ML Report 1

學號：B04507025 系級： 電機四 姓名：韓秉勳

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

使用model共同資訊：

optimizer = adagrad

iteration = 200000

learning rate = 1

data = 每個月分開，每9小時training data , 但其中若任一feature<0，即拋棄該組資料，故全部只有5311組data

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

	RMSE of kaggle public	RMSE of kaggle private
all feature	5.52295	7.05220
pm2.5 only	5.79979	7.09273

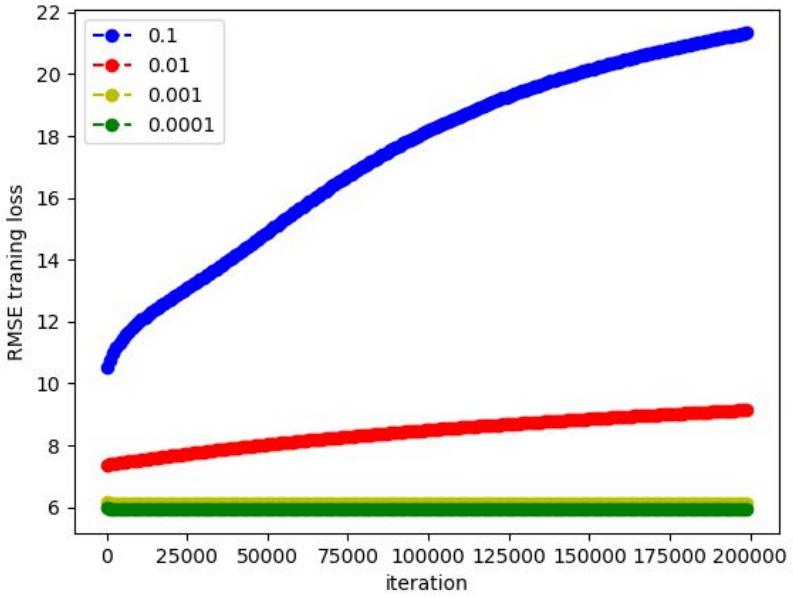
單獨pm2.5收斂較快，但public score比全取feature差。但可能由於pm2.5有多筆資料是負號，測試上data少了很多，故單取1feature沒有比較好的效果。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

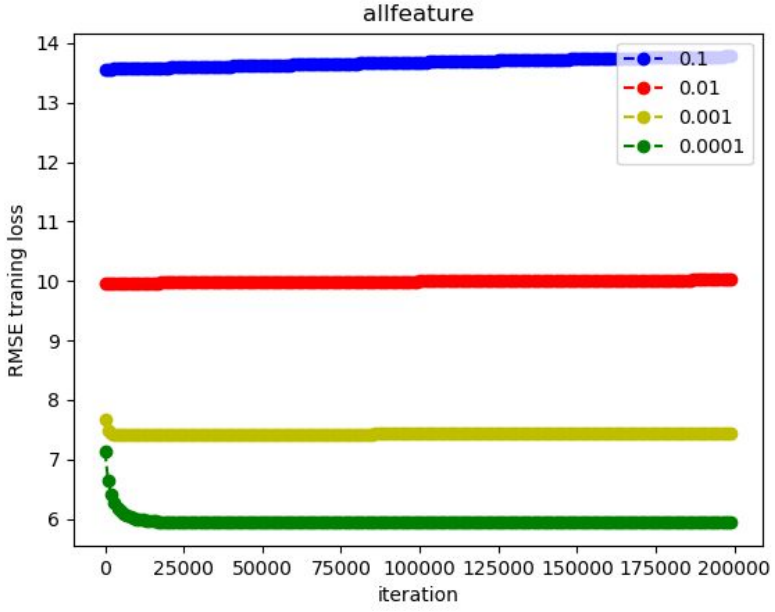
	public/private RMSE of kaggle - first 9 hr	public/private RMSE of kaggle - first 5 hr
all feature	5.52295 / 7.05220	5.91050 / 7.07862
pm2.5 only	5.79979 / 7.09273	6.18076 / 7.11060

由training loss來看，取5hr收斂比原本取9hr時快，如在只有pM2.5時 500 iteration即收斂，可能是參數少的關係，但也使取5hr的測試準確率並沒有像9hr這麼高。

3. (1%)Regularization on all the weight with λ=0.1、0.01、0.001、0.0001，並作圖
取9hr model為例（每1000 iteration 存一次點）:
PM2.5:



all feature:



lambda太大可能使model underfitting，而lambda太小則看不出幫助overfitting的方法。只取一個特徵由於資料點少波動大，regularization大時並無法幫助training loss下降，甚至

有上升趨勢。另外也看出regularization小時，loss平均較小（在pm2.5圖較明顯，綠線對藍線有顯著分別），但也可能是尚未train到overfitting，難看出上課投影篇中的loss到達低點後回升的現象。

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 $\sum_{n=1}^{N}(y^n - x^n \cdot w)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1\ x^2 \ldots x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1\ y^2 \ldots y^N]^T$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) $(X^TX)^{-1}X^Ty$
(d) $(X^TX)^{-1}yX^T$

Ans:(c)