學號：R05546022 系級： 工工所碩二 姓名：謝立成

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

|  |  |  |
| --- | --- | --- |
| 9hr | All feature | only pm2.5 |
| Private | **7.4663** | **7.4401** |
| public | **5.3011** | **5.6272** |
| RMSE | **6.4748** | **6.5962** |

We can see that although only PM2.5 can train the public data lower, the private data is better when all feature is taken into account.

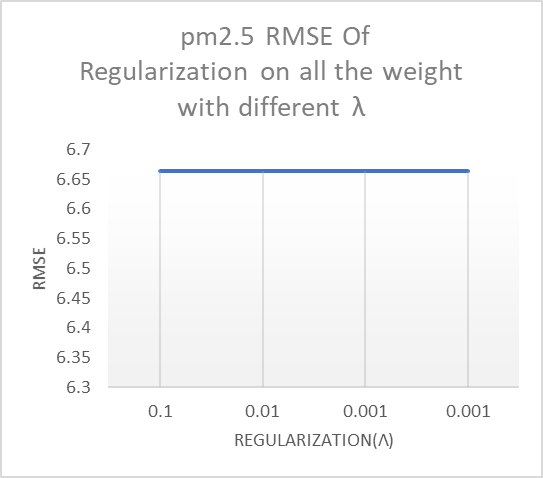
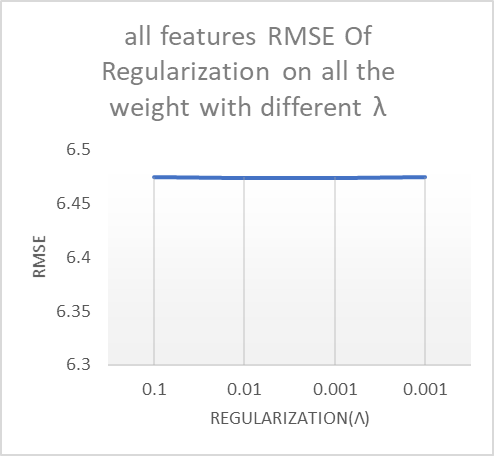
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

|  |  |  |
| --- | --- | --- |
| 5hr | All feature | only pm2.5 |
| Private | **7.7375** | **7.579** |
| public | **5.3782** | **5.7919** |
| RMSE | **6.6631** | **6.7449** |

Same result as above, but is poorer than that of 9 hour training

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

It appears to be little difference when λ is different



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Regularization(λ)** | **public** | **private** | **RMSE** |
| **all features** | **0.1** | **7.4663** | **5.301** | **6.4748** |
| **0.01** | **7.4663** | **5.301** | **6.4748** |
| **0.001** | **7.4663** | **5.301** | **6.4748** |
| **0.001** | **7.4663** | **5.3011** | **6.4748** |
| **only pm2.5** | **0.1** | **7.7375** | **5.3782** | **6.6631** |
| **0.01** | **7.7375** | **5.3782** | **6.6631** |
| **0.001** | **7.7375** | **5.3782** | **6.6631** |
| **0.001** | **7.7375** | **5.3782** | **6.6631** |

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)

ANS : (c)

