

# 作業一 Report

學號：B06901020 系級：電機二 姓名：張恆瑞

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 當作一次項(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
- c. 第 1-3 題請都以題目給訂的兩種 model 來回答
- d. 同學可以先把 model 訓練好，kaggle 死線之後便可以無限上傳。
- e. 根據助教時間的公式表示，(1) 代表  $p = 9 \times 18 + 1$  而(2) 代表  $p = 9 \times 1 + 1$

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

	$p = 9 \times 18 + 1$	$p = 9 \times 1 + 1$
public	5.48861	5.80086
private	7.03200	7.10380

使用較少 feature 之 model 不論在 public 還是 private data set 上表現都較差，可見其他的污染物也和 PM2.5 有關係，也就是說，把其他污染物考慮進預測模型，則可得到更好的預測結果。

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

下表為抽前 5 小時之結果(抽前 9 小時之結果在上一題)

	$p = 9 \times 18 + 1$	$p = 9 \times 1 + 1$
public	5.92162	6.17484
private	7.13021	7.10928

不論是在 public、private 或是抽 18 個 feature、1 個 feature 的結果，表現都很明顯降低了許多，可見如果知道更多先前的資料更能夠預測接下來的 PM2.5 量。

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

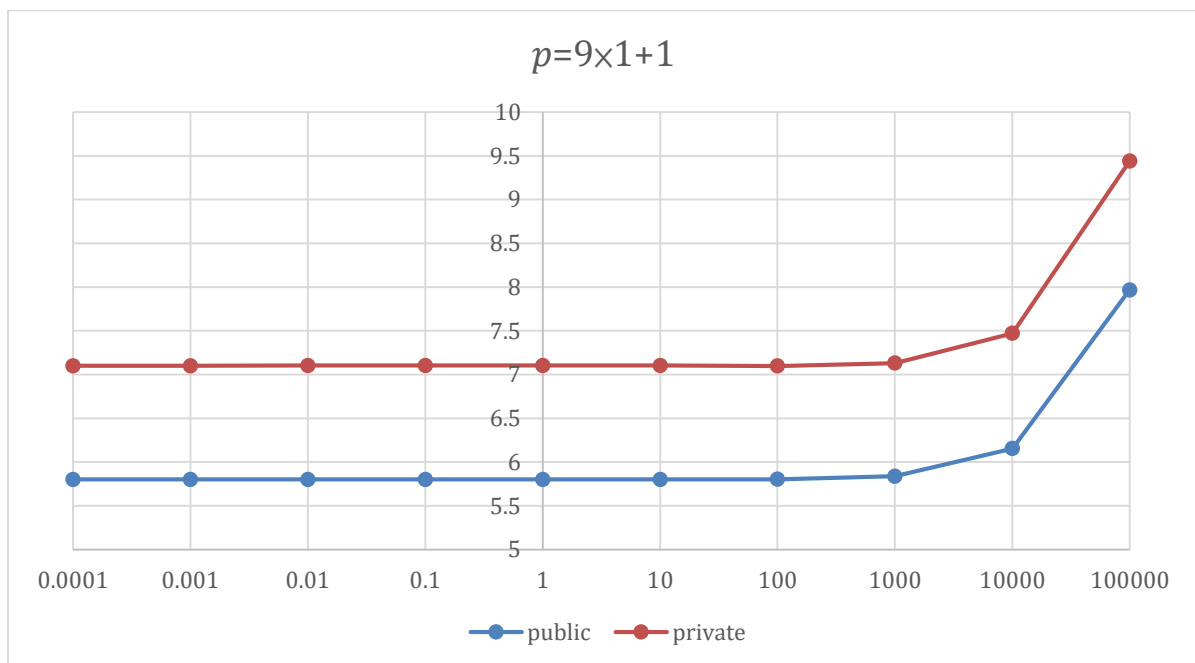
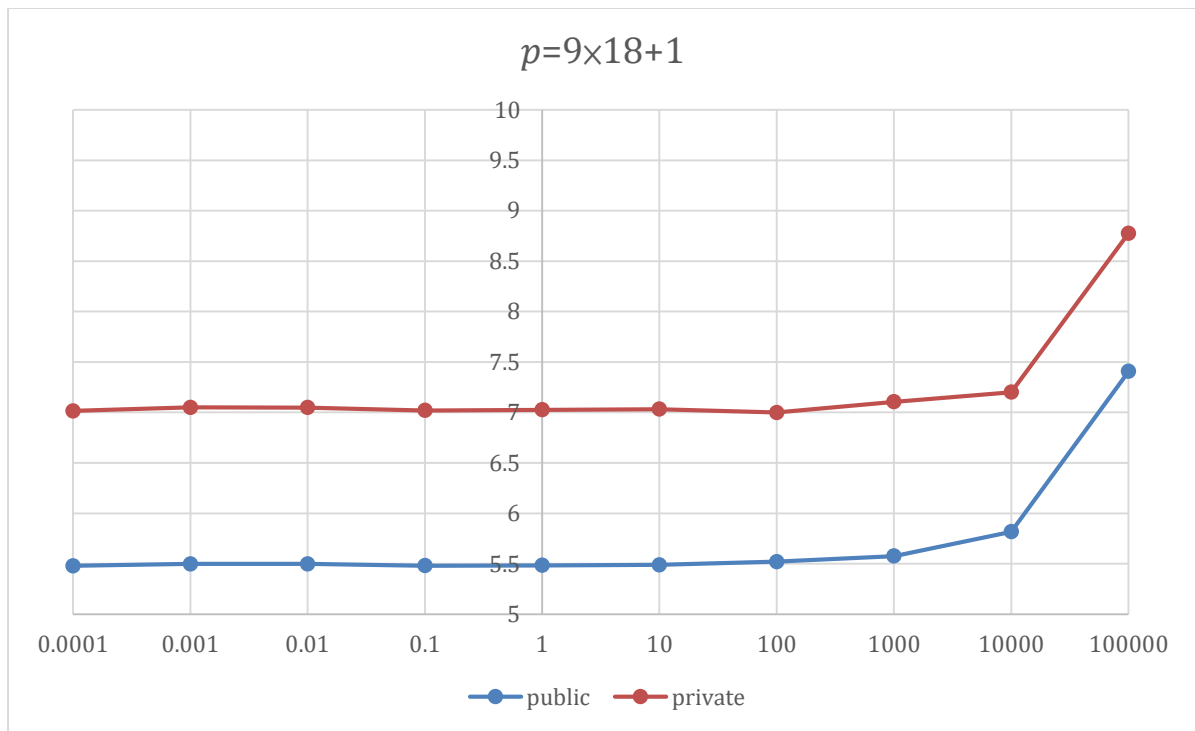
$p = 9 \times 18 + 1$ ：

	0.0001	0.001	0.01	0.1	1	10	100	1000	10000	100000
public	5.47893	5.49892	5.49824	5.48142	5.48405	5.48861	5.52034	5.57541	5.81772	7.40728
private	7.01496	7.04982	7.04886	7.02044	7.02571	7.03200	6.99904	7.10547	7.20117	8.77545

$p = 9 \times 1 + 1$ ：

	0.0001	0.001	0.01	0.1	1	10	100	1000	10000	100000
public	5.80052	5.80052	5.80078	5.80078	5.80079	5.80086	5.80241	5.83756	6.15334	7.96774
private	7.10006	7.10006	7.10411	7.10411	7.10408	7.10380	7.09883	7.13065	7.47044	9.43974

註：紅字為最小值



從結果得知， $\lambda$ 在很小時對於 public data 較有利，但在稍大時對 private data 較有利，然而在 $\lambda$ 過大時就使得 model 的精確度降低很多。

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)為一純量  $y^n$ ，模型參數為一向量  $\mathbf{w}$  (此處忽略偏權值  $b$ )，則線性回歸的損失函數(loss function)為  $\sum_{n=1}^N (y^n - \mathbf{x}^n \cdot \mathbf{w})^2$ 。若將所有訓練資料的特徵值以矩陣  $X = [\mathbf{x}^1 \ \mathbf{x}^2 \ \dots \ \mathbf{x}^N]^T$  表示，所有訓練資料的標註以向量  $\mathbf{y} = [y^1 \ y^2 \ \dots \ y^N]^T$  表示，請問如何以  $X$  和  $\mathbf{y}$  表示可以最小化損失函數的向量  $\mathbf{w}$ ？請選出正確答案。(其中  $X^T X$  為 invertible)

(c)  $(X^T X)^{-1} \mathbf{y}$