

學號：B04901069 系級：電機四 姓名：林志皓

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

RMSE of model(1): 5.77025

RMSE of model(2): 5.90274

model(1) 使用所有資料作為**feature**,...而其中有些**feature**與**pm2.5** 的值有正相關，因此有較低的**RMSE**

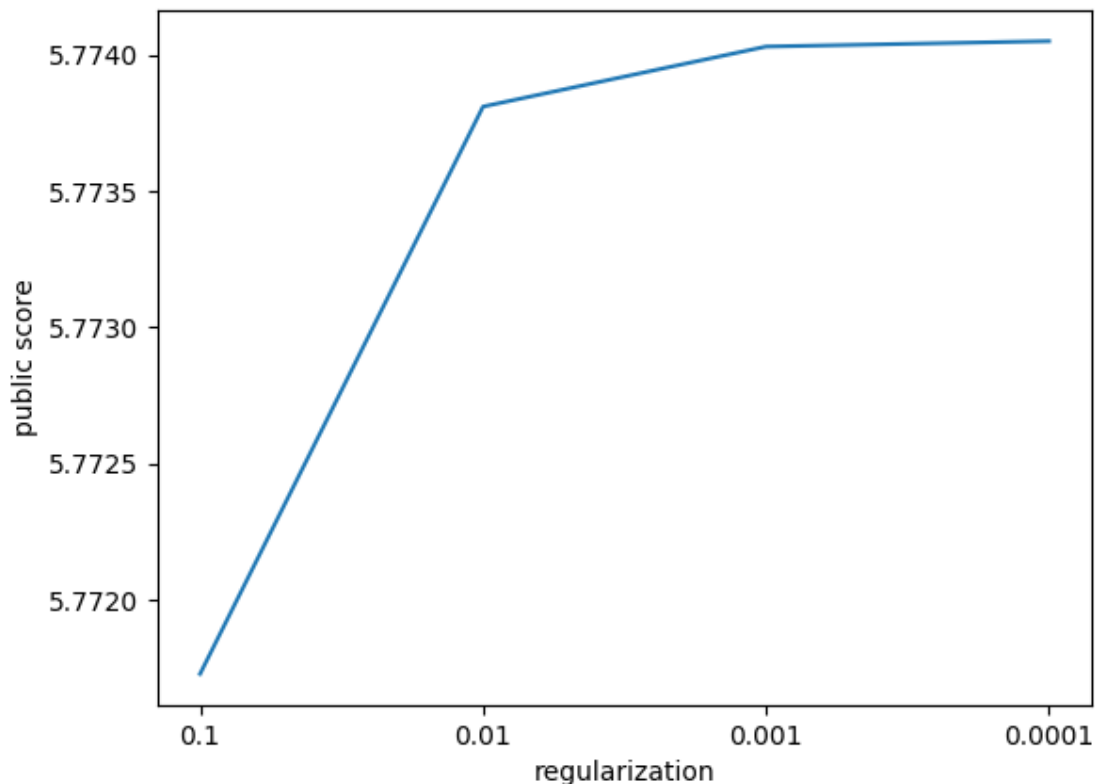
model(2)單純只使用前9天的**pm2.5**值進行訓練，雖然資料量少，但卻是最相關的**feature**，因此**RMSE**值雖比較高，卻也還在足以過**simple baseline**

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

RMSE 為5.99210,

較為不准。推測可能是因為參考9小時比5小時更能看出趨勢變化，做出更好的預測。

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



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

答：(C)