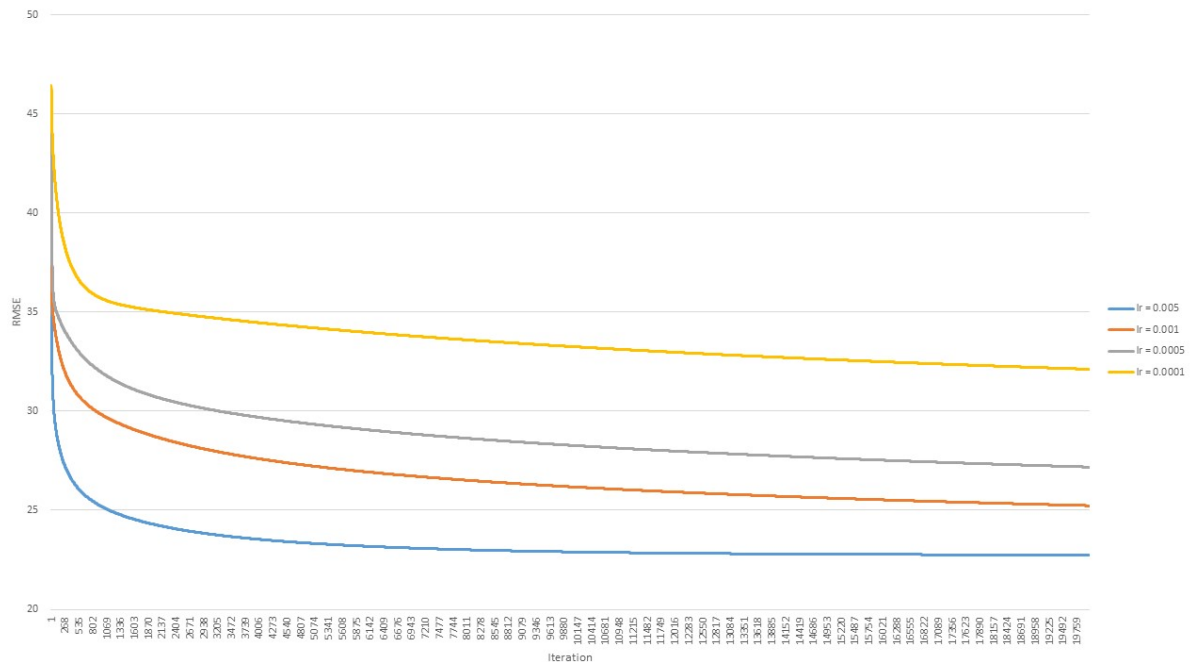


Homework 1 Report - PM2.5 Prediction

學號：R07921078 系級：電機所碩一 姓名：張廷維

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。

以下分別為Learning rate = 0.001/0.005/0.0001/0.0005的收斂過程，使用每筆data 9小時內所有feature的一次項 + bias項並且搭配Adagrad。



1. 當Learning rate較大時，可以較快達到收斂
2. Linear model上不會有local minima，所以在使用gradient descent時，不管Learning rate大小為何RMSE皆會收斂於同一個值
3. 當Iteration無窮大時，可以估計大約會收斂在RMSE=23左右
4. 由於做此題目時並未處理好原始資料的錯誤，因此training算出來的RMSE會與上傳到kaggle有些落差

2. (1%) 請分別使用每筆data 9小時內所有feature的一次項（含bias項）以及每筆data 9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）。

根據題目要求的feature, 利用closed form進行training找出最佳解, 所得結果如下:

每筆data 9小時內所有feature的一次項 + bias : 9.25926

每筆data 9小時內PM2.5的一次項 + bias : 9.47099

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(traning, testing)（testing根據kaggle上的public/private score）以及參數weight的L2 norm。

以下以Learning rate = 0.001 & Iteration = 20000並且使用每筆data 9小時內所有feature的一次項 + bias項進行training。

λ	training error	test error	L2 norm of weight
1	25.2340797412488	10.05929	0.566520594794594
10	25.2455085409669	10.10016	0.46532264745385
100	25.3192194844817	10.17584	0.403773052871745
1000	26.1389138602324	10.7827	0.298577757787156
10000	30.6106860130209	13.28453	0.122589595713393

1. 隨著regularization parameter(λ)的增加，training error也會增加
2. 隨著regularization parameter(λ)的增加，weight的L2 norm越來越小，也就是越來越smooth
3. 以上面圖表的狀態會選擇 $\lambda = 1$ ，並且也可知並不是越平滑(smooth)的function就會越好

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-a)

$$\frac{\partial E_p(w)}{\partial w} = \sum_{n=1}^N r_n (t_n - w^T x_n) \cdot (-x_n) \stackrel{w^*}{=} 0$$

$$\Rightarrow \sum_{n=1}^N r_n t_n x_n = \sum_{n=1}^N r_n w^T x_n x_n = \left(\sum_{n=1}^N r_n x_n x_n^T \right) w^*$$

$$\therefore w^* = \left(\sum_{n=1}^N r_n x_n x_n^T \right)^{-1} \sum_{n=1}^N r_n t_n x_n$$

(4-b)

$$w^* = \left(2 \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 3 \end{bmatrix} + 1 \cdot \begin{bmatrix} 5 \\ 1 \end{bmatrix} \begin{bmatrix} 5 & 1 \end{bmatrix} + 3 \cdot \begin{bmatrix} 5 \\ 6 \end{bmatrix} \begin{bmatrix} 5 & 6 \end{bmatrix} \right)^{-1} \cdot \left(2 \cdot 0 \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} + 1 \cdot 10 \cdot \begin{bmatrix} 5 \\ 1 \end{bmatrix} + 3 \cdot 5 \cdot \begin{bmatrix} 5 \\ 6 \end{bmatrix} \right)$$

$$= \left(2 \cdot \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix} + \begin{bmatrix} 25 & 5 \\ 5 & 1 \end{bmatrix} + 3 \cdot \begin{bmatrix} 25 & 30 \\ 30 & 36 \end{bmatrix} \right)^{-1} \cdot \left(\begin{bmatrix} 50 \\ 10 \end{bmatrix} + \begin{bmatrix} 75 \\ 90 \end{bmatrix} \right)$$

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

$$= \frac{1}{2267} \begin{bmatrix} 5175 \\ -2575 \end{bmatrix}$$