Homework 1 Report - PM2.5 Prediction

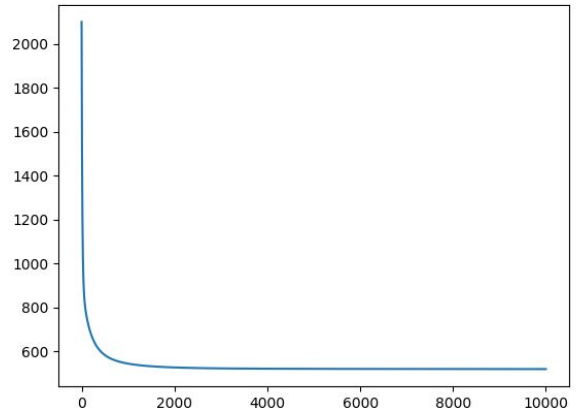
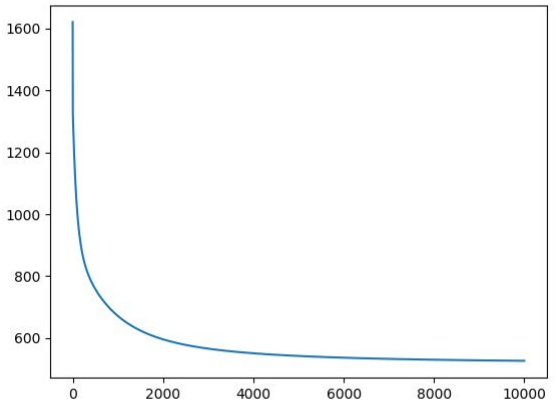
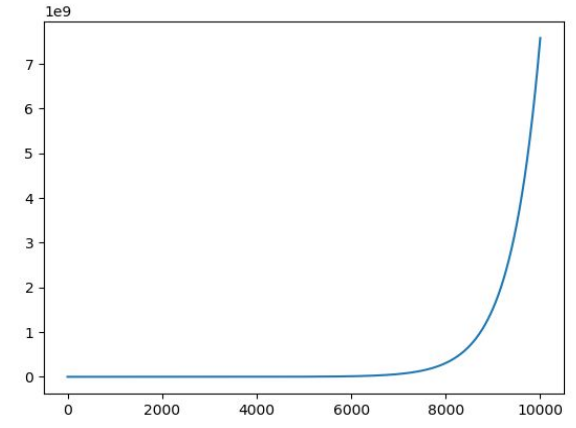
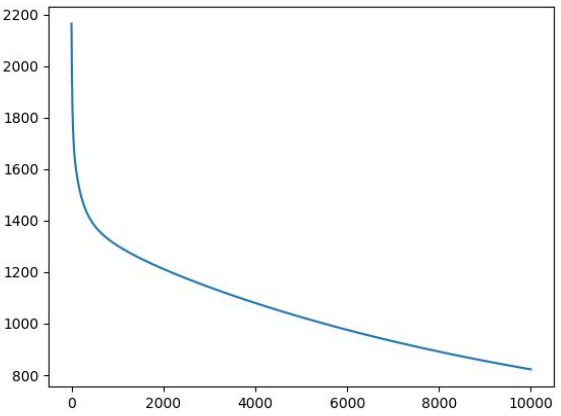
學號：r06921081 系級：電機所碩一 姓名：張邵瑀

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）。

	9小時內所有feature的一次項	9小時內PM2.5的一次項
public	8.83820	9.93353
private		

單純看PM2.5項太容易預測不準確，原因最主要是PM2.5的組成原因並不是單一污染源，所以他的變化量不能只看他自己本身的weight，若是如此這個實驗也不用做了，另外一個面向，就算他是單一污染源，還是有可能存在很多測量誤差，並且這時風雨這兩項也是可以當作參考的因素。

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:0.0000022	learning rate:0.0000001
	
其實在前面幾步就已經收斂到接進底部，但在後面的階段並沒有辦法再收斂了。	比起上一個更趨緩，但整體來說線條還是平滑的收斂。
learning rate:0.000002348	**learning rate:0.000002347**
	

測量只要超過0.000002348就無法收斂了	並且在0.000002347時反而收斂速度緩慢

結論：在learning rate 高的時候收斂速度非常的快速，但太大反而無法收斂，但adagrad不會產生無法收斂的情況，應該就是老師上課講的對於反差的處理，但是過低兩者都會遇到無法收斂的情況，所以選取適當的learning rate 是非常重要的。

3. (1%) 請分別使用至少四種不同數值的regulization 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）。

RMSE	λ = 0.1	λ = 10	λ = 100	λ = 1000
public	8.83820	8.83820	8.83822	8.83838

在這次的作業中作reglization 並沒有顯著的效果

4. (1%) 請這次作業你的best_hw1.sh是如何實作的？

這次的best_hw1.sh是仍然還是用linear regression來製作，而優化方法則是採用adagrad不是採用一般的gradient descent，因為在計算2次項的時候，若使用gradient descent 會常常算到數字過大以致於結果全部變成 nan，但使用adagrad就不會，其實我感覺兩者收斂的結果應該差不多，主要的差異在於feature的選擇與資料的選擇上，我試過所有選出來的feature做二次項並不會與單純取PM2.5的二次項好多少，也或許還會overfitting所以我只有取PM2.5的二次項，一開始我是把所有feature一個個拿掉看對於error的影響，但後來我考量到training data本身可能就有雜質，後來在選擇上我是用肉眼判斷法，把與PM2.5數值成正比的項次取出來使用，後來選用的資料是：

- AMB_TEMP、CH4、NMHC、NO2、O3、PM10、PM2.5、RAINFALL、SO2、$PM2.5^2$

後來我做n-fold把每個月份輪流當validation set 發現每個$h_i(\theta)$的差異太大，可能是資料不夠多，各個$h_i(\theta)$對各自的資料集overfitting，所以我把12個月合併成4個集合，接著我發現4個中前兩塊資料當validation set時，用其他的θ去算error都超高，我就確定某些月份的資料是有雜質的，所以我認為是在前6個月中有某些資料壞掉，所以用其他比較正常的資料做出來的θ驗證的時候error才這麼高，最後一個個嘗試選出來：

- 選用1,2,5,6,7,8,10,11月的資料

接著把他們放在一起train結果就過了strong base line

但經過檢討，我應該在取出較正常的月份之後再重新一個個feature測error變化量，這樣就可以避免用眼睛判斷相關性。