Machine Learning
Assignment 4

Name: Girish Chandar G
Roll No: 16110057
Department: Electrical Engineering

Codes for all questions has been coded in single jupyter notebook and headings given accordingly
https://gist.github.com/girish1511/66a90e579a52ca98a47ea6c8e68d316c#file-assignment4-ipynb

Q1. Attached as a continuation to the report after Q5

Q4.
(b)
Θ0= 1.09615385
Θ1=0.19230769
Θ2= 0.38461538

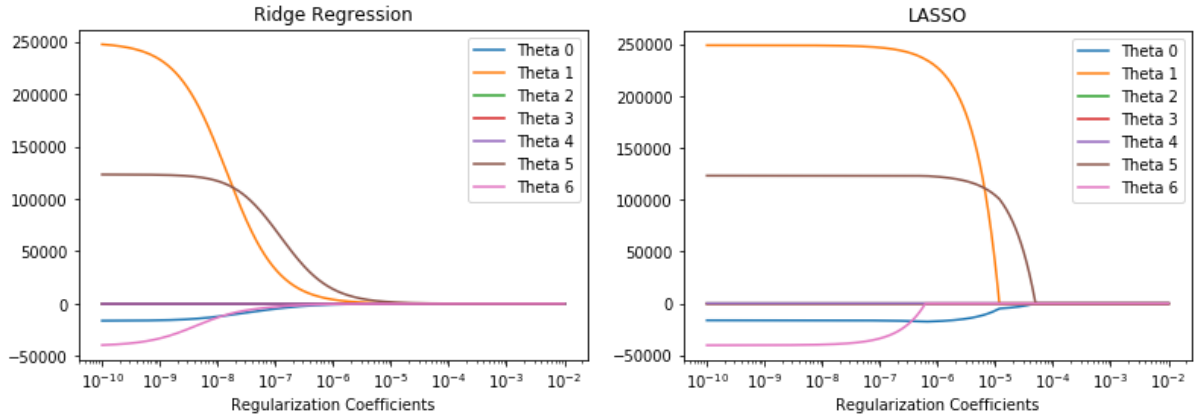
Q5.
(a)

```
Fold No  1
Optimal Penalty:  1
Train Error:  11.886619169173475
Validation Error:  10.627031885168384
Test Error:  11.990248315201793
Fold No  2
Optimal Penalty:  1
Train Error:  12.223306576509534
Validation Error:  10.352434889506265
Test Error:  10.478838282436534
Fold No  3
Optimal Penalty:  1
Train Error:  11.904432469726542
Validation Error:  11.595792687167666
Test Error:  10.309796390480598
Fold No  4
Optimal Penalty:  1
Train Error:  10.99248612966147
Validation Error:  13.415430618160944
Test Error:  11.694247098255621
Fold No  5
Optimal Penalty:  1
Train Error:  10.992130228562107
Validation Error:  12.033313548114242
Test Error:  13.519568205283628
```

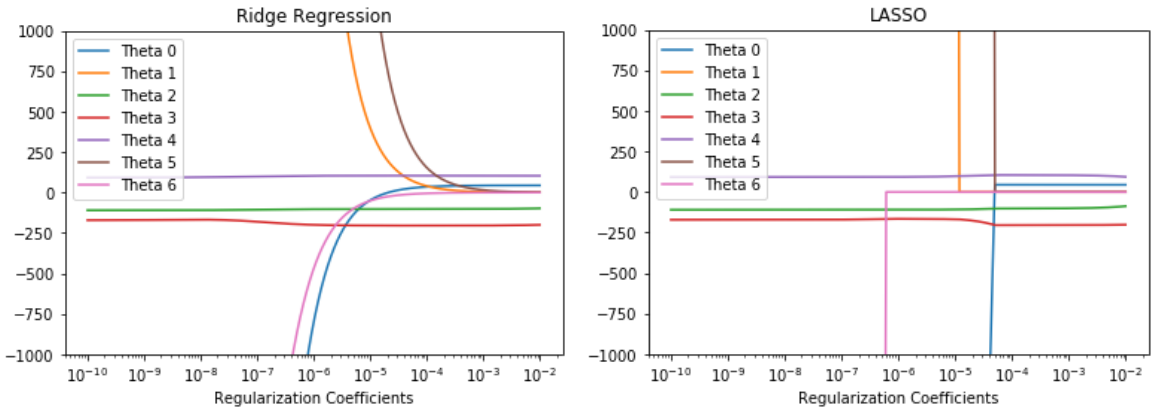
(b)

```
Fold No  1
Optimal Penalty:  1
Train Error:  13.484476916845283
Validation Error:  14.549148744706152
Test Error:  13.137785715655305
Fold No  2
Optimal Penalty:  1
Train Error:  13.863885620893614
Validation Error:  11.933781267314625
Test Error:  14.485371654502554
Fold No  3
Optimal Penalty:  1
Train Error:  14.637161794111762
Validation Error:  12.006714554567823
Test Error:  12.072655514450808
Fold No  4
Optimal Penalty:  1
Train Error:  13.217916232824804
Validation Error:  16.216780074311426
Test Error:  11.902226857365307
Fold No  5
Optimal Penalty:  1
Train Error:  12.821516284231857
Validation Error:  13.137841731145722
Test Error:  16.248122335933253
```

(c)



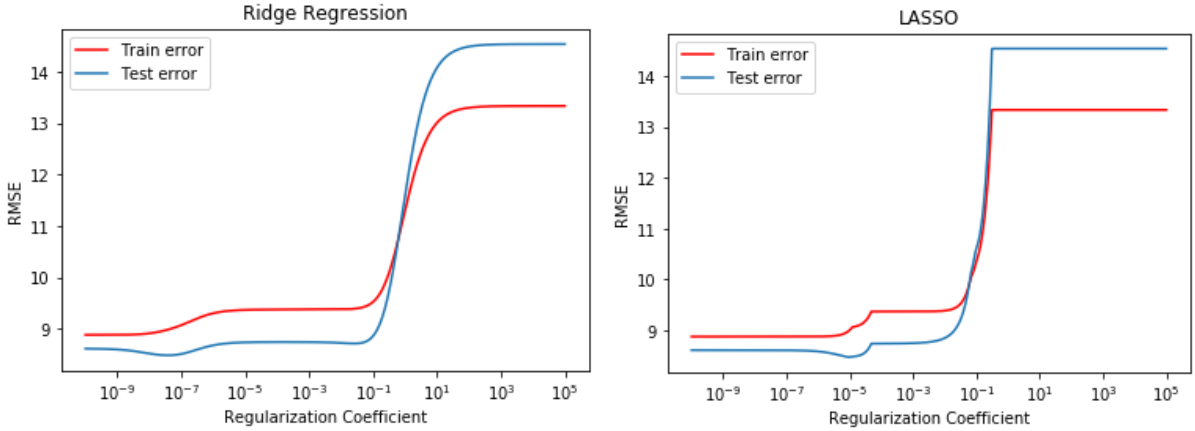
Range of λ Ridge Regression: logspace(-10,-2,1000)
Range of λ LASSO: logspace(-10,-2,1000)



On close inspection of ridge regression we can see that the coefficients tend to zero but never become zero. On the other hand the l1 norm in lasso forces some of the coefficients to zero.

Thus the sparsity of solution of LASSO is more than Ridge regression.

(d)



Range of λ Ridge Regression: logspace(-10,5,1000)
Range of λ LASSO: logspace(-10,5,1000)

(1) $X = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}$ $Y = \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix}$

(A)

$x^{(0)}$	$x^{(1)}$	y
1	1	6
1	3	10
1	6	16

$(\theta_0, \theta_1) = (0, 0)$ $N = 3$

Iteration 1 :-

Fix $\theta_0 = 0$

$$\theta_1 = \frac{\sum_{i=1}^{3} x_i^{(1)} (y_i - \hat{y}_i^{(-1)})}{\sum_{i=1}^{3} (x_i^{(1)})^2}$$

$$= \frac{1 \cdot (6 - \theta_{0_{init}} 1) + 3 \cdot (10 - \theta_{0_{init}} 3) + 6 \cdot (16 - \theta_{0_{init}} 6)}{1^2 + 3^2 + 6^2}$$

$$\hat{\theta}_1 = \frac{132}{46}$$

$$(\theta_0, \theta_1) = (0, 2.8695)$$

Iteration 2 :-

$(\theta_0, \theta_1) = (0, 2.8695)$

Fix $\theta_1 = 2.8695$

$$\theta_0 = \frac{\sum_{i=1}^{3} x_i^{(0)} (y_i - \hat{y}_i^{(0)})}{\sum_{i=1}^{3} (x_i^{(0)})^2}$$

$$= \frac{1 \cdot (6 - \theta_1 \cdot 1) + 1 (10 - \theta_1 \cdot 3) + 1 (16 - \theta_1 6)}{1^2 + 1^2 + 1^2}$$

$$\Rightarrow \theta_0 = \frac{32 - (2.8695)(10)}{3}$$

$$= \frac{3.305}{3}$$

$$\theta_0 \approx 1.1$$

$$(\theta_0, \theta_1) = (1.1, 2.8695)$$

Iteration 3:-

$(\theta_0, \theta_1) = (1.1, 2.8695)$

Fix $\theta_0 = 1.1$

$$\theta_1 = \frac{\sum_{i=1}^{3} x_i^{(1)} (y_i - \hat{y}_i^{(-1)})}{\sum_{i=1}^{3} (x_i^{(1)})^2}$$

$$= \frac{1.(6 - \theta_0.1) + 3(10 - \theta_0.1) + 6(16 - \theta_0.1)}{1^2 + 3^2 + 6^2}$$

$$\Rightarrow \theta_1 = \frac{132 - 10\theta_0}{46}$$

$$= \frac{132 - 10(1.1)}{46}$$

$$= \frac{132 - 11}{46}$$

$$\theta_1 = 2.63$$

$$\boxed{(\theta_0, \theta_1) = (1.1, 2.63)}$$

b) $(\theta_0, \theta_1) = (0, 0)$ $\alpha = 0.01$

$$X = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix} \quad Y = \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix} \begin{matrix} \rightarrow (1) \\ \rightarrow (2) \\ \rightarrow (3) \end{matrix}$$

Iteration 1:-

Sample ①

(1, 6)

$$RSS = (6 - (\theta_0 + \theta_1(1)))^2$$

$$\frac{\partial RSS}{\partial \theta_0} = 2.(6 - \theta_0 - \theta_1)(-1) = -12$$

$$\frac{\partial RSS}{\partial \theta_1} = 2(6 - \theta_0 - \theta_1)(-1) = -12$$

$$\begin{bmatrix} \theta_0^{(1)} \\ \theta_1^{(1)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.01 \begin{bmatrix} -12 \\ -12 \end{bmatrix}$$

$$\begin{bmatrix} \theta_0^{(1)} \\ \theta_1^{(1)} \end{bmatrix} = \begin{bmatrix} 0.12 \\ 0.12 \end{bmatrix}$$

Iteration 2:-

Sample ②

(3, 10)

$$RSS = (10 - (\theta_0 + \theta_1(3))^2$$

$$\frac{\partial RSS}{\partial \theta_0} = 2(10 - \theta_0 - 3\theta_1)(-1) \qquad \frac{\partial RSS}{\partial \theta_1} = 2(10 - \theta_0 - 3\theta_1)(-3)$$

$$= -19.04 \qquad = -57.12$$

$$\begin{bmatrix} \theta_0^{(2)} \\ \theta_1^{(2)} \end{bmatrix} = \begin{bmatrix} 0.12 \\ 0.12 \end{bmatrix} - 0.01 \begin{bmatrix} -19.04 \\ -57.12 \end{bmatrix}$$

$$\begin{bmatrix} \theta_0^{(2)} \\ \theta_1^{(2)} \end{bmatrix} = \begin{bmatrix} 0.3104 \\ 0.6912 \end{bmatrix}$$

Iteration 3 :-

Sample (3)

(6, 16)

$$RSS = (16 - \theta_0 + \theta_1(6))^2$$

$$\frac{\partial RSS}{\partial \theta_0} = 2(16 - \theta_0 - 6\theta_1)(-1) = -2(16 - 0.3104 - 6(0.6912)) = -23.0848$$

$$\frac{\partial RSS}{\partial \theta_1} = 2(16 - \theta_0 - 6\theta_1)(-6) = -12(16 - 0.3104 - 6(0.6912)) = -138.5088$$

$$\begin{bmatrix} \theta_0^{(3)} \\ \theta_1^{(3)} \end{bmatrix} = \begin{bmatrix} 0.3104 \\ 0.6912 \end{bmatrix} - 0.01 \begin{bmatrix} -23.0848 \\ -138.5088 \end{bmatrix}$$

$$\begin{bmatrix} \theta_0^{(3)} \\ \theta_1^{(3)} \end{bmatrix} = \begin{bmatrix} 0.541248 \\ 2.076288 \end{bmatrix}$$

(c) $X = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}$ $y = \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix}$ $\lambda = 1$

$$\theta = \left(X^T X - \lambda \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{(d+1)\times(d+1)}\right)^{-1} X^T y$$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix} = \begin{bmatrix} 3 & 10 \\ 10 & 46 \end{bmatrix}$$

$$X^T X - \lambda \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 10 \\ 10 & 45 \end{bmatrix}$$

$$\left(X^T X - \lambda \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right)^{-1} = \begin{bmatrix} 3 & 10 \\ 10 & 45 \end{bmatrix}^{-1} = \frac{1}{135-100} \begin{bmatrix} 45 & -10 \\ -10 & 3 \end{bmatrix}$$

$$\theta = \begin{bmatrix} 1.28 & -0.285 \\ -0.285 & 0.086 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 6 \end{bmatrix} \begin{bmatrix} 6 \\ 10 \\ 16 \end{bmatrix}$$

$$= \begin{bmatrix} 1.28 & -0.285 \\ -0.285 & 0.086 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ 1 & 3 & 2 \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} = \begin{bmatrix} 3.428 \\ 2.1714 \end{bmatrix}$$