

# ML\_HW1\_Report

學號：B05901005 系級：電機三 姓名：賴沂謙

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 當作一次項(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  
我實作了 train 不同 iteration 次數訓練出來的 model , test 後的結果如下 :

iteration	1000	5000	10000	20000	50000	100000
(1)所有 features	17.88262	14.18077	13.46444	13.06063	12.86501	12.85325
(2)只有 pm2.5	13.33782	13.12689	13.12619	13.12619	13.12619	13.12619

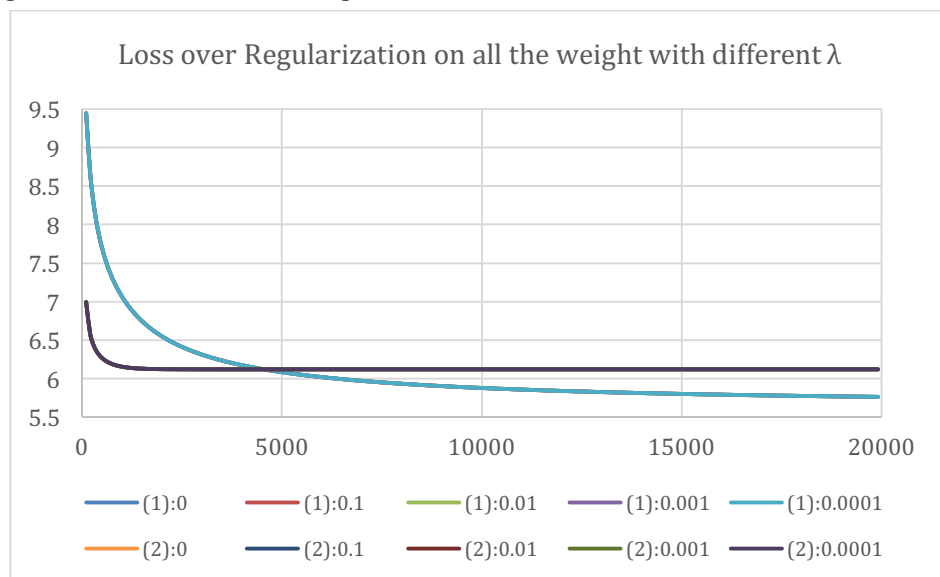
所有 features 都使用的模型會比較晚收斂，而只有 pm2.5 的模型則比較快，這應該是因為 data 量大小的關係。就結果而言，所有 features 的模型，結果較好。

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

iteration	1000	5000	10000	20000	50000	100000
(1)所有 features	15.95824	13.58529	13.27428	13.17327	13.1508	13.14958
(2)只有 pm2.5	13.47326	13.45284	13.45284	13.45284	13.45284	13.45284

把 9 個小時的縮短成 5 個小時後，train 的 data 量更少了，所以收斂的時間也都有變短，但結果就變差了一些。而所有 features 的模型一樣比只有 pm2.5 的模型好。

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



Training Loss at iteration 1000	No regularization	$\lambda = 0.1$	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 0.1$
(1)所有 features	7.059961425	7.059961547	7.059961437	7.059961426	7.059961425
(2)只有 pm2.5	6.15661662	6.156616872	6.156616646	6.156616623	6.156616621

因為我用的模型 feature 的次數都是一次，用了 Regularization 的結果也沒有多大的差別，從 loss 的圖看不太出來不同  $\lambda$  間的差距，所以另外附上一張表，顯示在 iteration 1000 次的時候的 loss，看出某些微的差別。

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  $w$ ？請選出正確答案。(其中  $X^T X$  為 invertible)

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

Ans: (c)