

# Machine Learning 2017 Fall, Homework 1

## DUE DATE: OCTOBER 13, 2017

系級: 資工四 學號: b03902129 姓名: 陳鵬宇

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

$$(1) \text{ Public Score} = 7.46631, \text{ Private Score} = 5.30105 \Rightarrow \sqrt{\frac{7.46631^2 + 5.30105^2}{2}} \approx 6.47483$$

$$(2) \text{ Public Score} = 7.44013, \text{ Private Score} = 5.62719 \Rightarrow \sqrt{\frac{7.44013^2 + 5.62719^2}{2}} \approx 6.59624$$

可以看出，不論是抽全部污染源的 feature，或是只抽 pm2.5 的 feature，結果都非常接近，由此我們可以猜測，單單 pm2.5 項來當 training data 就可以預測地滿準的，但是 (1) 的 RMSE 仍比 (2) 小些，所以也可以知道，其它 17 項 feature 還是跟第 10 小時的 pm2.5 有關，不然 (1) 的 RMSE 應該會變大。

在考慮所有污染源時，Private 的準確度遠優於 Public，在寫作業時，不太知道 Private 和 Public 的關係都在衝刺排名，當參考了一些 paper，知道 PM2.5 主要是受到 NO<sub>2</sub>, SO<sub>2</sub>, O<sub>3</sub> 和 PM2.5 的影響。

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

$$(1) \text{ Public Score} = 7.66477, \text{ Private Score} = 5.32990 \Rightarrow \sqrt{\frac{7.66477^2 + 5.32990^2}{2}} \approx 6.60138$$

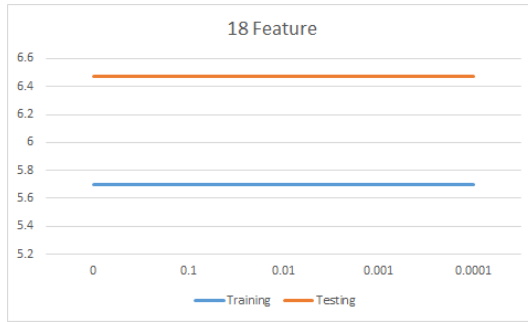
$$(2) \text{ Public Score} = 7.57904, \text{ Private Score} = 5.79187 \Rightarrow \sqrt{\frac{7.57904^2 + 5.79187^2}{2}} \approx 6.74491$$

將前 9 小時改成只抽前 5 小時，(1) 和 (2) 的 RMSE 都變大了，由此我們可以知道，連續的 9 小時預測的會比連續 5 小時要再精準一些。

代表隨著維度的增加，是可以提高精準度的，但提升的幅度不大，可能也代表連續 5 小時的 training data，已經是有一定程度的準確度的。

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

(1)



(2)



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) 為一純量  $y^n$ ，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)  $(X^T X)X^T y$
- (b)  $(X^T X)^{-0} X^T y$
- (c)  $(X^T X)^{-1} X^T y$
- (d)  $(X^T X)^{-2} X^T y$

Loss function  $L$ :

$$L = \sum_{i=1}^N (y_i - x_i w)^2 = (y - Xw)^2$$

求  $L$  最小值，經過微分 (對  $w$ ) 可得：

$$\frac{\partial L}{\partial w} = 2X^T(y - Xw) = 0$$

$$2X^T Xw = 2X^T Y$$

$\because X^T X$  is invertible  $\therefore w = (X^T X)^{-1} X^T y$

故選 (c)