

# Machine Learning (2017,Fall)

## Assignment 1 - Predicting PM2.5

學號：b03901086 系級：電機四 姓名：楊正彥

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)

1 (2%)記錄誤差值 (RMSE) (根據 kaggle public+private 分數)，討論兩種 feature 的影響：

本題實作的兩個模型使用的相關參數與訓練方法為：iteration = 12000、learning rate = 0.01 以及對從 5760 筆中隨機取出 4000 筆 training sets 以 *adagrad* 的方式實作 linear regression：

	Public Score	Private Score	RMSE
取前 9 小時內所有污染源的 一次項	6.89520	6.05058	6.48665
取前 9 小時內 PM2.5 的一次項	6.92350	6.08831	6.51929

根據 kaggle 上的 public score 以及 private score 可以發現，取前 9 小時內 PM2.5 一次項的模型（共 9 個 features）performance 比取前 9 小時內所有污染源一次項的模型（共 162 個 features）要來的差，可以推測在一定範圍內，選用較多的 features 將能得到較準確的模型。

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

本題實作的兩個模型除了 features 的選用之外，使用的相關參數與訓練方法皆與前一題相同：

	Public Score	Private Score	RMSE
取前 5 小時內所有污染源的 一次項	6.89026	5.57072	6.26533
取前 5 小時內 PM2.5 的一次項	6.96678	5.89037	6.46728

根據 kaggle 上的 public score 以及 private score 可以發現，只取 PM2.5 作為 features 的模型仍然在 performance 不如取所有污染源當 features 的模型，其根本原因可能與上題類似，仍和 features 多寡有高度關聯。

除此之外，與上題結果比較亦可以發現僅取前 5 小時污然源資料當作測資的模型皆比取前 9 小時污然源資料當作測資的模型要來的準確。

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

本題實作的三個模型除了  $\lambda$  的值之外，使用的訓練方法與前一題相同，而 features 改為 O3、PM2.5、SO2 以及與風有關的四個指數（同 hw1\_best.py）：

	Public Score	Private Score	RMSE
$\lambda = 0.1$	6.70703	5.30433	6.04649
$\lambda = 0.01$	6.74885	5.31950	6.07634
$\lambda = 0.001$	6.74922	5.31972	6.07665

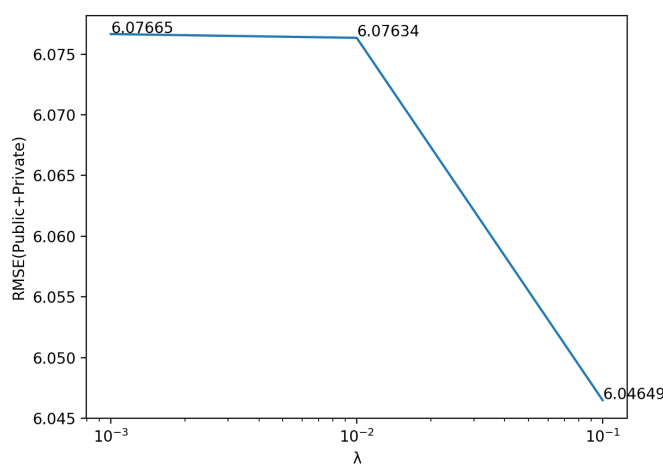


Fig 1 : Regularization on all weight with different  $\lambda$

從結果可以發現  $\lambda = 0.1$  的模型會比  $\lambda = 0.001$  的模型 performance 要來得好一點點，但是效果沒有想像中顯著，最後也於 hw1\_best.py 中選用  $\lambda=2.5$  來做訓練。

4 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  $w$ ？

$$\sum_{n=1}^N (y^n - x^n \cdot w)^2 = (y^1 - x^1 \cdot w \ \dots \ y^N - x^N \cdot w) \cdot \begin{pmatrix} y^1 - x^1 \cdot w \\ \vdots \\ y^1 - x^1 \cdot w \end{pmatrix} = (y - X \cdot w)^T (y - X \cdot w)$$

$$\frac{\partial}{\partial w} (y - X \cdot w)^T (y - X \cdot w) = \frac{\partial}{\partial w} (y^T y - y^T X w - w^T X^T y + w^T X^T X w) = 0$$

$$-2X^T (y - X \cdot w) = 0$$

$$X^T X \cdot w = X^T y$$

$$w = (X^T X)^{-1} X^T y \quad \therefore \text{選 (C)}$$