

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

Model 1:

public RMSE: 5.63862

private RMSE: 7.21873

Model 2:

public RMSE: 5.90263

private RMSE: 7.22356

由以上紀錄的誤差值可以發現，model 1 無論是在 public 或是 private 都取得了較好的分數，所以我們可以推論影響第 10 小時的 PM2.5 的濃度的污染源可能不只 PM2.5 一種，其他的污染源也會影響，因此只考慮了 PM2.5 一種 feature 的 model 2 的誤差會較大，而將全部 18 種污染源都納入考慮的 model 1 會得到較準確的預測結果。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

Model 1:

public RMSE: 5.97997

private RMSE: 7.17131

Model 2:

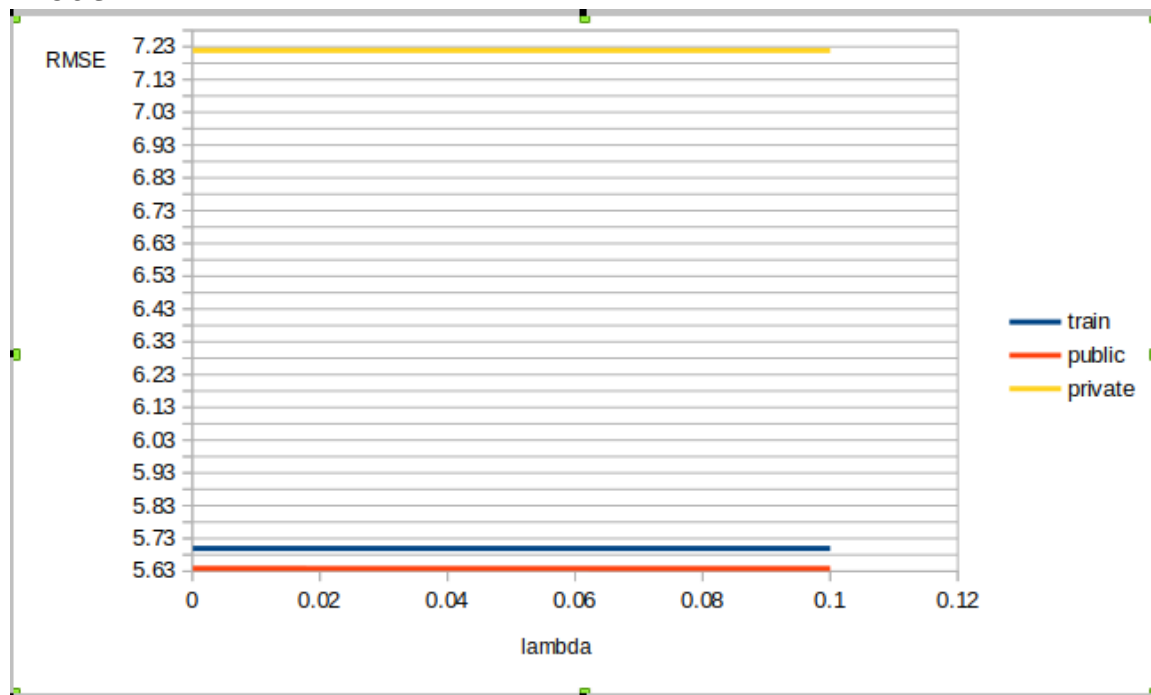
public RMSE: 6.22732

private RMSE: 7.22552

與第 1 題的結果比較可以發現若將 feature 抽取的時間減少為前 5 個小時，大致上兩個 model 的準確率都會下降，因此我們可以推測若將 feature 抽取的時間延長，可能可以有效降低誤差，得到更準確的預測結果。

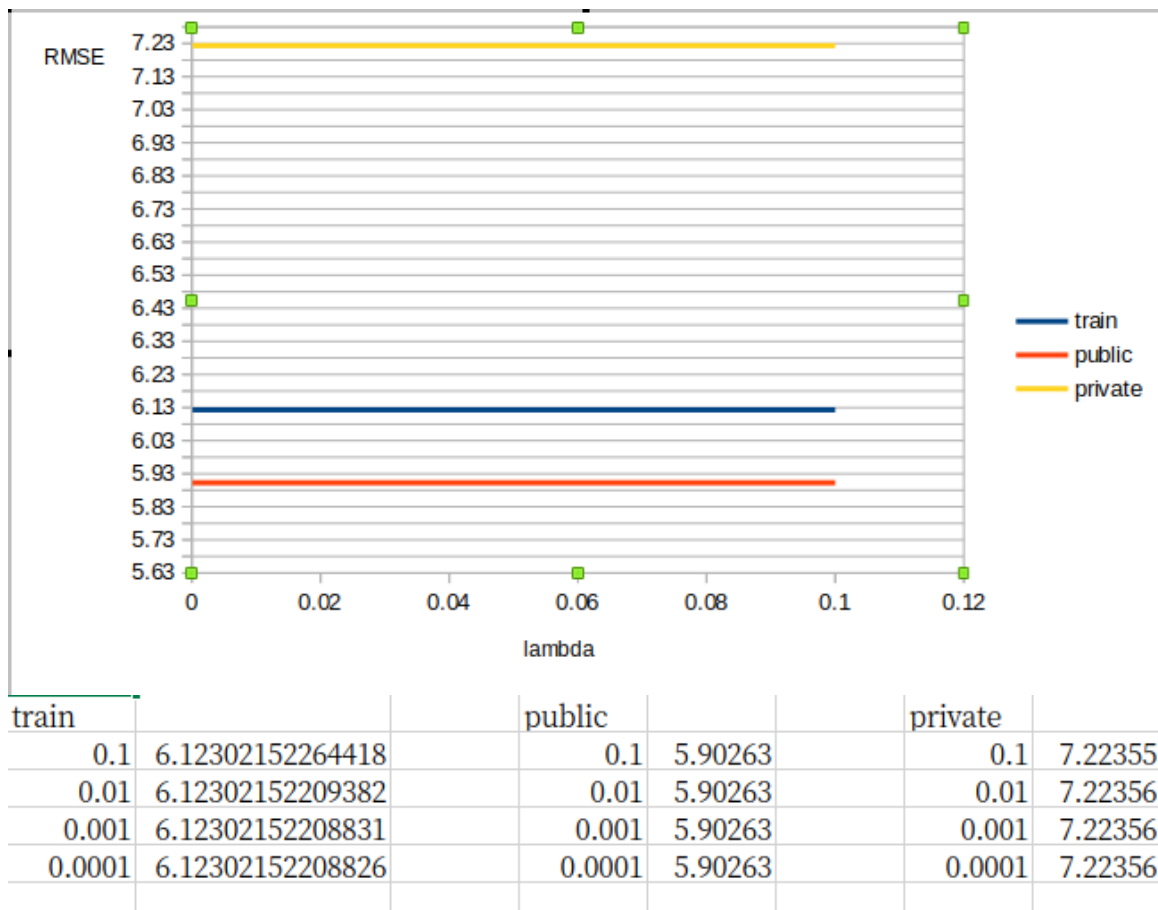
3. (1%)Regularization on all the weight with $\lambda=0.1$ 、 0.01 、 0.001 、 0.0001 ，並作圖

Model 1:



train		public		private	
lambda	RMSE				
0.1	5.69955815586152	0.1	5.63848	0.1	7.2185
0.01	5.69955883244734	0.01	5.6386	0.01	7.2187
0.001	5.69955890851874	0.001	5.63862	0.001	7.21872
0.0001	5.69955891621135	0.0001	5.63862	0.0001	7.21873

Model 2:



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 $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^N]^T$ 表示，所有訓練資料的標註以向量 $\mathbf{y} = [y^1 y^2 \dots y^N]^T$ 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) $(\mathbf{X}^T \mathbf{X}) \mathbf{y} \mathbf{X}^T$
- (c) $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$
- (d) $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{y} \mathbf{X}^T$

Ans: (c)