學號：R06725005 系級： 資管碩一 姓名：郝思喬

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的一次項(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

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

|  |  |
| --- | --- |
| Model | RMSE(Private+Public) |
| (1) | 5.50427 + 7.83403 = 13.3383 |
| (2) | 5.62719 + 7.44013 = 13.06732 |

Model(1)的誤差較Model(2)的大，有可能的原因有某些feature在這Model中並非重要的因素，或是某些feature的資料是誤差較大的。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

|  |  |
| --- | --- |
| Model | RMSE(Private+Public) |
| (1) | 5.37824+ 7.73769 = 13.11593 |
| (2) | 5.79187 + 7.57904 = 13.37091 |

Model(1)的誤差反而較Model(2)的小，有可能的原因有某些污染源的資料在連續5小時內是較可以預測的，相對於只有用PM2.5的可以預測的較準確，也有可能是因為Testing Data本身資料誤差的一些關係導致結果相反。

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

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)為一存量 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 ？請寫下算式並選出正確答案。(其中XTX為invertible)

1. (XTX)XTy
2. (XTX)-0XTy
3. (XTX)-1XTy
4. (XTX)-2XTy

L(w) = = || y – Xw ||2

→ (XTX)w = XTy

→ w = (XTX)-1XTy

Ans: (c)