

Homework 1 Report - PM2.5 Prediction

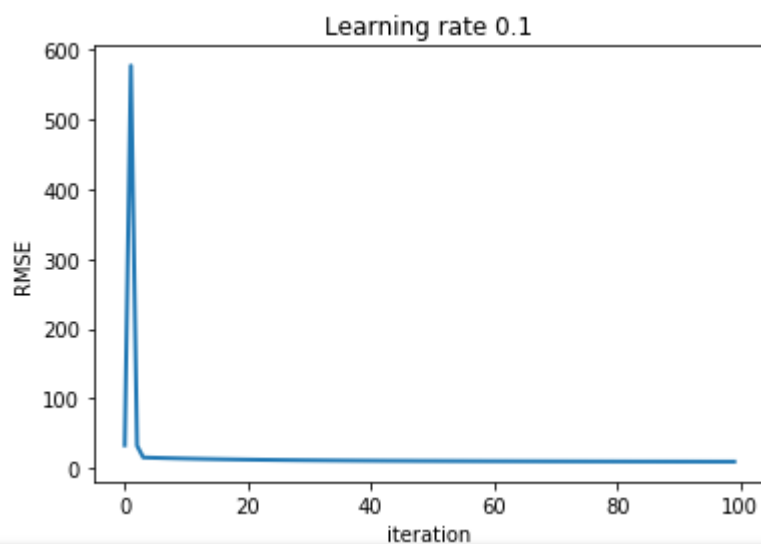
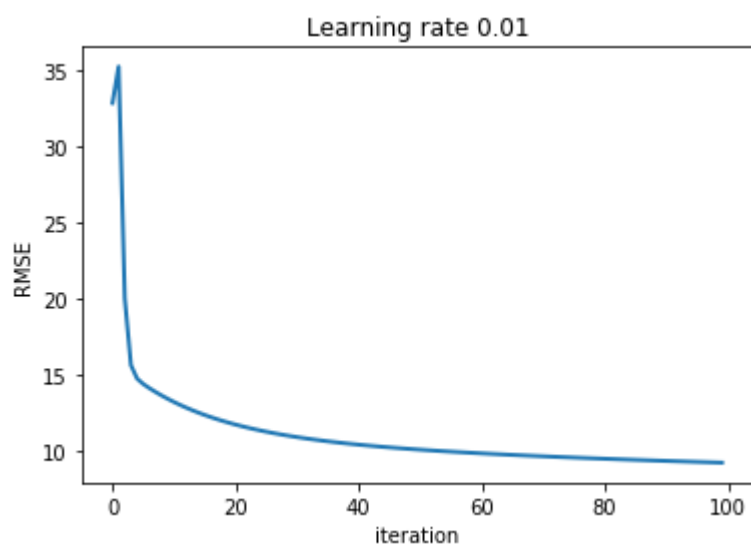
學號：R06922134 系級：資工碩二 姓名: 葉沛陽

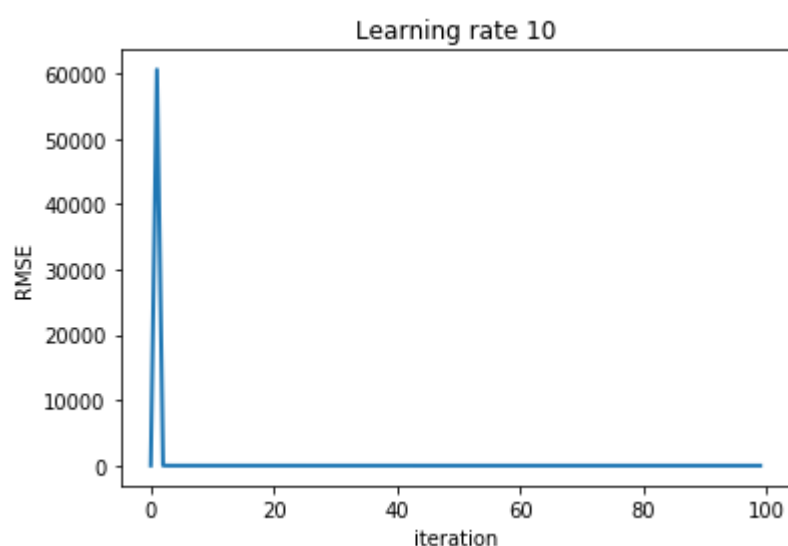
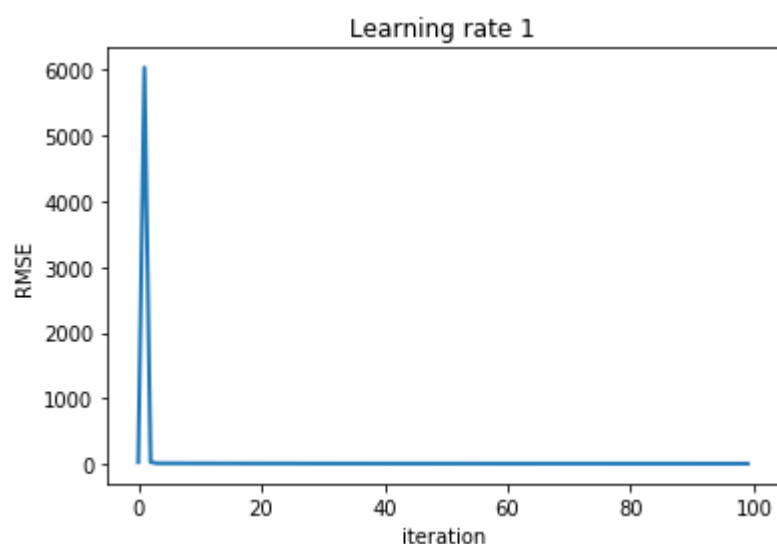
- Report.pdf 檔名錯誤 (-1%)
- 學號系級姓名錯誤 (-0.5%)

1. (1%) 請分別使用至少 **4** 種不同數值的 learning rate 進行 training（其他參數需一致），對其作圖，並且討論其收斂過程差異。

以下為示意圖：

Ans:





Learning rate 越大，在第 2 個 iteration 的 RMSE 會飆越高，不過 4 種 learning rate 在第 100 個 iteration 的時候，分別收斂到{learning rate 10 : RMSE:9.398166};{learning rate 1 : RMSE:9.391467};{learning rate 0.1 : RMSE:9.328023};{learning rate 0.01 : RMSE:9.215379};RMSE 都在 9 左右，只有些微差距，learning rate 越小 在第 100iteration 時 RMSE 越小。

2. (1%) 請分別使用每筆 data9 小時內所有 feature 的一次項（含 bias 項）以及每筆 data9 小時內 PM2.5 的一次項（含 bias 項）進行 training，比較並討論這兩種模型的 root mean-square error（根據 kaggle 上的 public/private score）。

	public	private
所有 feature	5.63946	
PM2.5	6.35282	

只用 PM2.5 雖然直觀上，PM2.5 是跟答案最相關的 feature，不過少了其他那麼多 feature 還是大大的影響了準確率，證明了其他 feature 也是跟 PM2.5 的變化有相關性的。

3. (1%)請分別使用至少四種不同數值的 regularization parameter λ 進行 training（其他參數需一至），討論及討論其 RMSE(training, testing)（testing 根據 kaggle 上的 public/private score）以及參數 weight 的 L2 norm。

Ans:

其他參數

Iteration = 10000 ; Learning rate = 10 ;

λ	RMSE(training)	RMSE(public)	RMSE(private)	Weight L2 norm
0.01	22.603668	9.00778		1.87981517
0.1	22.603668	9.00778		1.87971463
1	22.603671	9.00776		1.87870999
10	22.603703			1.86873505

regularization parameter λ 越大 RMSE(training)有稍微變大 不過 Weight L2 norm 會變小來調控，因此改變不大。

使用 regularization parameter λ 越大 會讓 Weight 的改變對 Loss 的影響放大 所以 Weight L2 norm 會隨著 λ 變大而變小

4~6 (3%) 請參考數學題目（連結：），將作答過程以各種形式（latex 尤佳）清楚地呈現在 pdf 檔中（手寫再拍照也可以，但請注意解析度）。

4.

(4-a)

Ans:

設 x_1, \dots, x_N 是 d 維。

$$E_D(\mathbf{w}) = \frac{1}{2} (r_1(t_1 - (w_1x_{1,1} + \dots + w_dx_{1,d}))^2 + \dots + r_N(t_N - (w_1x_{N,1} + \dots + w_dx_{N,d}))^2)$$

$$\frac{\partial E_D(\mathbf{w})}{\partial w_i} = (r_1(w_1x_{1,1} + \dots + w_dx_{1,d} - t_1)x_{1,i} + \dots + (r_N(w_1x_{N,1} + \dots + w_dx_{N,d} - t_N)x_{N,i})$$

$$\nabla E_D(w) = \begin{bmatrix} \frac{\partial E_D(\mathbf{w})}{\partial w_1} \\ \frac{\partial E_D(\mathbf{w})}{\partial w_2} \\ \vdots \\ \frac{\partial E_D(\mathbf{w})}{\partial w_d} \end{bmatrix} = \begin{bmatrix} r_1x_1 & r_2x_2 & \dots & r_Nx_N \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix} - \begin{bmatrix} r_1x_1 & r_2x_2 & \dots & r_Nx_N \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}$$

令

$$X' = \begin{bmatrix} r_1x_1 \\ r_2x_2 \\ \vdots \\ r_Nx_N \end{bmatrix}, X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}, W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}, T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}$$

For

$$\nabla E_D(W^*) = 0$$

$$(X')^T XW^* = (X')^T T$$

$$W^* = ((X')^T X)^{-1} (X')^T T$$

(4-b)

ANS:

$$W^* = ((X')^T X)^{-1} (X')^T T = \begin{bmatrix} 4 & 5 & 15 \\ 6 & 1 & 18 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 5 & 1 \\ 5 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 4 & 5 & 15 \\ 6 & 1 & 18 \end{bmatrix} \begin{bmatrix} 0 \\ 10 \\ 5 \end{bmatrix}$$

$$= \begin{bmatrix} 108 & 107 \\ 107 & 127 \end{bmatrix}^{-1} \begin{bmatrix} 125 \\ 100 \end{bmatrix} = \begin{bmatrix} \frac{127}{2267} & \frac{-107}{2267} \\ \frac{-107}{2267} & \frac{108}{2267} \end{bmatrix} \begin{bmatrix} 125 \\ 100 \end{bmatrix} = \begin{bmatrix} \frac{5175}{2267} \\ \frac{-2575}{2267} \end{bmatrix}$$