ECE-363 Assignment 1

Lavanya Verma (lavanya18155@iiitd.ac.in)

April 1, 2021

1. Problem: Linear Regression

Part 1:
Plot of Mean R.M.S.E. over 5-folds vs G.D. iterations for train and val set.

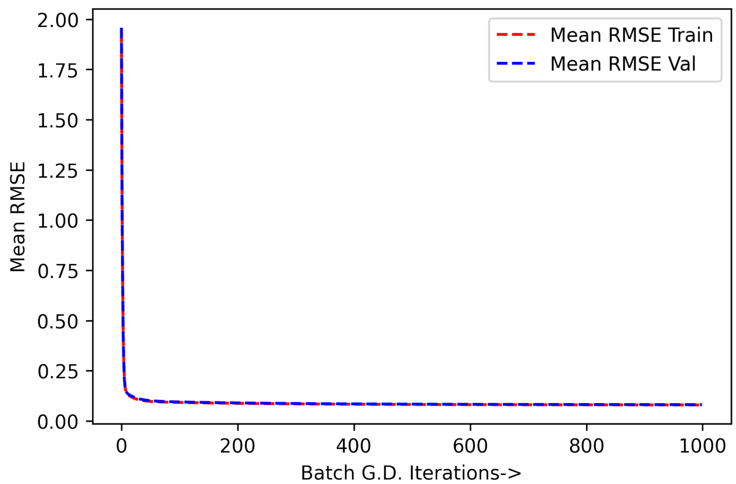

Figure 1: Mean RMSE vs iterations

RMSE computed on least square based linear regression model.

$$\text{Train set} : \begin{bmatrix} 1.6053891719466324 & nan & nan & 1.1367230553375038 & 2.0903646378568403 \end{bmatrix}$$
$$\text{Val set} : \begin{bmatrix} 1.3330744000039038 & nan & nan & 0.9520675553153591 & 1.5580328992910049 \end{bmatrix}$$

In some cases, as for the fold 2 and 3, $X^T X$ conditional number was very high around 10^{17}, so does matrix X was not full rank upto acceptable machine precision(epsilon for IEEE 745 double precision). Therefore, in those cases, inverse of $X^T X$ is not computable.

Comparision: RMSEs computed using least square based regression method are higher than their gradient counterpart even on same data. This seems to be counter intuitive since least square solution guarantees optimality, its reason could be that the approximations done in computation of optimal parameters. Since the data matrix($X^T X$) has very high conditional number, any small change in input(y) will reflect as large change in output(parameters), leading to optimal parameters and same reflected in RMSE.

Part 1.2:
Optimal parameters computed using grid search on grid

$$\{"lam" : [0.01, 0.1, 0.5, 1], "lr" : [1e-5, 1e-4, 1e-3], "reg" : ['l1'], 'epochs' : [500]\}$$

are

$$dict(lam = 1, lr = 0.0001, reg = "l1", epochs = 500)$$

L1 RMSE (test set) : 0.062933

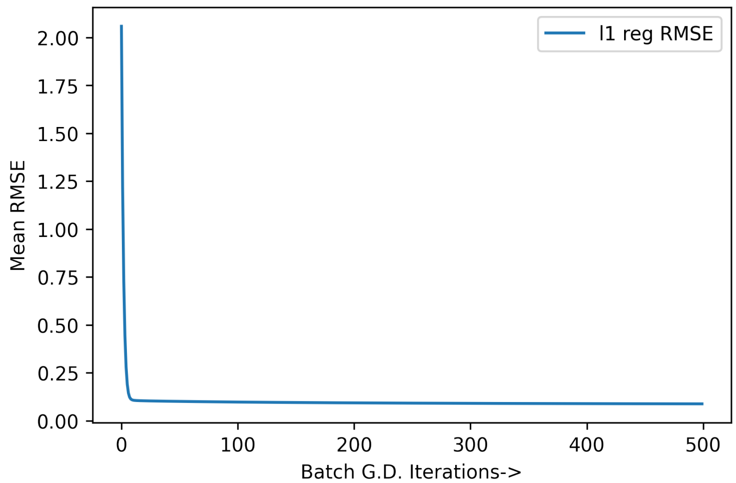

Figure 2: Mean RMSE vs iterations(L1 Reg. Optimal Parameters)

Optimal parameters computed using grid search on grid

$$\{"lam" : [0.01, 0.1, 0.5, 1], "lr" : [1e-6, 1e-4, 1e-3], "reg" : ['l2'], 'epochs' : [500]\}$$

are

$$dict(lam = 0.01, lr = 0.0001, reg = "l2", epochs = 500)$$

L2 RMSE (test set) : 0.065522

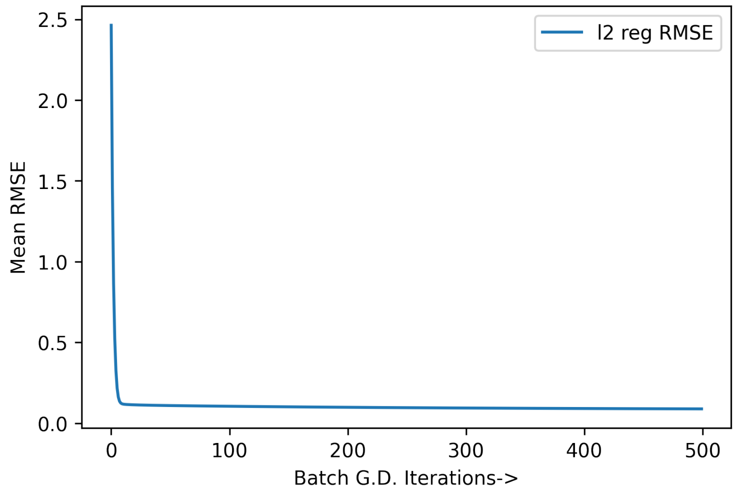

Figure 3: Mean RMSE vs iterations(L2 Reg. Optimal Parameters)

Part 1.3:
L1 regularisation seems to better fit the linear model than L2 and no regularisation.

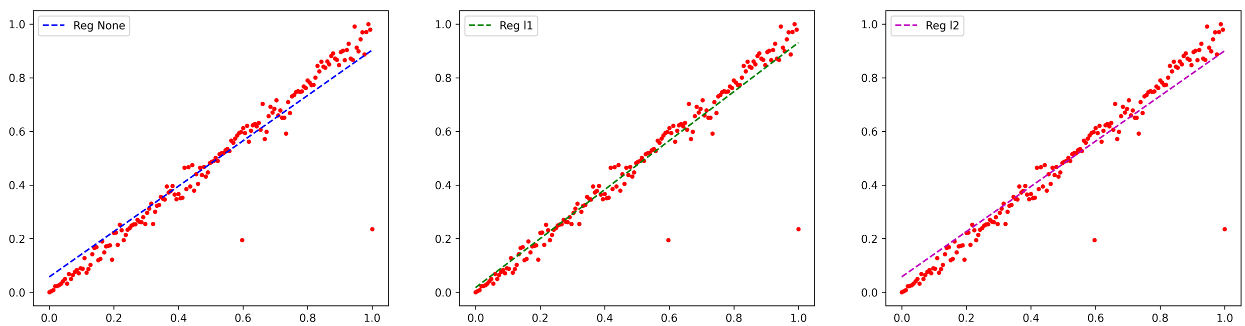

Figure 4: Mean RMSE vs iterations

2. Problem: Theory 1

Linear Model Relation:

$$
\begin{aligned}
&y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \epsilon_i \\
&Y = X\beta + \epsilon
\end{aligned}
$$

where,

$$
Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}; X = \begin{bmatrix} 1 & x_{21} & \dots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2n} & \dots & x_{kn} \end{bmatrix}; \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}; \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}
$$

$$
\begin{aligned}
Loss &= \sum_{i=1}^{n} \epsilon_i^2 \\
&= \begin{bmatrix} \epsilon_1 & \dots & \epsilon_n \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} \\
&= \epsilon^T \epsilon \\
&= (Y - X\beta)^T (Y - X\beta) \\
&= Y^T Y - \beta^T X^T Y - Y^T X\beta + \beta^T X^T X\beta
\end{aligned}
$$

Since,

$$
\begin{aligned}
Y^T X\beta \in \mathbb{R} &\Rightarrow (Y^T X\beta)^T = Y^T X\beta \\
Loss &= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta \\
\frac{\partial Loss}{\partial \beta} &= -2X^T Y + 2X^T X\beta
\end{aligned}
$$

Equating to zero(for extrema)

$$
\begin{aligned}
\frac{\partial Loss}{\partial \beta} &= 0 \\
\Rightarrow X^T X\beta &= X^T Y \\
\frac{\partial^2 Loss}{\partial \beta^2} &= 2X^T X
\end{aligned}
$$

$$
\begin{aligned}
Q(z) &= z^T X^T X z \\
&= (Xz)^T X z \\
&= \|Xz\|_2 \\
\textit{Q(z) is positive definite, if Q(z)} > \textit{0} \ \forall \ z \ &\neq 0 \\
min(n, k) &= rank(X) + dim(Nul\ X) \\
\Rightarrow dim(Nul\ X) &= 0 \\
ker(X) &= \{0\} \\
\Rightarrow Q(z) &> \ 0 \ \forall \ z \notin ker(X) \\
Q(z) &> \ 0 \ \forall \ z \neq 0
\end{aligned}
$$

Since $X^T X$ *is a positive definite matrix* $\forall X \in \mathbb{R}^{nxk}$. *Hence,* $\frac{\partial Loss}{\partial \beta} = 0$ *yields minima. Solution only exists if and only if* $X^T X$ *is full rank* $\Rightarrow$ *X is full rank.*
Optimal Solution:

$$
\beta = (X^T X)^{-1} X^T Y
$$

3. Problem: Theory 2

Assumption:

- Interaction between variable is taken as multiplication of variables

$$
\begin{aligned}
Income &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \\
&= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 \\
&= 50 + 20X_1 + 0.07X_2 + 35X_3 + 0.01X_1X_2 - 10X_1X_3 \\
\textit{for Males, } Income &= 50 + 20X_1 + 0.07X_2 + 0.01X_1X_2 \\
\textit{for Females, } Income &= 85 + 10X_1 + 0.07X_2 + 0.01X_1X_2
\end{aligned}
$$

1. for a fixed G.P.A. x_1 *and IQ* x_2*,*

$$
\begin{aligned}
\textit{(Male Income) } I_1 &= 50 + 20x_1 + 0.07x_2 + 0.01x_1x_2 \\
\textit{(Female Income) } I_2 &= 85 + 10x_1 + 0.07x_2 + 0.01x_1x_2 \\
I_1 == I_2; &(x_1 = 3.5\ ) \\
I_1 > I_2; &(x_1 > 3.5\ ) \\
I_1 < I_2; &(x_1 < 3.5\ )
\end{aligned}
$$

Selected Option:(c)
2. Salary of female IQ 115 GPA 3.5:

$$
\begin{aligned}
I &= 85 + 10 * 3.5 + 0.07 * 115 + 0.01 * 3.5 * 115 \\
I &= 132.075 \textit{ thousand dollars}
\end{aligned}
$$

*3. False, contribution of GPA/IQ term also depends upon GPA*IQ*