

學號：R06944034 系級：網媒碩一 姓名：黃禹程

請實做以下兩種不同 feature 的模型，回答第 (1) ~ (3) 題：

抽全部 9 小時內的污染源 feature 的一次項(加 bias)

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

	9 hours all feature	9 hours pm 2.5
RMSE	6.865409627	6.5962795224

所有 feature 的版本比只使用 PM2.5 的版本較差。

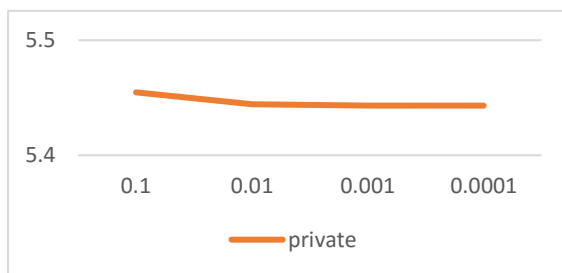
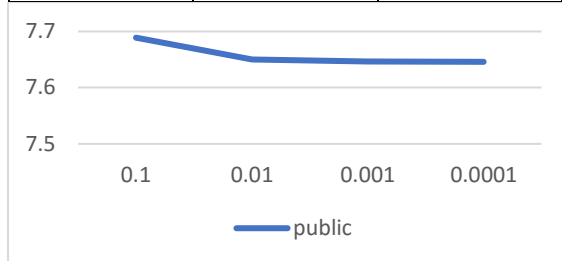
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

	5 hours all feature	5 hours pm 2.5
RMSE	6.7019210574	6.7449136857

使用 5 小時的所有 feature 的版本，比 9 小時還要好
但 pm2.5 比 9 小時版本差

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

λ	0.1	0.01	0.001	0.0001
Public	7.68897	7.65026	7.64638	7.64599
Private	5.45470	5.44426	5.44324	5.44314



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)

- $(X^T X) X^T y$
- $(X^T X)^{-1} X^T y$
- $(X^T X)^{-1} X^T y$
- $(X^T X)^2 X^T y$

(C)

最小化損失函數

$$\Rightarrow y - Xw = 0$$

$$\Rightarrow \langle y - Xw, b \rangle = 0, \forall b \in \mathbb{R}^N$$

\therefore 所有 b 為 Real number.

$$\therefore \text{令 } b = Xw$$

$$\Rightarrow \langle y - Xw, Xw \rangle = 0$$

$$\Rightarrow \langle y, Xw \rangle - \langle Xw, Xw \rangle = 0$$

$$\Rightarrow \langle y, Xw \rangle = \langle Xw, Xw \rangle$$

$$\Rightarrow (Xw)^T y = (Xw)^T Xw$$

$$\Rightarrow w^T X^T y = w^T X^T Xw$$

$$\Rightarrow X^T y = X^T Xw$$

$$\Rightarrow w = (X^T X)^{-1} X^T y$$