

ML2017FALL HW1 Report

學號：B04902090 系級：資工三 姓名：施長元

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) 都可以用

個人的訓練狀況：取 15 萬次，初始 Training Rate = 0.5，使用 Adagrad，捨去七月的資料，前兩題 $\lambda = 0$

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

	Training Loss	(RMSE) : Kaggle Public+Private
All feature	5.424189	$5.24088 + 7.40240 = 12.64328$
Only PM 2.5	5.852992	$5.62131 + 7.26272 = 12.88403$

討論：

全部的污染源中，總有幾項污染源會跟 PM 2.5 有相關性(例如雨量)，但也有不相關的污染源，今天將他們全部放入 feature，可以見得，有相關性的比不相關的影響上高了一些，造成 All feature 比單 PM2.5 的資料來得準確。

2. (1%)將feature從抽前9小時改成抽前5小時，討論其變化

	Training	(RMSE) : Kaggle Public+Private
All feature	5.544544	$5.30421 + 7.61738 = 12.92159$
Only PM 2.5	5.931668	$5.79114 + 7.41670 = 13.20784$

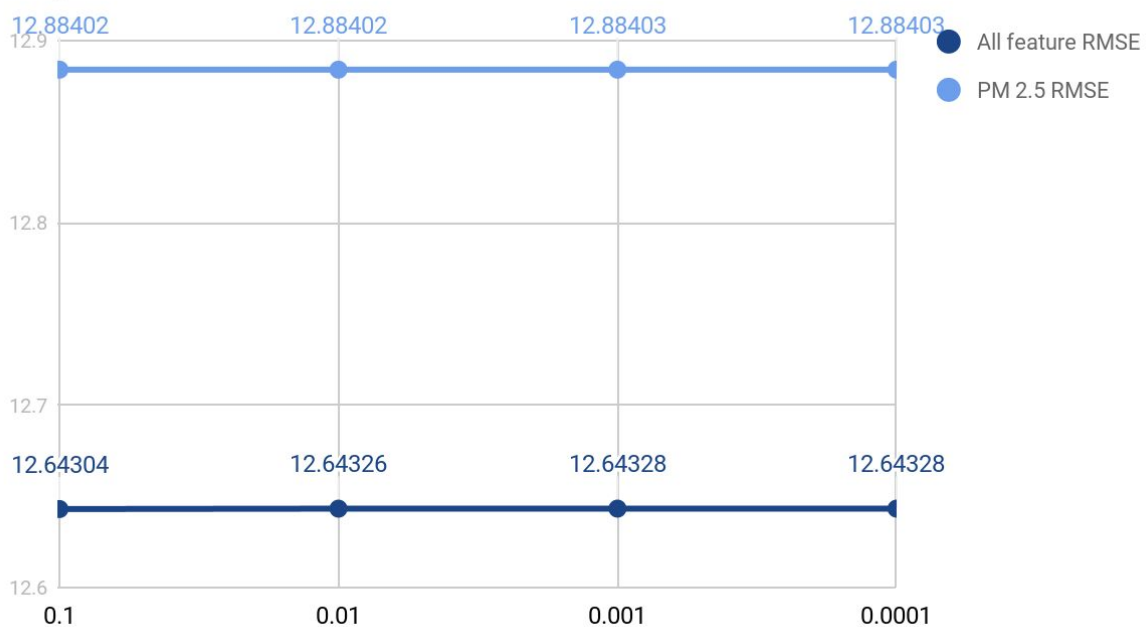
討論：

五小時的資料代表時間連續性的因素相較九小時降低，因此在全部污染源以及 PM 2.5 都有正確率下降的趨勢。

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

	All feature		Only PM 2.5	
λ	training loss	RMSE	training loss	RMSE
0.1	5.424187	5.24091 +7.40213 = 12.64304	5.852992	5.62132 +7.26270 =12.88402
0.01	5.424189	5.24089 +7.40237 =12.64326	5.852992	5.62131 +7.26271 =12.88402
0.001	5.424189	5.24088 +7.40240 =12.64328	5.852992	5.62131 +7.26272 =12.88403
0.0001	5.424189	5.24088 +7.40240 =12.64328	5.852992	5.62131 +7.26272 =12.88403

Regularization



討論：

Regularization 似乎沒有造成太大的影響

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

Ans : C

$$\begin{aligned} \text{cost} &= \sum_{n=1}^N (y^n - x^n \cdot w)^2 = \|y^T - X^T w\|^2 \\ \frac{\partial}{\partial w} \|y^T - X^T w\|^2 &= 2X(y^T - X^T w) \quad \therefore f=0 \\ w &= (X^T X)^{-1} X^T y \end{aligned}$$