

學號：R06922116 系級：資工碩一 姓名：賴柏恩

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

(1)kaggle public: 7.83378 private: 5.50413 RMSE(in training set): 5.701671

(2)kaggle public: 7.44013 private: 5.62719 RMSE(in training set): 6.123022

在考慮所有 **feature** 一次項的 **model** 裡，我們可以看出他比只取 **pm2.5** 的一次項的 **model** 更加的 **fit train data**，但是在 **test data** 中單取 **pm2.5** 的 **model** 分數卻沒有比考慮所有 **feature** 的 **model** 差，應該是因為考慮所有 **feature** 的 **model** 會對 **train data** 造成 **overfit** 的作用，導致在 **test data** 中反而無法取得相對更好的成績。

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

(1)kaggle public: 7.66477 private: 5.32990 RMSE(in train data): 5.816742

(2)kaggle public: 7.57904 private: 5.79187 RMSE(in train data): 6.207004

在(1)中，雖然只取前 5 小時的 **model** 在 **train data** 中並沒有比 9 小時來得更 **fit** 但是在 **test data** 中卻表現比 9 小時更加合適，因此我們可以推測在 5 小時的 **model** 反而比 9 小時的更加 **fit** 整體的 **model**。

在(2)中，在取 5 小時的 **model** 中在所有的表現中都比 9 小時的 **model** 差，因此可知道 9 小時的 **model** 比 5 小時的更可以代表此 **data**。

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

(1)

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**)為一存量  $y^n$ ，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  $\sum_{n=1}^N (\hat{y}^n - y^n)^2$ 。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  $\mathbf{w}$ ？請寫下算式並選出正確答案。(其中  $\mathbf{X}^T \mathbf{X}$  為 **invertible**)

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

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

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

Ans:  $Xw = Y$

$$X^T X w = X^T Y$$

$$W = (X^T X)^{-1} X^T Y$$

(c)