學號：B06901087　系級：　電機二　姓名：翁瑋襄

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

Kaggle score (9 hours)	public	private
18 features	5.49185	7.04399
1 feature	5.80505	7.10989

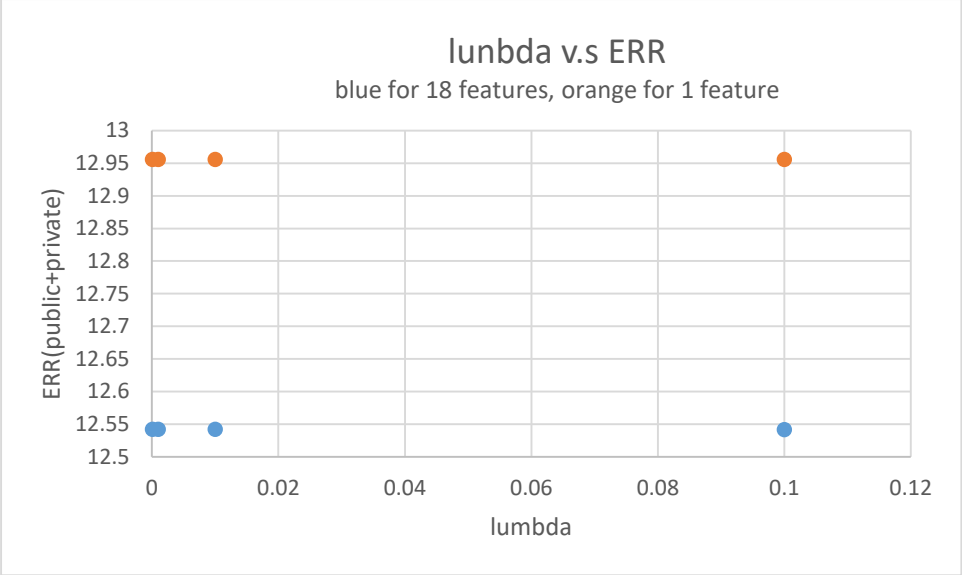
從表格中可以看出有 18 種 features 時預測的準確度比只靠 pm2.5 預測來的精確，顯示除了過去的 pm2.5 濃度外，還有其他各項指標(如是否下雨、其他粒子濃度)也會對未來的 pm2.5 數值造成影響。

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

Kaggle score (5 hours)	public	private
18 features	5.90539	7.05244
1 feature	6.17801	7.11682

可觀察出兩種 feature 進行 train 後，參考前九個小時預測的 pm2.5 結果都比只參考前五個小時預測出來的結果來的更加精確，顯示 pm2.5 的數值變化受到長時間波動而影響，非短暫五個小時內的資料就可預測下一個小時 pm2.5 的變化。

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



我認為助教提供的 lumbda 數值太小了，我自己 train 的時候 lumda 的數值會給到 5 上下才會明顯降低預測的 RMSE

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 n=1Nyn-xnw2 。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. $(X^TX)^{-1}X^Ty$
d. $(X^TX)^{-1}yX^T$

Ans. c