學號：B05901111 系級： 電機三 姓名：陳建成

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

所有模型learning rate = 300，跑20000 個iterations，初始的weight全部為零

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

1. 抽全部污染源feature

RMSE: public = 5.68190， private = 7.26508

1. 只抽取全部pm2.5的feature

RMSE: public = 5.90263， private = 7.22356

　　由結果顯示，只取pm2.5訓練的model在public data上表現不如private data，但在private data卻反之。可以推測其他污染源的features對於pm2.5的濃度預測確實存在影響，但影響的關係可能並非簡單的線性迴歸可以模擬。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

1. 抽全部污染源feature

Training final loss = 5.8055

Kaggle RMSE: public = 5.97570， private = 7.22232

1. 只抽取全部pm2.5的feature

Training final loss = 6.2070

Kaggle RMSE: public = 6.22732， private = 7.22552

　　若單純就結果而論，表現都明顯變得比較差，甚至連training的loss也高出許多（至少高出0.5以上）。但是在training的過程中最顯而易見的變化是，大概才到不到四分之一的iteration過程loss幾乎就收斂很難再降低了。從跟前面比較起來，原因應該是因為參數本身就比較少，使得收斂的速度加快，但是也比較難得到較低的loss（結果比較不準確）。

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

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)

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 xn，其標註(label)為一純量 yn，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 。若將所有訓練資料的特徵值以矩陣 X = [x1 x2 … xN]T 表示，所有訓練資料的標註以向量 y = [y1 y2 … yN]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XTX為invertible)

1. (XTX)XTy
2. (XTX)yXT
3. (XTX)-1XTy
4. (XTX)-1yXT