

Machine Learning HW1

學號：D05921027 系級：電機博一 姓名：張鈞閔

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：

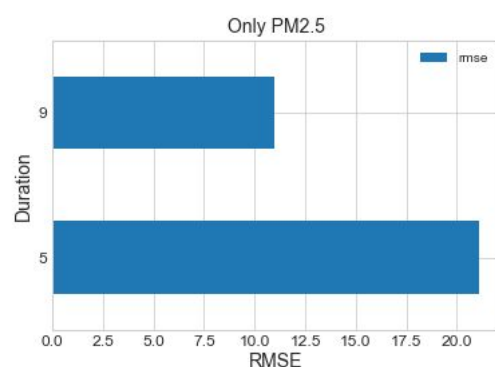
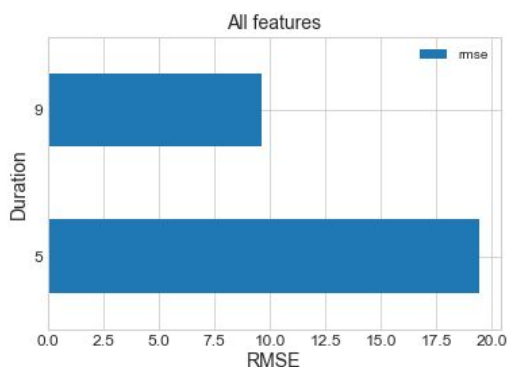
- a. NR請皆設為0，其他的數值不要做任何更動
- b. 所有 advanced gradient descent 技術(如: adam, adagrad 等) 都是可以用的

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

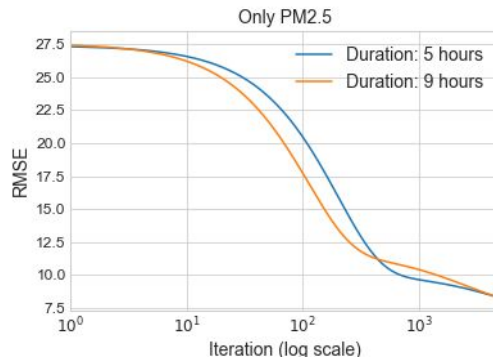
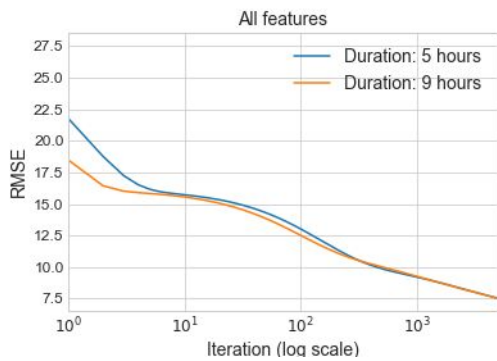
	Public	Private	若使用前 9 小時內的 PM2.5 值，類似於只依照歷史資料猜測現在數值，而忽略了其他環境因素的影響，如：風速、降雨。所以使用全部的 feature 可增加模型預測力改善 public 和 private score
All features	6.4208	5.3253	
Only PM2.5	9.1032	12.5513	

2. (1%)將 features 從抽前9小時改成抽前5小時，討論其變化

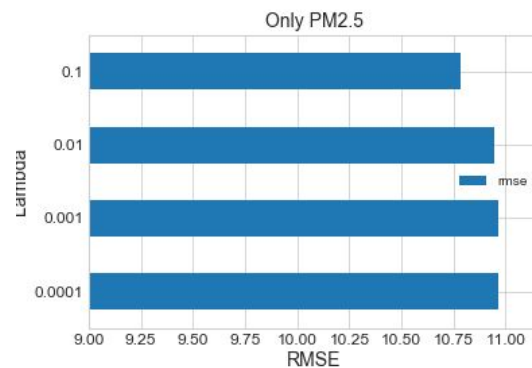
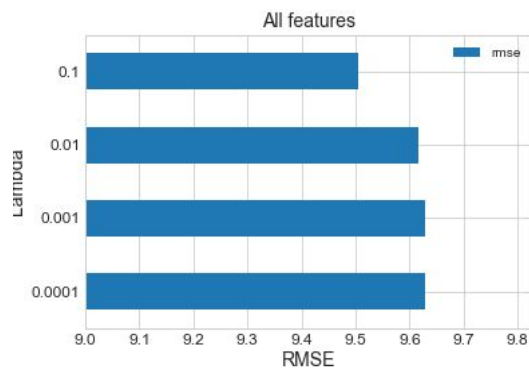
Duration	9 hours	5 hours	若參考 9 小時的 features，可以得到較佳的 performance，預測誤差差異近兩倍。
All features	9.6288	19.444	
Only PM2.5	10.963	21.082	



- Loss changes through iterations: 參考的時間長度越長，loss 下降較快



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



Lambda	0.1	0.01	0.001	0.0001
All features	9.5046	9.6164	9.6275	9.6286
Only PM2.5	10.783	10.946	10.962	10.964

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。(其中 $X^T X$ 為invertible)

- (a) $(X^T X) X^T y$
- (b) $(X^T X)^{-1} X^T y$
- (c) $(X^T X)^{-1} X^T y \Rightarrow$ answer
- (d) $(X^T X)^{-2} X^T y$

The analytical solution to linear regression problem with a least-squares loss function

$$\text{minimize } f(w) = (Xw - y)^T (Xw - y)$$

$$\rightarrow f(w) = ((Xw)^T - y^T)(Xw - y) = (Xw)^T Xw - (Xw)^T y - y^T Xw + y^T y$$

To find the minimum of $f(w)$, we derive by w and compare to 0,

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

Finally, we obtain $w = (X^T X)^{-1} X^T y$.

