

# Machine Learning

## Homework 1 Report

學號：R06922117 系級：資工碩一 姓名：李岳庭

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

(1)  $RMSE = 7.44992 + 5.28679 = 12.73671$

(2)  $RMSE = 7.33508 + 5.79906 = 13.13414$

抽取全部的 feature 在 public 表現比較差，但 private 表現比較好，相加之後比較，抽取全部 feature 的表現較佳，可見挑 feature 是這次作業的關鍵。

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

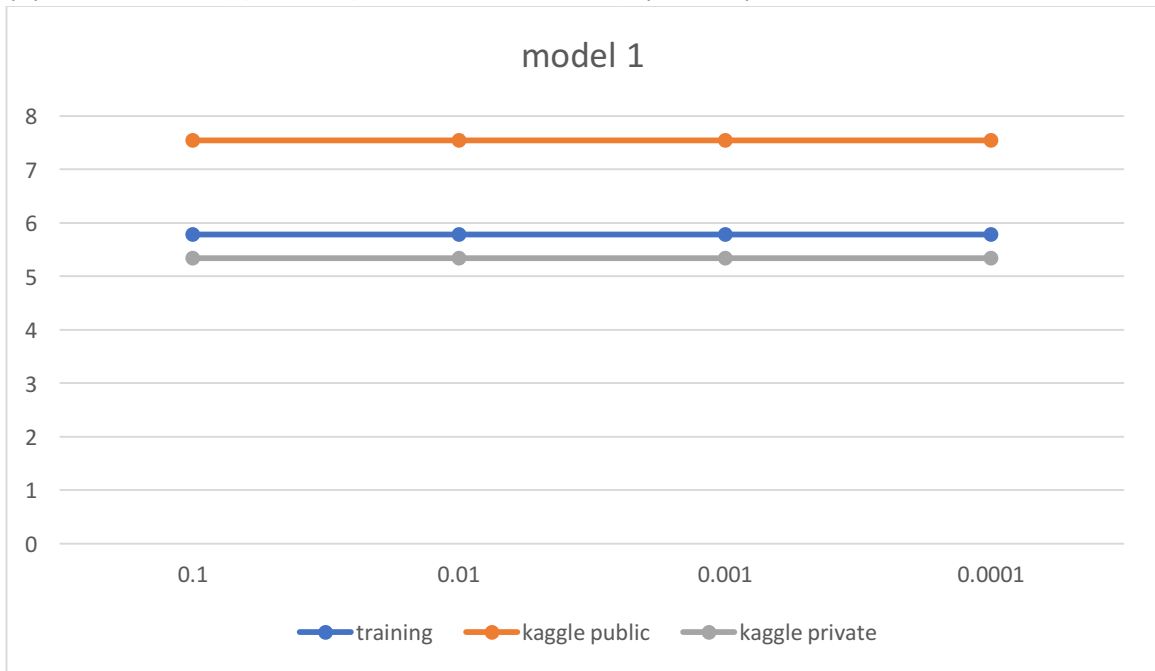
(1)  $RMSE = 7.63920 + 5.34941 = 12.98861$

(2)  $RMSE = 7.44395 + 5.91309 = 13.35704$

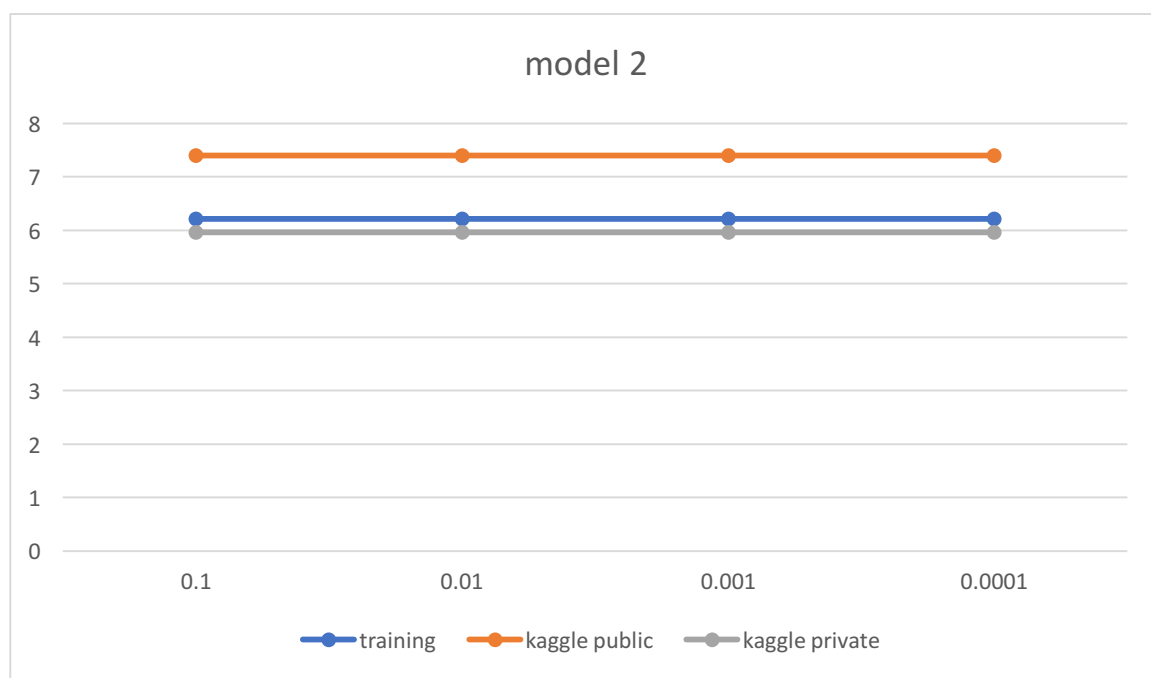
相較於 9 小時的結果，只取 5 小時的表現較差。

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

(1) 抽全部 9 小時內的污染源 feature 的一次項(加 bias)



(2)抽全部 9 小時內的污染源 feature 的一次項(加 bias)



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  $\mathbf{x}^n$ ，其標註(label)為一存量  $y^n$ ，模型參數為一向量  $\mathbf{w}$  (此處忽略偏權值  $\mathbf{b}$ )，則線性回歸的損失函數(loss function)為  $\sum_{n=1}^N (\mathbf{y}^n - \mathbf{x}^n \cdot \mathbf{w})^2$ 。若將所有訓練資料的特徵值以矩陣  $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^N]^T$  表示，所有訓練資料的標註以向量  $\mathbf{y} = [\mathbf{y}^1 \mathbf{y}^2 \dots \mathbf{y}^N]^T$  表示，請問如何以  $\mathbf{X}$  和  $\mathbf{y}$  表示可以最小化損失函數的向量  $\mathbf{w}$ ？請寫下算式並選出正確答案。(其中  $\mathbf{X}^T \mathbf{X}$  為 invertible)

- (a)  $(\mathbf{X}^T \mathbf{X}) \mathbf{X}^T \mathbf{y}$
- (b)  $(\mathbf{X}^T \mathbf{X})^{-0} \mathbf{X}^T \mathbf{y}$
- (c)  $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$
- (d)  $(\mathbf{X}^T \mathbf{X})^{-2} \mathbf{X}^T \mathbf{y}$

$$\begin{aligned}
 E = \text{loss function} &= \| \mathbf{X} \mathbf{w} - \mathbf{y} \|^2 \\
 &= (\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w} - 2 \mathbf{w}^T \mathbf{X}^T \mathbf{y} + \mathbf{y}^T \mathbf{y}) \\
 \xrightarrow[\text{偏微}]{\text{對 } \mathbf{w}} \quad 2(\mathbf{X}^T \mathbf{X} \mathbf{w} - \mathbf{X}^T \mathbf{y}) &\stackrel{!}{=} 0 \\
 \mathbf{X}^T \mathbf{X} \mathbf{w} - \mathbf{X}^T \mathbf{y} &= 0 \\
 \mathbf{X}^T \mathbf{X} \mathbf{w} &= \mathbf{X}^T \mathbf{y} \\
 \mathbf{w} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}
 \end{aligned}$$

ANS: (c)