Homework 1 Report - PM2.5 Prediction

學號：b04902063　系級：資工三　姓名：陳昱儒

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）。

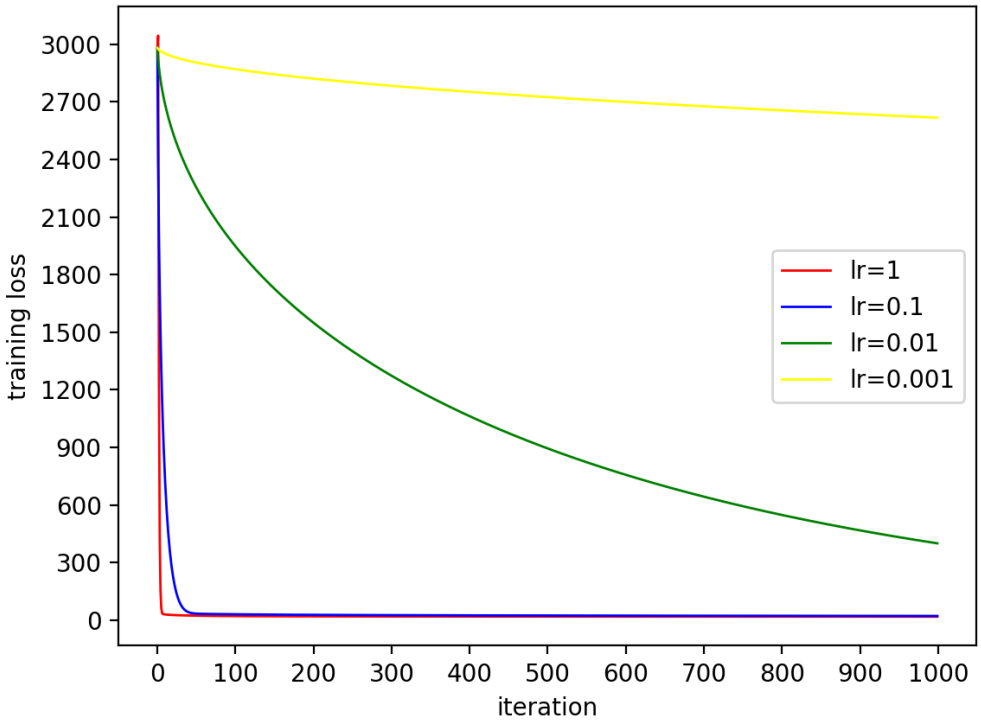
	Public score	Private score	RMSE
PM2.5	10.87565	10.40124	10.64109
All feature	9.36670	10.59109	9.99765

RMSE = $\sqrt{\frac{private^2+public^2}{2}}$

終端條件：training loss 與前次誤差 < 5*1e-7

由表中可知，若只取 PM 2.5 的 Root Mean Square Error 會比把 18 種污染物都拿下去 train 的結果來得高，代表這 18 種污染物裡也存有會影響預測出第 10 小時 PM2.5 的污染物，加入考慮能得到更好的結果。而對於在 private score 上 18 種污染物的預測結果比單純取 PM2.5 還高的狀況，我覺得可能是因為我的 model 參數的關係，對於那些與 PM2.5 零相關甚至負相關的污染物產生較明顯的反應，可能造成對第 10 小時的 PM2.5 預測結果產生較大的誤差而蓋過了較相關污染物的準確度，才造成這樣的結果。

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。



附註：實際上是訓練 10000 次，但取前 1000 次的資料來做圖以凸顯差異

由圖中可以很直接地發現，learning rate 為 0.001 的黃色線條下降速度非常緩慢，而 learning rate 為 0.1 的綠色線條下降速度中等，learning rate 為 0.1 與 1 的藍色與紅色線條下降速度均非常快，但又以 learning rate 為 1 的紅色下降速度最快，由此可見 learning rate 越高，下降速度（收斂速度）越快。

因 Adagrad 算法中 w 減少的值正比於 learning rate，所以 learning rate 設定值越高，下降的速度會越快。

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）。

	Public score	Private score	RMSE
λ = 1	8.68864	9.26661	8.98227
λ = 0.1	8.95387	9.90700	9.4424
λ = 0.01	9.00182	10.01915	9.52407
λ = 0.001	9.00697	10.03113	9.53281

由表中可知，若 regularization parameter λ 越高，其 public score、private score 的結果均較低，代表 model 本身可能就有 overfitting 的問題，若將 regularization parameter 提高，regularization 的效果更好，對於 overfitting 的改善更佳，使曲線更平滑，做出來的預測結果也會比較好。

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) 0 附近的 Preprocessing，因發現 train data 中有幾處一整大塊都是 0 的狀況，所以對於值為 0 的資料，使用其左右的值相加除以 2 代替它的值，而若遇到邊界則直接使用其前一個或後一個值代替他的值。

(2)有部分 train data 的資料值為負的，若是遇到負的值就把它調為零，並和上面的情況做一樣的處裡。

(3) Features 的選用是看相關係數，把所有 feature 和 pm2.5 的相關係數算過一遍，相關係數為負的踢掉，在測試怎麼樣的組合可以獲得較低的 loss

(4)使用 Adagrad 算法

(5)訓練相關參數則是從 0.5 和 0.1 開始測試，找到比較好的組合