學號：R07942091 系級： 電信碩一 姓名：許博閔

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (3) 題：

(1) 抽全部 9 小時內的污染源 feature 當作一次項(加 bias)

(2) 抽全部 9 小時內 pm2.5 的一次項當作 feature(加 bias)

備註 :

a. NR 請皆設為 0，其他的數值不要做任何更動

b. 所有 advanced 的 gradient descent 技術(如: adam, adagrad 等) 都是可以用的

c. 第 1-3 題請都以題目給訂的兩種 model 來回答

d. 同學可以先把 model 訓練好，kaggle 死線之後便可以無限上傳。

e. 根據助教時間的公式表示，(1) 代表 p = 9x18+1 而(2) 代表 p = 9*1+1

1~2 題的皆用 adagrad，learning rate = 1.5，80000 epoch

第三題只畫到 epoch = 10000

1. (2%)記錄誤差值 (RMSE)(根據 kaggle public+private 分數)，討論兩種 feature 的影響

全部 feature : train RMSE : 5.70389， kaggle public : 5.63588， kaggle private : 7.21184

只用 pm2.5 當 feature : train RMSE : 6.123021，public : 5.90263，private : 7.22356

當只用 pm2.5 當 feature 的時，RMSE 比較高，而且在 training 的時候很快就走到最佳解(loss 不再下降)，因此參數太少的時候，會有 underfitting 的問題出現

當用全部 feature 時，RMSE 較低，但如果 update 參數太多次，即使 training loss 還可以繼續下降，但 kaggle 的分數反而會下降，因此用全部 feature 可能會導致 overfitting

2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

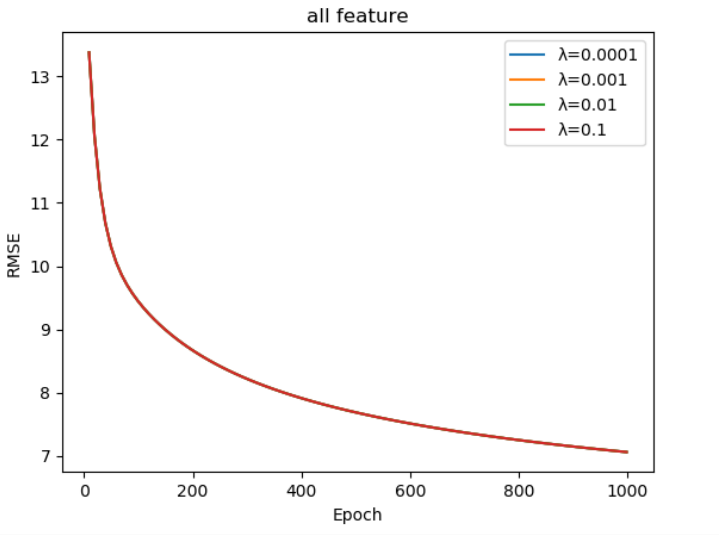
全部 feature : train RMSE : 5.829136，kaggle public :5.98341，kaggle private : 7.16508

只用 pm2.5 當 feature: train RMSE : 6.219417，public : 6.22749，private : 7.22464

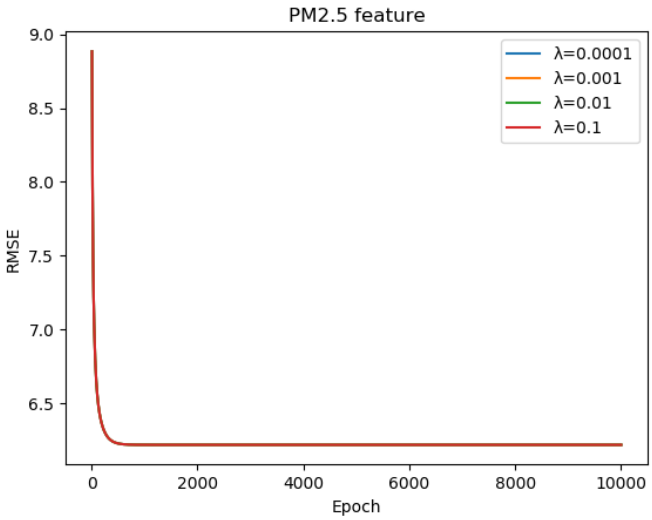
和第 1 題相比，無論是用全部 feature 或只用 pm2.5，改成只抽前 5 小時的情況下，train RMSE 都變高了，而 kaggle 分數只有全部 feature 的 private RMSE 變低，但我認為這是因為 model overfit public data 的原因，因此改只抽前 5 小時的情況下，model 的 feature 的數量變少了，預測結果會變比較差

3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖

(1)全部 feature :



(2) 只用 pm2.5 當 feature:



無論是全部 feature 或是只用 pm2.5 當 feature，lambda 的小或大畫出來的 loss 曲線都重疊，我認為是因為 model 是線性的，最初每個 weight = 1，若去看 training 的過程，都不會出現 weight 特別大的情況，導致 lambda 不夠大的時候，regularization 對結果沒有影響。

4. (1%)在線性回歸問題中，假設有 N 筆訓練資料，每筆訓練資料的特徵 (feature) 為一向量 x^n，其標註(label)為一純量 y^n，模型參數為一向量 w (此處忽略偏權值 b)，則線性回歸的損失函數(loss function)為$\sum_{n=1}^{N} (y^n - w \cdot x^n)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1\ x^2 \dots x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2 \dots y^N]^T$ 表示，請問如何以 X 和 y 表示可以最小化損失函數的向量 w ？請選出正確答案。(其中 X^TX 為 invertible)

(a) $(X^TX)X^Ty$
(b) $(X^TX)yX^T$
(c) $(X^TX)^{-1}X^Ty$　　答案 : (C) $(X^TX)^{-1}X^Ty$
(d) $(X^TX)^{-1}yX^T$