

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

	public	private	total
Model 1	7.44992	5.28679	6.45954551
Model 2	7.33508	5.79906	6.61182635

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。

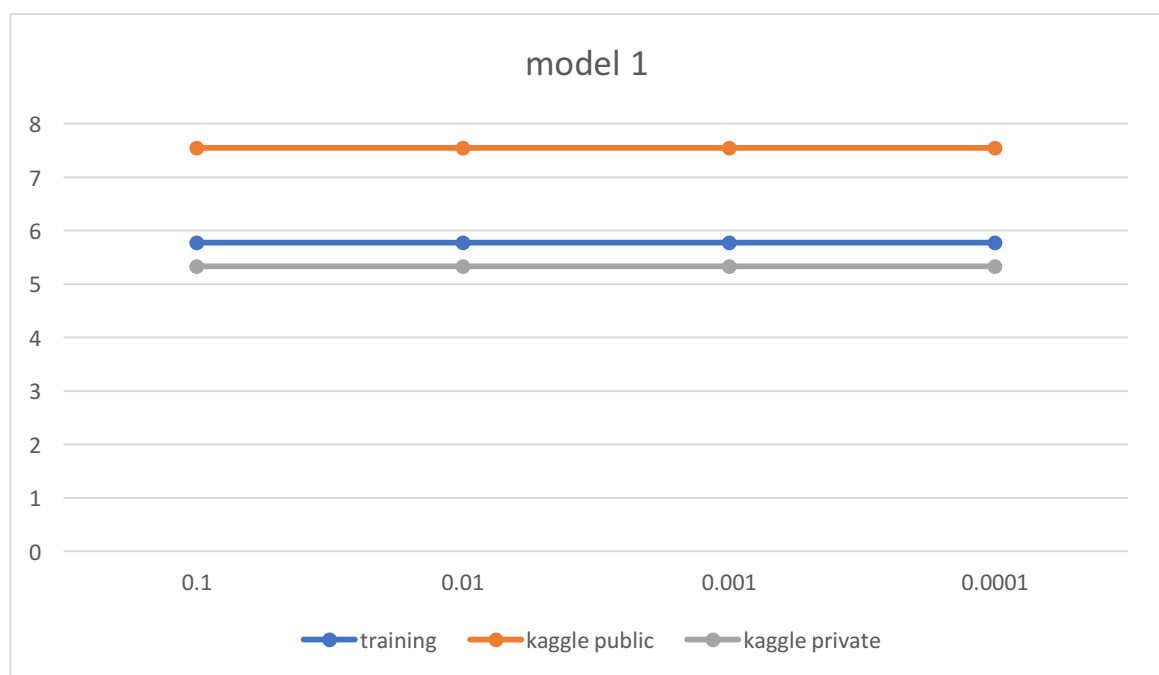
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

	public	private	total
Model 1	7.63920	5.34941	6.59445085
Model 2	7.44395	5.91309	6.72224014

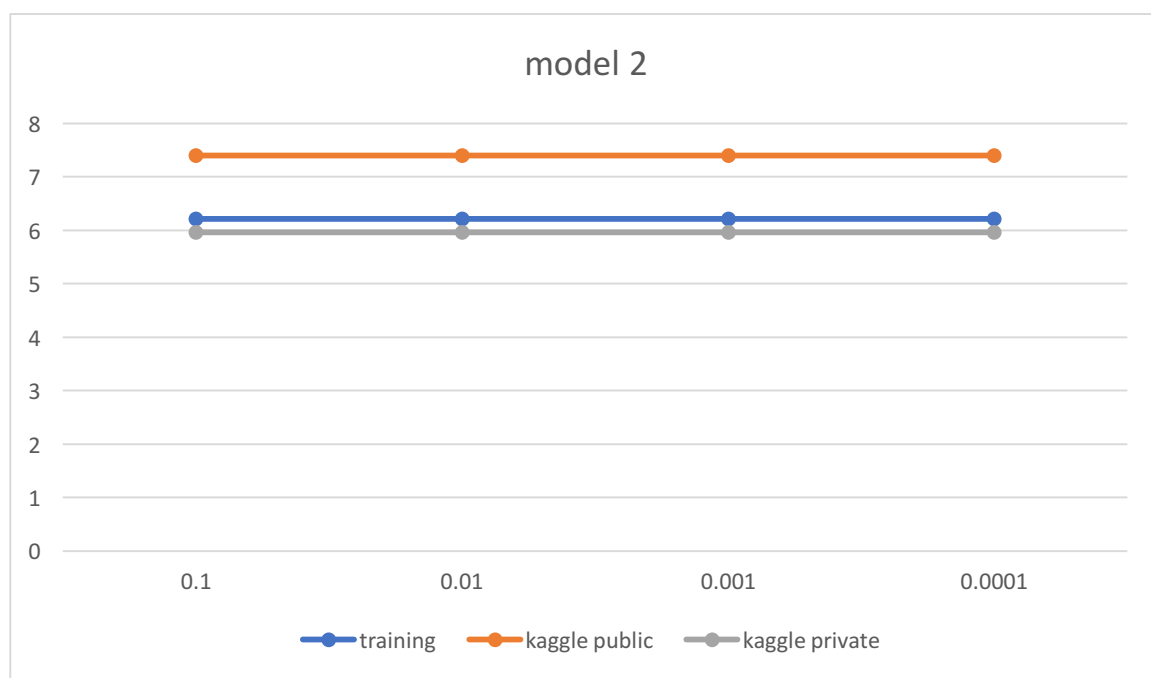
相較於 9 小時的結果，只取 5 小時的表現較差。

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 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)

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

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

ANS: (c)