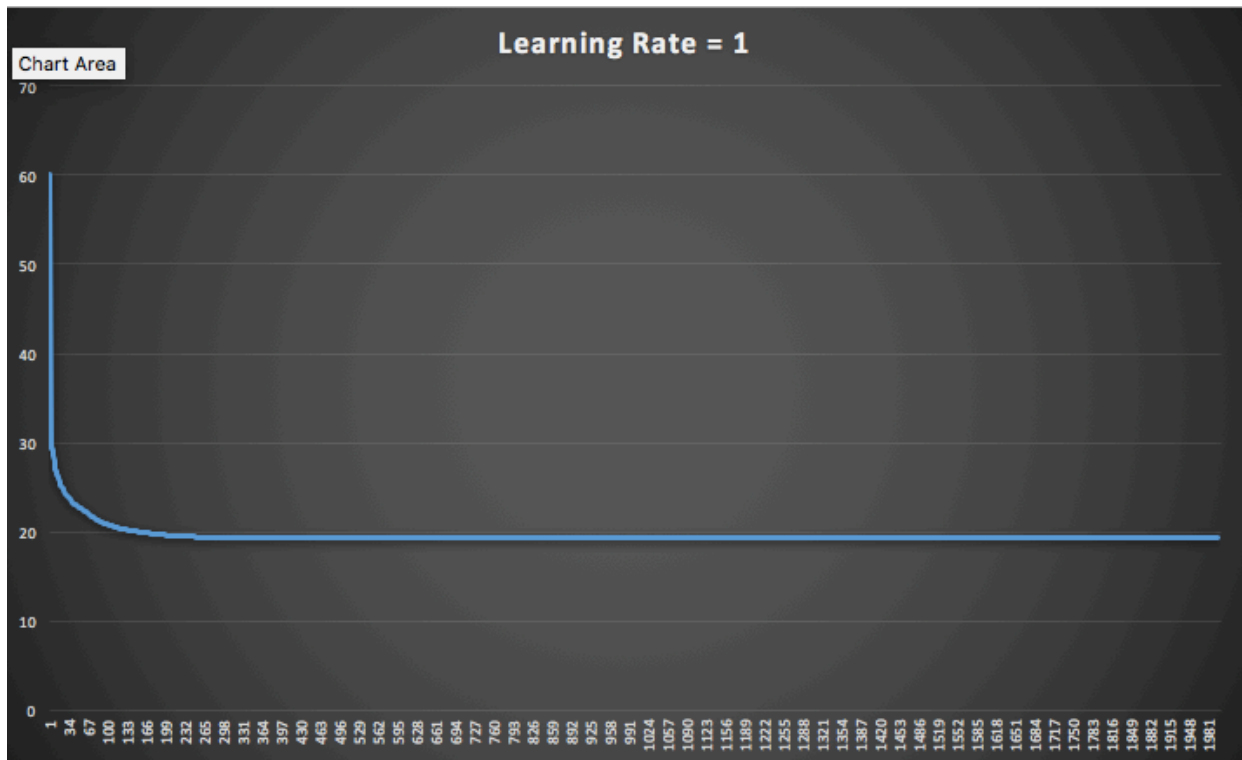
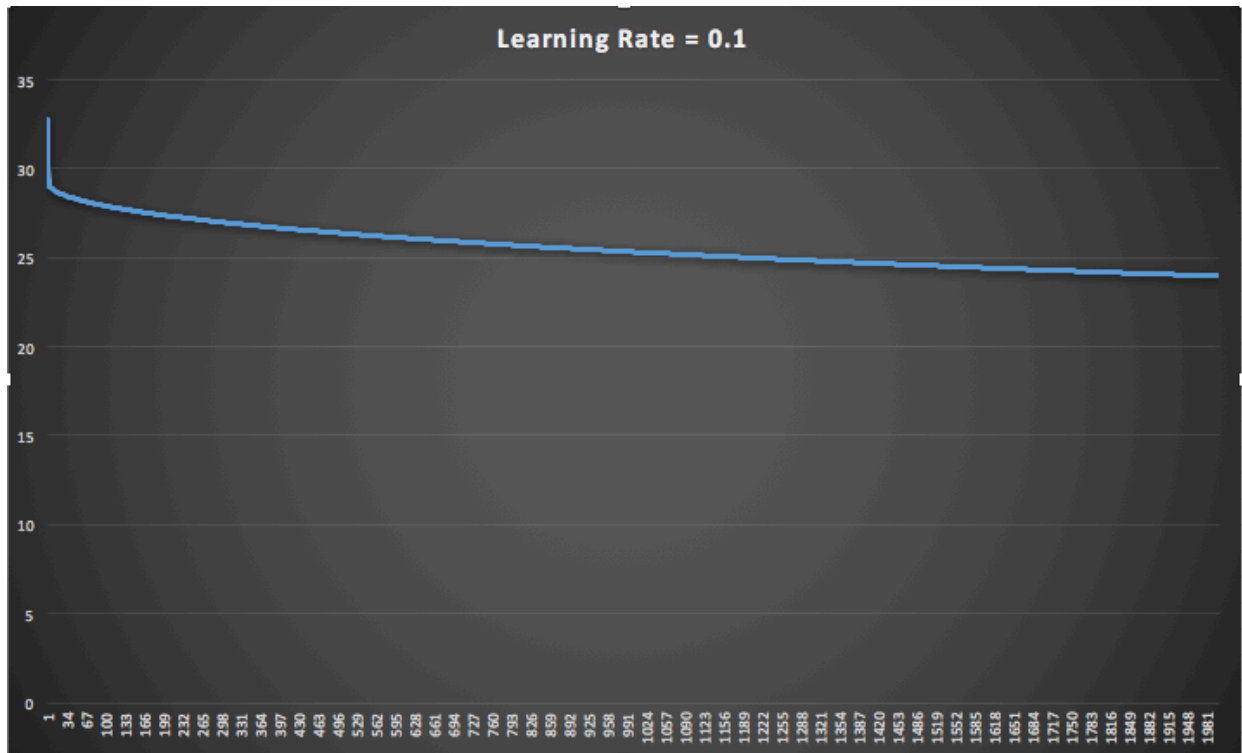
以下實作有使用到Adagrad,且learning rate = 1, 各做100000次的gradient descent, kaggle 分數如下：

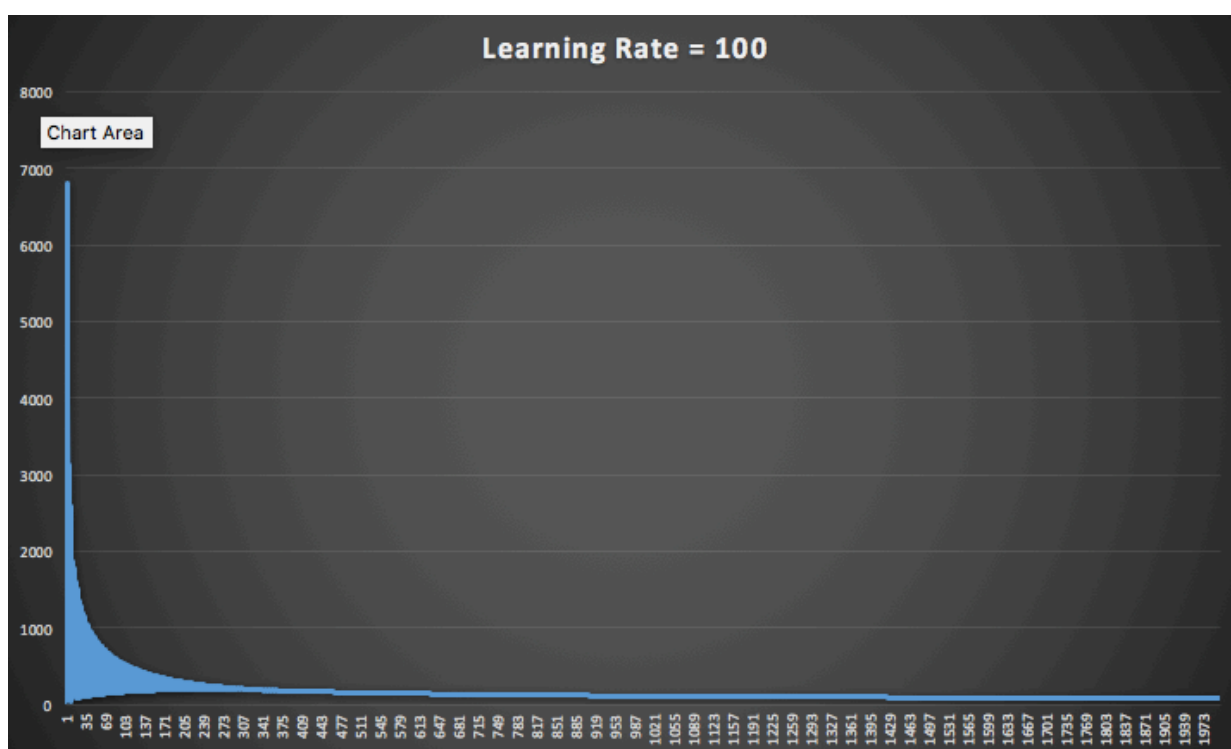
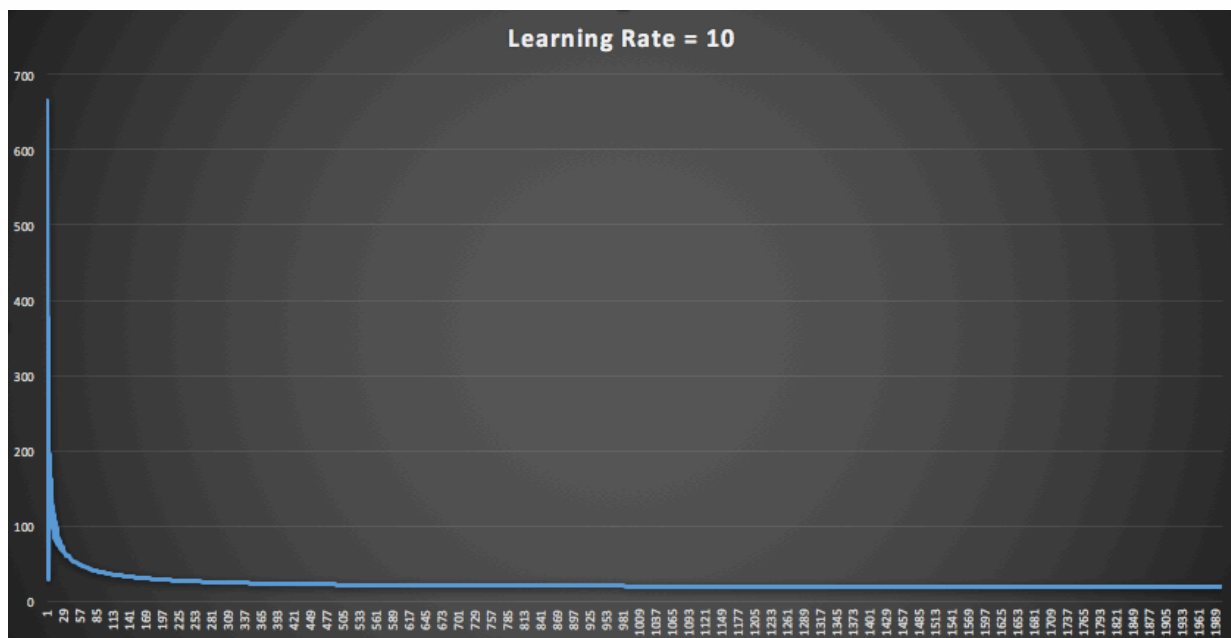
	僅含PM2.5	含所有feature
Private Score	10.55044	23.51726
Public Score	13.35283	29.86413

很明顯，僅含PM2.5的root mean-square error比較小。許多非PM2.5的feature可能跟我們要的結果沒有太大關係，但由於我們的training set不夠大，高估了其對於結果的影響，導致計算所有feature的時候產生了較為偏頗的model。

**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。**

以下實作有使用到Adagrad, 各做2000次的gradient descent, 且feature只取前一小時的PM2.5的一次項, Err收斂過程如下：





在

$\eta = 0.1$ 的時候，descenting的速度比較緩慢，因此超過最適值所產生的震盪也比較不明顯，然而卻比較慢才收斂到最佳值。而 $\eta = 100$ 的時候，雖然每次descent的幅度很大，但因為超過最適值而不斷震盪導致較晚才收斂。 $\eta = 1, 10$ 的時候，後者雖然有明顯的震盪但幅度比 $\eta = 100$ 時小，跟descenting較緩慢的前者比，兩者收斂的時間點差不多。

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

以下實作有使用到Adagrad, 各做100000次的gradient descent, feature取前一小時的PM2.5三次項, kaggle 分數如下:

$\lambda$	0.2	2	20000	2000000
Private Score	11.90348	11.91081	20.48295	38.68775
Public Score	12.59100	12.60700	21.93124	39.28392

$\lambda > 2$ 時, 由於model的參數受到 $\lambda$ 的影響被壓的越來越低, 遠離最佳model, 導致error 越來越大。

**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?)**

由於我並沒有對data做preprocessing, 沒有將training set裡不合理的PM2.5的值去掉(小於0或大於130), 因此這次的best\_hw1.sh我是將兩份不一樣的linear regression的結果merge而得 兩份都只有選用前一小時的PM2.5和PM10下去train, 但一份是三次式另外一份是一次式。由於3次式受到testing set裡不合理的值影響更大, 所以一旦遇到不合理的testing input我都用一次式的model來算 (兩次train的過程都有事先用closed form solution求值在做gradient descent,  $\lambda = 0$ , learning\_rate = 0.0005)