

學號：B03505031 系級：工科四 姓名：邱昱軒

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) · 討論兩種 feature 的影響

全部	9 小時	RMSE = 6.39087203617
PM2.5	9 小時	RMSE = 6.43376586149

在這兩組數據中，抽全部污染原的 RMSE 較小。由於 PM2.5 的濃度與其他污染源及降雨等參數有關，且若只抽取 PM2.5 當作 feature，bias 會比較大。因此抽取全部污染源的 feature 來 train，誤差較小是合理的。

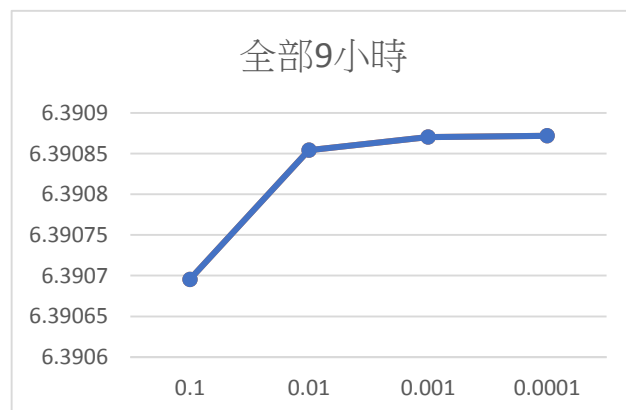
2. (1%)將 feature 從抽前 9 小時改成抽前 5 小時，討論其變化

	9 小時 RMSE	5 小時 RMSE	變化
全部	6.58340081617	6.39087203617	進步 0.193
PM2.5	6.59194162446	6.43376586149	進步 0.148

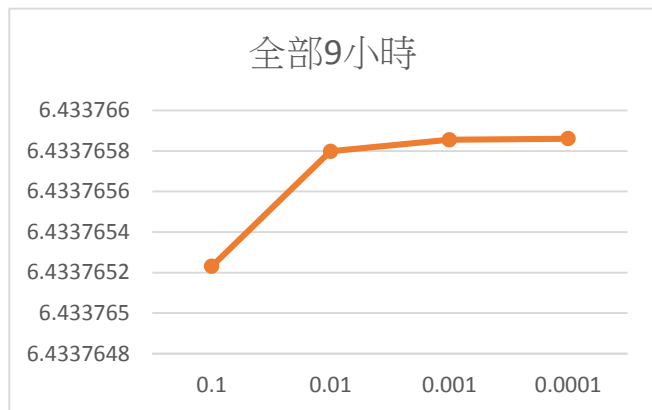
將 feature 改為 5 小時後，誤差皆變小。推測是因為太久以前的數據與要預測的數值關係較小。若將以前的數據拿來 train，反而會造成 overfitting train data 的反效果。

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

全部九小時	
λ	RMSE
0.1	6.39069527023
0.01	6.39085419644
0.001	6.39087025055
0.0001	6.39087185759



PM2.5 九小時	
λ	RMSE
0.1	6.43376523089
0.01	6.43376579837
0.001	6.43376585518
0.0001	6.43376586086



因為原本的 model 使用到一些與結果相關係數較小的 feature，所以該 model 的 variance 較大，若加上 lamda 則可以將 model 平滑化。

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。

- (a) $(X^T X) X^T y$
- (b) $(X^T X)^{-0} X^T y$
- (c) $(X^T X)^{-1} X^T y$
- (d) $(X^T X)^{-2} X^T y$

$$L = (Y - XW)^T (Y - XW)$$

$$\frac{\partial L}{\partial W} = 2X^T(Y - XW) = 2X^TY - 2X^TXW$$

$$\text{let } \frac{\partial L}{\partial W} = 0, \quad 2X^TY = 2X^TXW \quad W = (X^TX)^{-1} X^TY$$

答案為 C