**Machine Learning (2017,Fall)**

**Assignment 1 - Predicting PM2.5**

學號：b03901086  系級： 電機四  姓名：楊正彥

1

|  |  |  |  |
| --- | --- | --- | --- |
|  | Public Score | Private Score | Error |
| 取前9小時內所有污然源的一次項當features |  |  |  |
| 取前9小時內PM2.5的一次項當features |  |  |  |

2

|  |  |  |
| --- | --- | --- |
|  | Public Score | Error |
| 取前9小時內所有污然源的一次項當features |  |  |
| 取前5小時內所有污然源的一次項 |  |  |

3 Regularization

4

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

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

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