學號：B06902030 系級： 資工二 姓名：邱譯

請實做以下兩種不同feature的模型，回答第 (1) ~ (3) 題：
抽全部9小時內的污染源feature當作一次項(加bias)
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

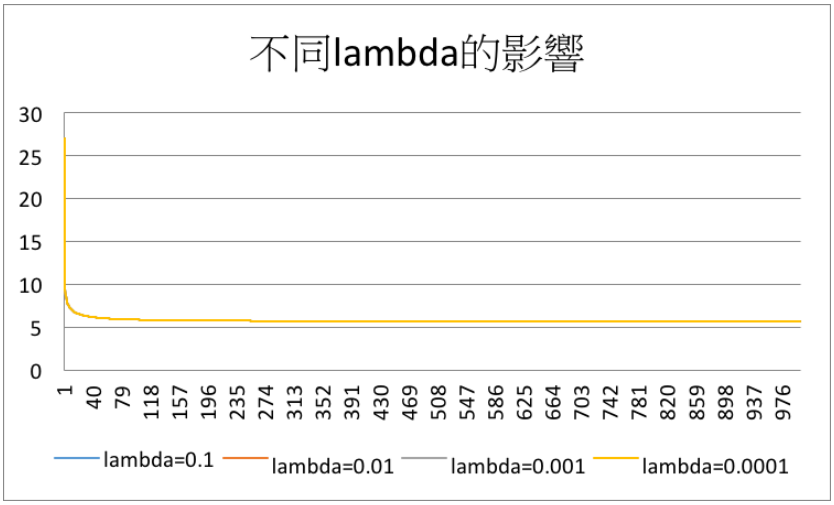
1. (2%)記錄誤差值 (RMSE)(根據kaggle public+private分數)，討論兩種feature的影響
(1)抽全部9小時內的污染源feature當作一次項(加bias)
RMSE = 5.63404(Public) 7.21573(Private)
(2)抽全部9小時內pm2.5的一次項當作feature(加bias)
RMSE = 5.90263(Public) 7.22355(Private)

抽取所有污染源比起只抽取pm2.5，無論在public set或是private set都會做得比較好，我認為是因為pm2.5的值不僅與過去pm2.5相關，也與許多因素相關，包含溫度、風、空氣中的許多種粒子等等，因此如果不僅僅利用過去的pm2.5，也加入許多其他的因素，應當可以得到更好的error。

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5小時內的污染源feature當作一次項(加bias)
RMSE = 5.98104(Public) 7.16810(Private)
(2)抽全部5小時內pm2.5的一次項當作feature(加bias)
RMSE = 6.22732(Public) 7.22548(Private)

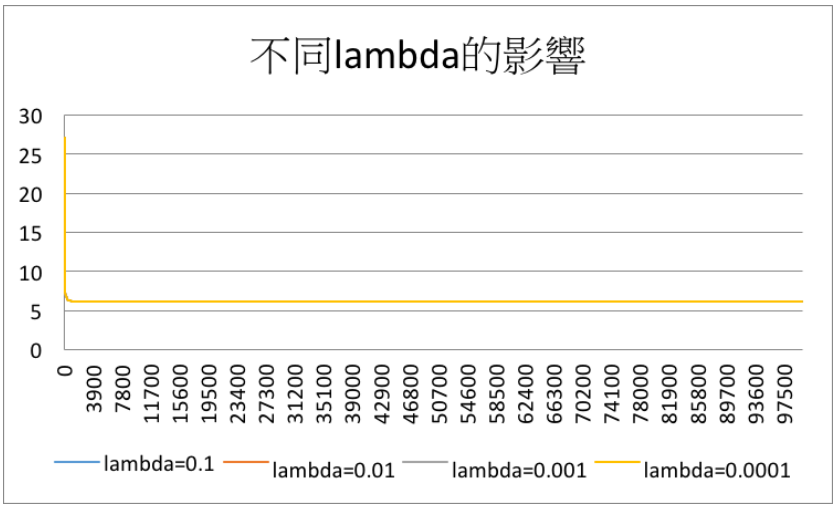
抽取9小時比起只抽取5小時，在public set及private set都可以做得比較好，因為9小時可以考慮更多的因素進來，如果我們說pm2.5的值會受過去影響，那麼就不可能只受過去5小時影響，只是時間遠近可能影響得多或少，因此考慮就9小時應得到比5小時還要好的error。

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)



縱軸為Ein
橫軸為iteration（單位為100次）

(2)抽全部9小時內pm2.5的一次項當作feature(加bias)



縱軸為Ein
橫軸為iteration（單位為1次）

可以發現不同λ的影響並不明顯。

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n=1Nyn-xnw2 。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)
(a) (XTX)XTy
(b) (XTX)yXT
(c) (XTX)-1XTy
(d) (XTX)-1yXT

Ans: (c)