

# Machine Learning Homework Report

## PM2.5 Predict

學號：r05922145 系級：資工碩二 姓名：郁錦濤

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

model	training data RMSE	kaggle public RMSE	kaggle private RMSE
(1)	5.698564	7.53453	5.39402
(2)	6.226667	7.85510	5.87430

解答：

模型 (1) 抽取全部 9 小時內的污染源 feature 的一次項 (加 bias)，共有  $12 \times (480 - 9) = 5652$

筆測資，每筆測資則有  $18 \times 9 = 162$  個 feature。

模型 (2) 抽取全部 9 小時內的 pm2.5 的一次項 (加 bias)，共有  $12 \times (480 - 9) = 5652$  筆測資，

每筆測資則有 9 個 feature。

模型（1）抽取的 feature 比較多，所以在 RMSE 表現上好於模型（2），但是計算速度很慢於模型（2）。相反，模型（2）抽取的 feature 比較少，雖然在 RMSE 的表現上比模型（1）差，但是計算速度比較快。

根據兩種模型的對比，發現選取特徵較多的模型一般比選取模型較少的 performance 更好。但也不是一定的，我的模型只選取 18 類 feature 中的 10 類，最後的效果比選取全部 feature 的模型還要好。

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

model	(1)Training	(1) Public	(1) Private	(2)Training	(2)Public	(2)Private
	RMSE	RMSE	RMSE	RMSE	RMSE	RMSE
前 9 小時	5.698564	7.53453	5.39402	6.226667	7.85510	5.87430
前 5 小時	5.823100	7.71000	5.39192	6.340770	8.08842	6.03934

解答：

將 feature 從抽前 9 小時改成抽前 5 小時，這樣每個測資的 feature 數會由原來的 162 個減少到 90 個，這樣造成的結果是 public set 和 private set 的 RMSE 會變大，也就是 performance 不好。

但是在抽取前 9 小時和前 5 小時的全部 feature 的兩種模型中，抽取前 5 小時全部 feature 模型的 public set 的 RMSE 雖然沒有前 9 個小時的模型好，但是在 private set 的 RMSE 則好於抽取前 9 個小時的模型。這說明，在某些情況下，選取特定的 feature 會有更好的 performance，每個 feature 的權重是不一樣的。

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

解答：

My Model:

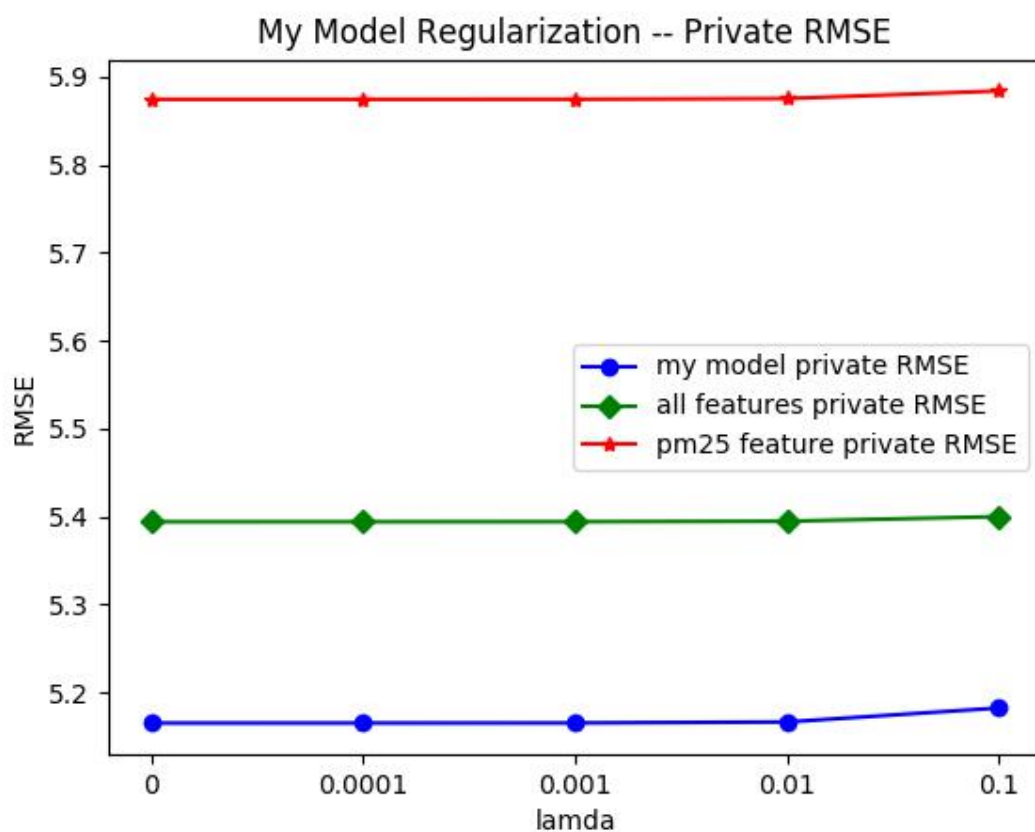
$\lambda =$	0.1	0.01	0.001	0.0001	0
Train RMSE:	5.687335	5.682595	5.682388	5.682372	5.682370
Public RMSE:	6.50708	6.49006	6.48836	6.48820	6.48818
Private RMSE:	5.18226	5.16634	5.16522	5.16512	5.16510

Model(1):

$\lambda =$	0.1	0.01	0.001	0.0001	0
Train RMSE:	5.701145	5.698791	5.698584	5.698564	5.698562
Public RMSE:	7.55890	7.53694	7.53475	7.53453	7.53450
Private RMSE:	5.39984	5.39458	5.39407	5.39402	5.39401

Model(2):

$\lambda =$	0.1	0.01	0.001	0.0001	0
Train RMSE:	6.230814	6.227074	6.226704	6.226667	6.226663
Public RMSE:	7.87279	7.85685	7.85526	7.85510	7.85508
Private RMSE:	5.88383	5.87516	5.87437	5.87430	5.87429



從測資中可以看出，lamda 對於模型(1)和(2)的兩種 feature 選取的一次方的模型最後 public RMSE 和 private RMSE 結果影響不大，對於我的二次模型，當 lamda 取 0.1 和 0.01 時，private RMSE 會從 5.18226 降至 5.16634。

分析如下。由於 regularization 所添加的是  $\lambda \cdot w^2$ ，這個對於選取 feature 的一次方的模型 performance 是不明顯的。但是對於有選取 feature 的二次方的模型是有幫助的，能夠降低 RMSE，當 lamda 越小時，RMSE 表現更好。

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。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)  $(X^T X)^{-2} X^T y$

解答：選擇(c)。

理由：

在取出特徵後，我們的預測結果為  $Xw$ ，因此我們想找到  $w$  使  $(y - Xw)^T (y - Xw)$  最小(也就是讓我們定義的 loss function 最小)，將 loss function 對  $w$  做偏微分後得到  $-2X^T(y - Xw)$ 。令其等於 0，得到  $w = (X^T X)^{-1} X^T y$ 。