學號：R07942115 系級： 電信碩一 姓名：謝硯澤

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

|  |  |  |
| --- | --- | --- |
|  | Pubilc Score | Private Score |
| 前9小時的所有feature | 8.02776 | 9.03806 |
| 前9小時只有PM2.5 | 6.09551 | 7.04791 |

在我實作的程式中，只抽取PM2.5的模型，在kaggle上的分數是遠好於抽取全部feature的。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

|  |  |  |
| --- | --- | --- |
|  | Pubilc Score | Private Score |
| 前5小時的所有feature | 18.47992 | 18.90672 |
| 前5小時只有PM2.5 | 17.79629 | 19.89730 |

將feature從抽前9小時改成抽前5小時，只取PM2.5的模型，在Kaggle上的分數更低，可能的原因是feature數太少了(p=5+1)，抽取全部feature的模型分數會稍微好一點，但兩種模型皆不理想。

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| λ | AllFeature\_public | AllFeature\_private | PM2.5\_public | PM2.5\_private |
| 0.1 | 7.54748 | 8.68879 | 5.91274 | 7.2816 |
| 0.01 | 6.41488 | 10.8005 | 6.14957 | 7.41655 |
| 0.001 | 8.12723 | 8.36323 | 6.04844 | 7.44476 |
| 0.0001 | 7.23828 | 8.38745 | 5.73555 | 7.0238 |

加入Regularization後，表現並沒有提升，推測可能是因為我的model並沒有overfitting的問題，所以有沒有加入Regularization影響不大。

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

Ans. (C)