

HW1: Report

學號：R06942074 系級：電信所碩一 姓名：李宇哲

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

| | RMSE (training) 10000 iteration | RMSE (public) 10000 iteration | RMSE(private) 10000 iteration |
|-------|------------------------------------|----------------------------------|----------------------------------|
| 18種 | 5.955 | 7.81 | 5.47 |
| PM2.5 | 6.20 | 7.47 | 5.64 |

從這裡可以發現:

兩個model都跑了10000個iteration，雖然18種的feature的training的RMSE比較小，但是在testing dataset上沒有明顯的進步反而且在public退步，理論上162個feature應該要比9個feature還要好，但是會發生這樣的狀況可以推論這18種feature讓model提早收斂到local minimum，影響model的performance。再者，在162個feature提早達到很低的loss也可以推論增加的153個feature只會更容易讓model發生overfitting的現象。

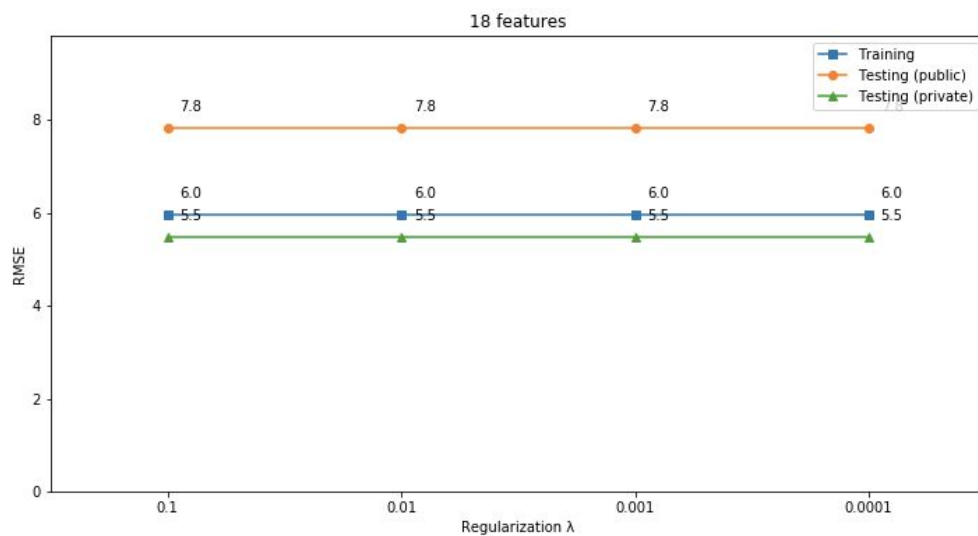
2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

| 9 hours | RMSE (training) 10000 iteration | RMSE (public) 10000 iteration | RMSE(private) 10000 iteration |
|---------|------------------------------------|----------------------------------|----------------------------------|
| 18種 | 5.955 | 7.81 | 5.47 |
| PM2.5 | 6.20 | 7.47 | 5.64 |

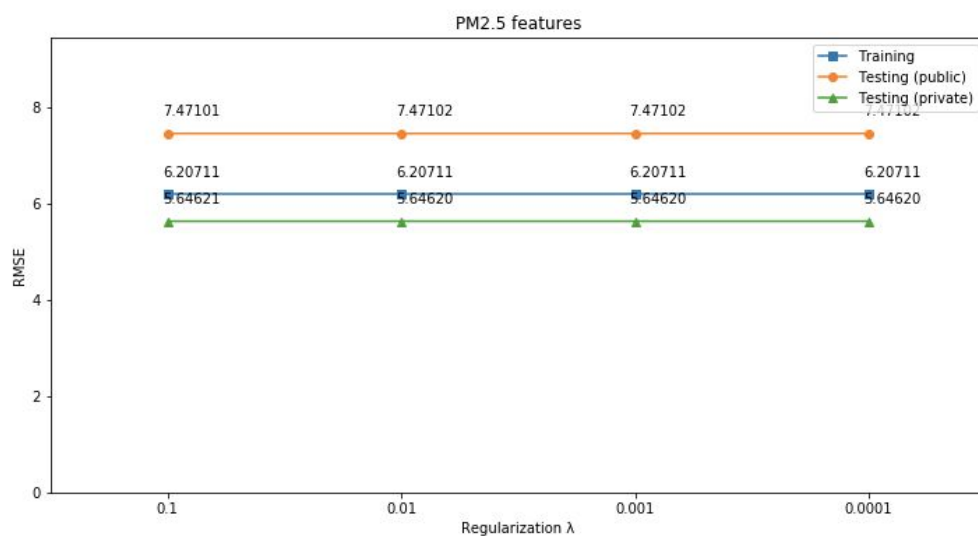
| 5 hours | RMSE (training) 10000 iteration | RMSE (public) 10000 iteration | RMSE(private) 10000 iteration |
|---------|------------------------------------|----------------------------------|----------------------------------|
| 18種 | 5.946 | 7.72778 | 5.37111 |
| PM2.5 | 6.293 | 7.59792 | 5.80192 |

這裡可以發現，相較於9hour，5hours在feature減少後應該會比較差或者是一樣的 performance 像是PM2.5的狀況，但是18種feature在testing dataset有上升的趨勢只能說減少feature候在10000個iteration下比較不容易overfitting，這裡不能說明前5hours會比較好。

3. (1%)Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖 training & testing 18種features:



PM2.5:



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 x^n ，其標註 (label) 為一存量 y^n ，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 $\sum_{n=1}^N (y^n - x^n \cdot w)^2$ 。若將所有訓練資料的特徵值以矩陣 $X = [x^1 \ x^2 \ \dots \ x^N]^T$ 表示，所有訓練資料的標註以向量 $y = [y^1 \ y^2 \ \dots \ y^N]^T$ 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 $X^T X$ 為 invertible)

$$\begin{aligned}
 \text{Loss} &= (y - Xw)^T (y - Xw) \\
 \nabla_w \text{Loss} &= \frac{\partial}{\partial w} (y - Xw)^T (y - Xw) \\
 &= \frac{\partial}{\partial w} (\cancel{y^T y} - \cancel{w^T X^T y} - \cancel{y^T X w} + w^T X^T X w) \\
 &= 2X^T X w - 2X^T y \\
 &= 0 \\
 X^T X w &= X^T y \\
 \underline{w} &= \underline{(X^T X)^{-1} X^T y}
 \end{aligned}$$

$$w = (X^T X)^{-1} X^T y$$