學號：B04901060 系級： 電機三 姓名：黃文璁

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**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 9小時 | public | private |
| 全部汙染源 | 7.46237 | 5.53562 |
| PM2.5 | 7.44013 | 5.62719 |
|  |  |  |

**答：** 9小時全部汙染源的public RMSE比只抽PM2.5的RMSE來得高（約高0.02）。

9小時全部汙染源的private RMSE比只抽PM2.5的RMSE來得低（約低0.1）。

由於public、private沒有絕對高低，兩種feature就本題來說不容易分出好壞。

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

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 5小時 | public | private |
| 全部汙染源 | 7.65925 | 5.44092 |
| PM2.5 | 7.57904 | 5.79187 |
|  |  |  |

**答：** 5小時全部汙染源的public RMSE比只抽PM2.5的RMSE來得高（約高0.08）

5小時全部汙染源的private RMSE比只抽PM2.5的RMSE來得低（約低0.35）

綜合public和private分數來看，抽5小時全部汙染源的結果較好。

另外值得注意的是，抽5小時全部汙染源在private set的結果要比抽9小時來得好。

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
| Regularization | λ | 0.0001 | 0.001 | 0.01 | 0.1 |
| 全部汙染源 | public | 7.46237 | 7.46236 | 7.46233 | 7.46198 |
|  | private | 5.53562 | 5.53561 | 5.53553 | 5.53477 |
| PM2.5 | public | 7.44013 | 7.44013 | 7.44013 | 7.44012 |
|  | private | 5.62719 | 5.62719 | 5.62719 | 5.62720 |
|  |  |  |  |  |  |

**答：** 在本題指定的λ範圍中，regularization的效果不大（如上圖）。

**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**

**答：** (c)

Proof:

Using matrix calculus: