

學號：R05525096 系級：工海碩二 姓名：郭捷

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

Feature	public	private	public+private
全部	7.46237	5.53562	6.57001
PM2.5	7.44013	5.62719	6.59624

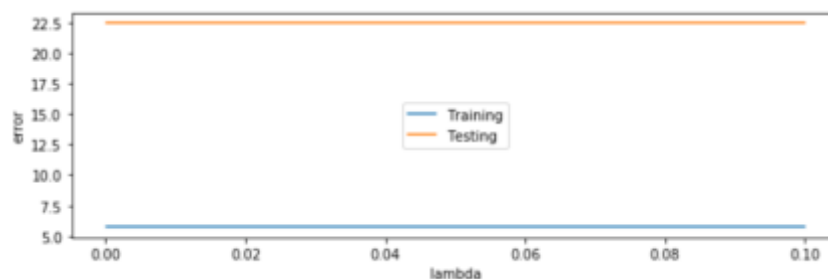
討論：根據所有的 feature 之間的皮爾森係數可見其中有些 feature 和 PM2.5 的相關度很低，如果把這些 feature 也拿去 train 可能會影響結果的準確度。但是根據 public 和 private 的分數可見相差度並不是很大，可能是因為 SGD 有挑選 feature 的能力，對於比較重要的 PM2.5 的 feature 會給予比較高的權重。

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

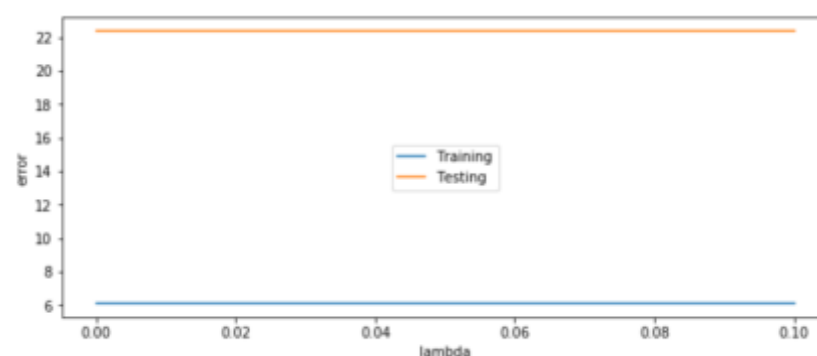
Feature	public	private	public+private
全部	7.65925	5.44092	6.64333
PM2.5	7.57904	5.79187	6.74491

討論：將 9 小時抽成 5 小時，全部 feature 在 private 的分數會高於 9 小時。但是總體來說，9 小時的表現好於 5 小時。

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



(1)All Feature:



(2)Only PM2.5

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)  $(X^T X)X^T y$

證明:

$$\begin{aligned} \text{LossFunction} &= \sum_{n=1}^N (y^n - x^n \cdot w)^2 \\ \text{對 } w \text{ 偏微分, } \frac{\partial \text{loss}}{\partial w} &= -2 \sum_{n=1}^N (y^n - x^n \cdot w) x^n = 0 \\ \sum_{n=1}^N x^n y^n &= w \sum_{n=1}^N (x^n)^2 \\ w &= \frac{\sum_{n=1}^N x^n y^n}{\sum_{n=1}^N (x^n)^2} \\ w &= \frac{X^T y}{X^T X} = (X^T X)^{-1} X^T y \end{aligned}$$

(b)  $(X^T X)^{-1} X^T y$

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

(d)  $(X^T X)^{-2} X^T y$

**Ans (c)**