MACHINE LEARNING LAB 2
22nd September 2017

Manish Singh (2017CSM1003)
Sujit Rai (2017CSM1006)

REGULARISED LOGISTIC REGRESSION

- Dataset Plot for the two classes.Feature x1 are represented on x-axis and feature x2 are represented on y-axis

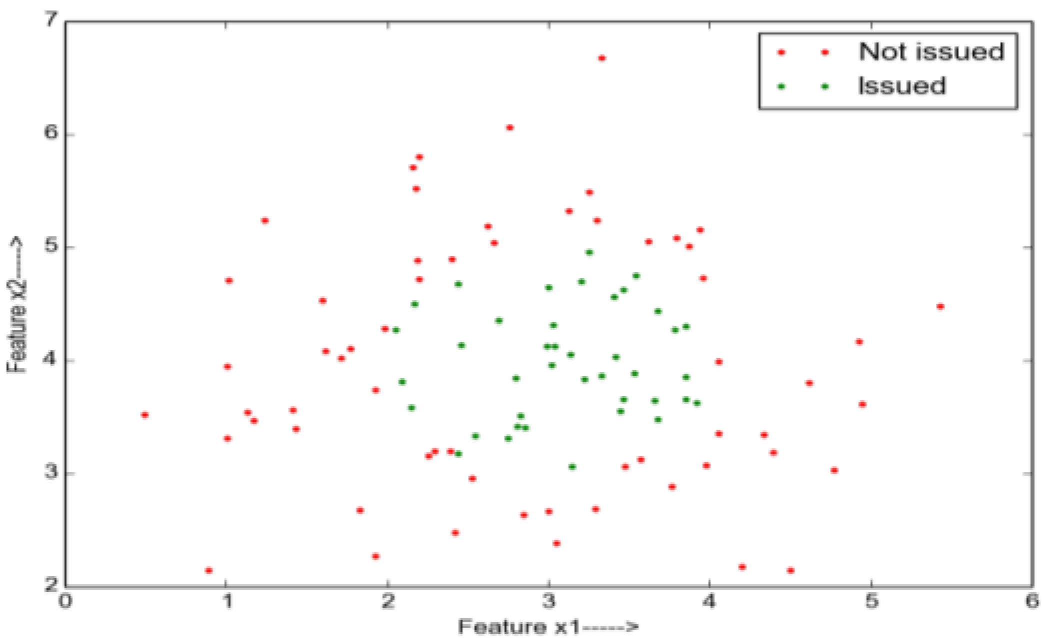
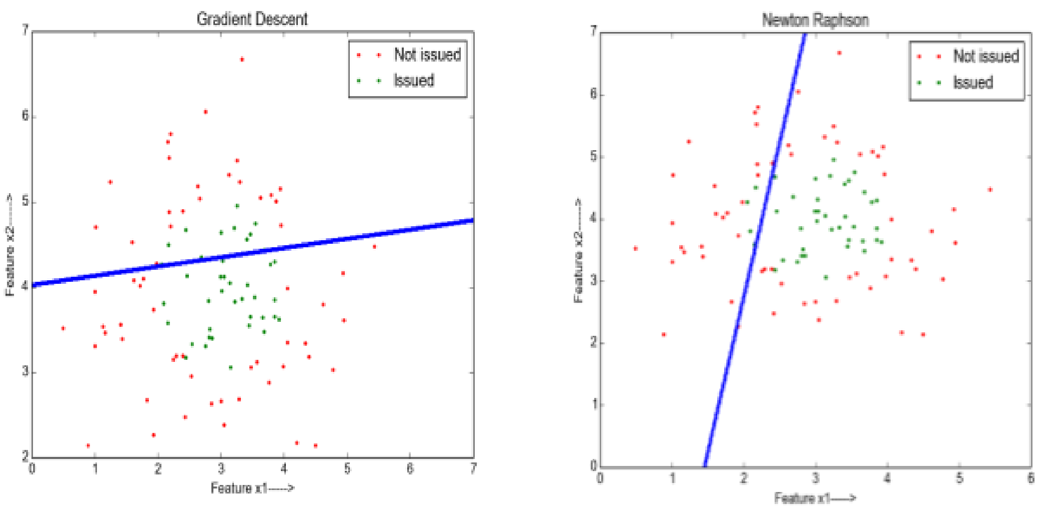


Fig 1. Green dots denotes credit card can be issued and red dots denotes that credit card cannot be issued

- Regularised logistic regression has been implemented with gradient descent and Newton Raphson as the optimization method. The following is the output of the line plotted with both of them.Clearly the line is unable to classify the two classes of data separately.

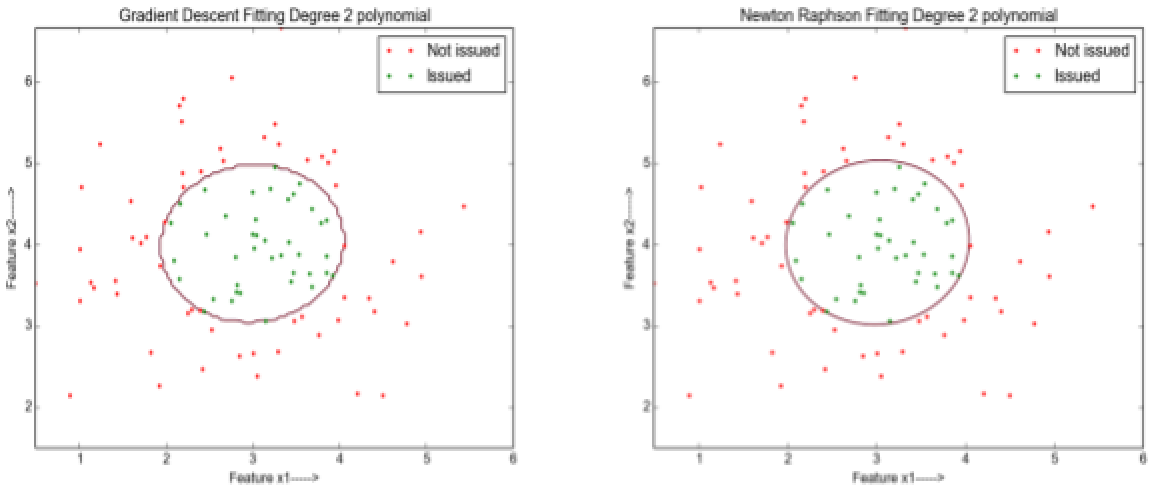


(a) (b)

Fig.2 a) on left shows gradient descent output line b) on right shows Newton Raphson output.

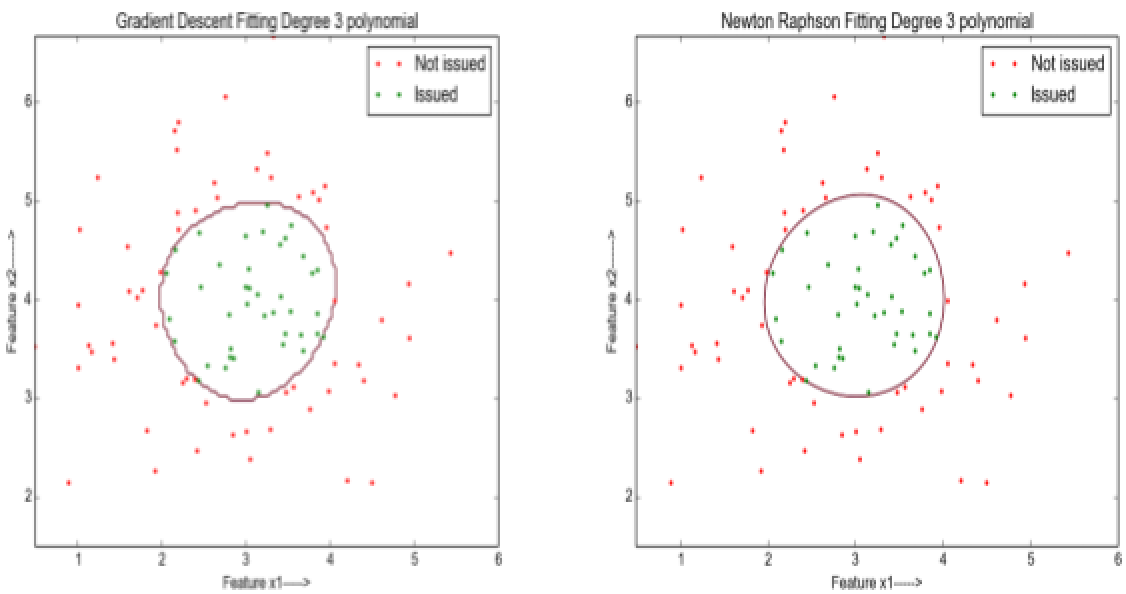
- The data given in credit.txt file is plotted in Fig .1 cannot be classified linearly as is shown in the Fig. 2 a) Gradient Descend and Fig. 2 b) Newton Raphson.

- Since the data was not linearly separable so features have been increased using the existing features to increase the degree of the curve that can be fit through the data.It has been noted that degree two polynomial fits the data accurately. Gradient Descend and Newton Raphson has been used as the optimization methods. Gradient Descend takes large number of iteration(in lakhs) to converge while Newton Raphson takes only a few iterations(5-20) to converge. Both of them gives 100 percent accuracy on the training data.It has also been observed that any polynomial with degree greater than 2 will accurately classify the two classes of data.
- Decision boundaries for the polynomial of different degree has been fit through the data shown in Fig. 1 and shown in the figure below



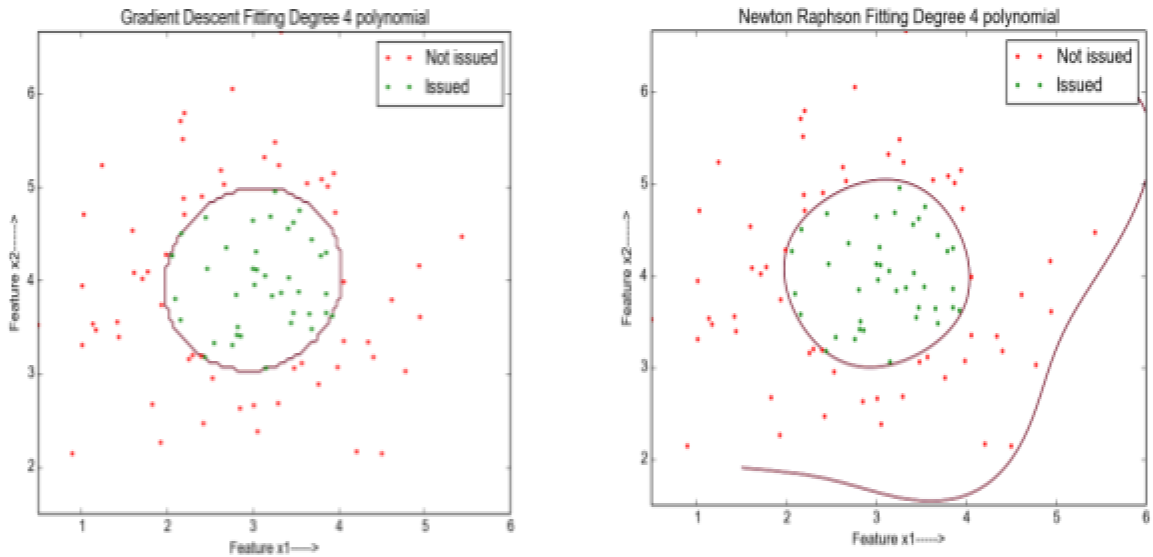
(a) (b)

Fig.3 a) on left shows decision boundary using Gradient Descend b) on right shows decision boundary using Newton Raphson



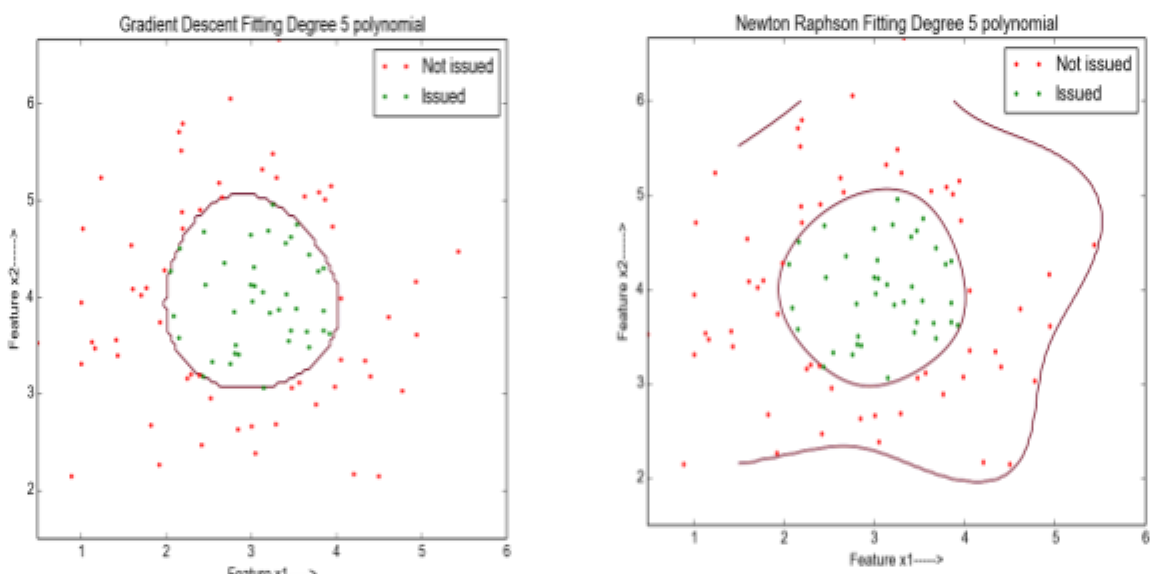
(a) (b)

Fig.4 a) on left shows decision boundary for degree 3 polynomial using Gradient Descend and Newton Raphson respectively



(a) (b)

Fig. 5 a) on left and b) on right shows decision boundary plotted using degree 4 polynomial



(a) (b)

Fig. 5 a) on left and b) on right shows decision boundary of Gradient Descend and Newton Raphson respectively for a degree 5 polynomial

Observations: It has been observed that as we increase the degree of the polynomial the decision boundary accurately and more precisely fits the training data.Degree 2 polynomial classifies the training examples with 100 percent accuracy. For optimal performance the polynomial with degree two is good as polynomial with higher degree leads to overfitting the training data.

- Various plots of decision boundary has been drawn with different values of lamda (the regularization parameter) and it has been observed that with lamda value as zero the decision boundary overfits the training data and as we gradually increase the value of lamda the decision boundary underfits the training data. The plots below summarize the results observed.

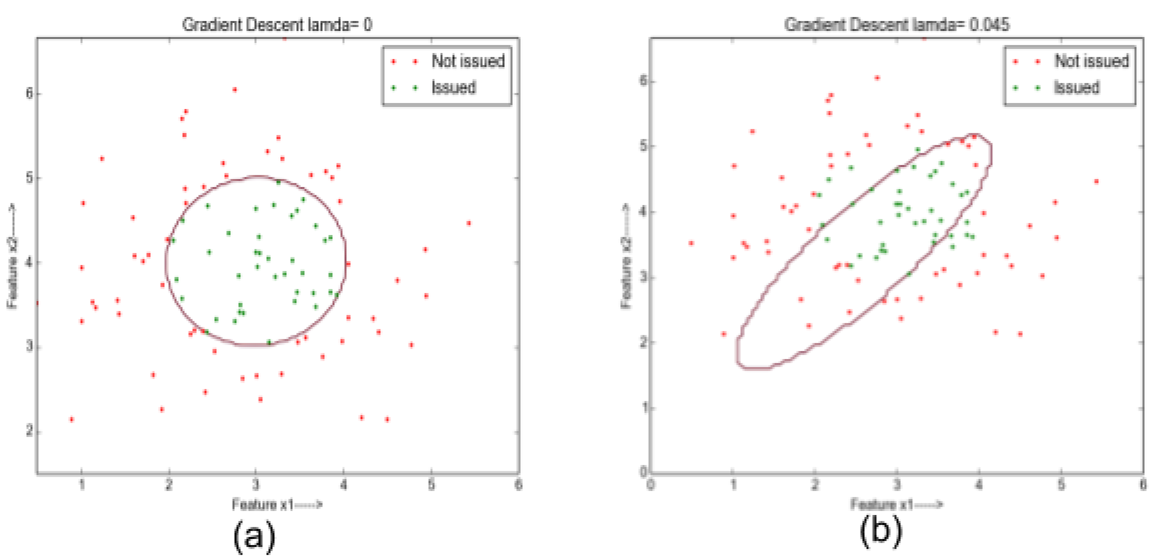


Fig. 6 a) denotes Overfitting with lamda value as zero and b) denotes Underfitting with lamda value 0.045 for a degree two polynomial using gradient descend

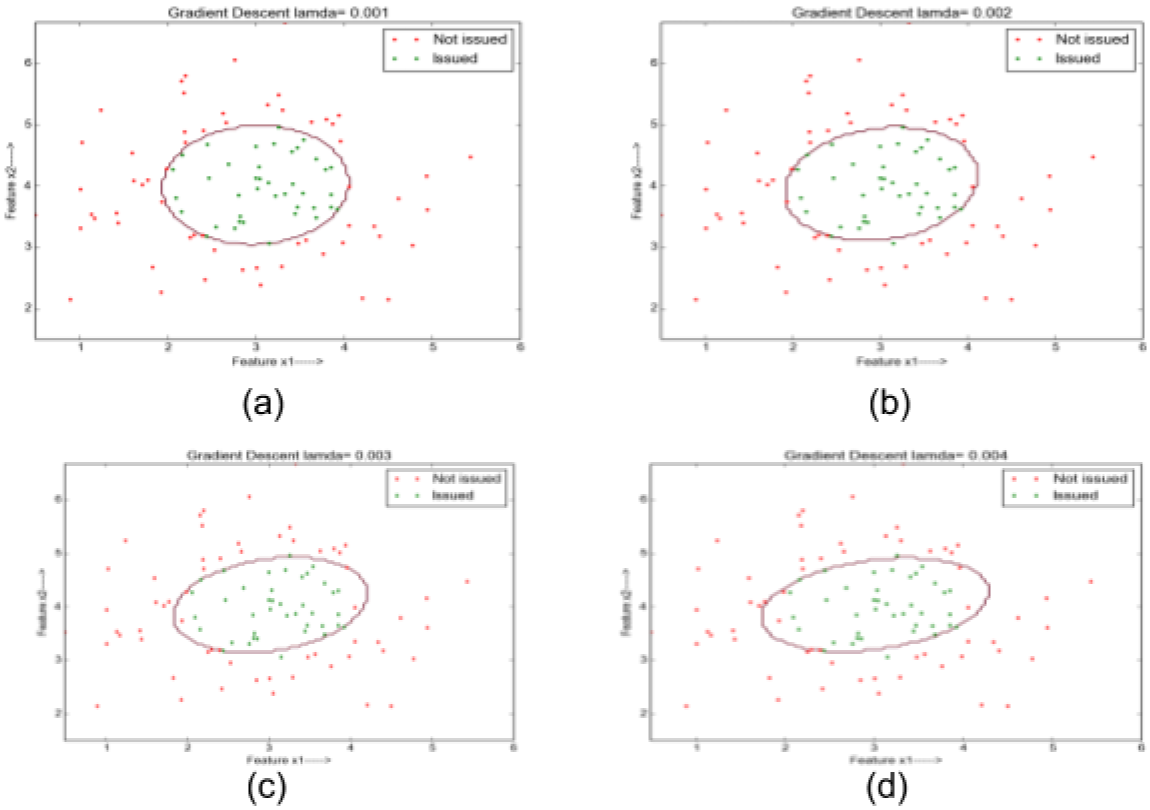


Fig. 6 (a) ,(b),(c) and (d) shows decision boundary with lamda's in increasing order

Observation : It has been observed that by increasing the values of lamda we get a decision boundary which underfits tha data and by keeping the value of lamda negligible the decision boundary classifies the data with 100 percent accuracy.Another observation is that adding the regularization paramter lamda increses the convergence rate of gradient descend.

Regularized Linear Regression :

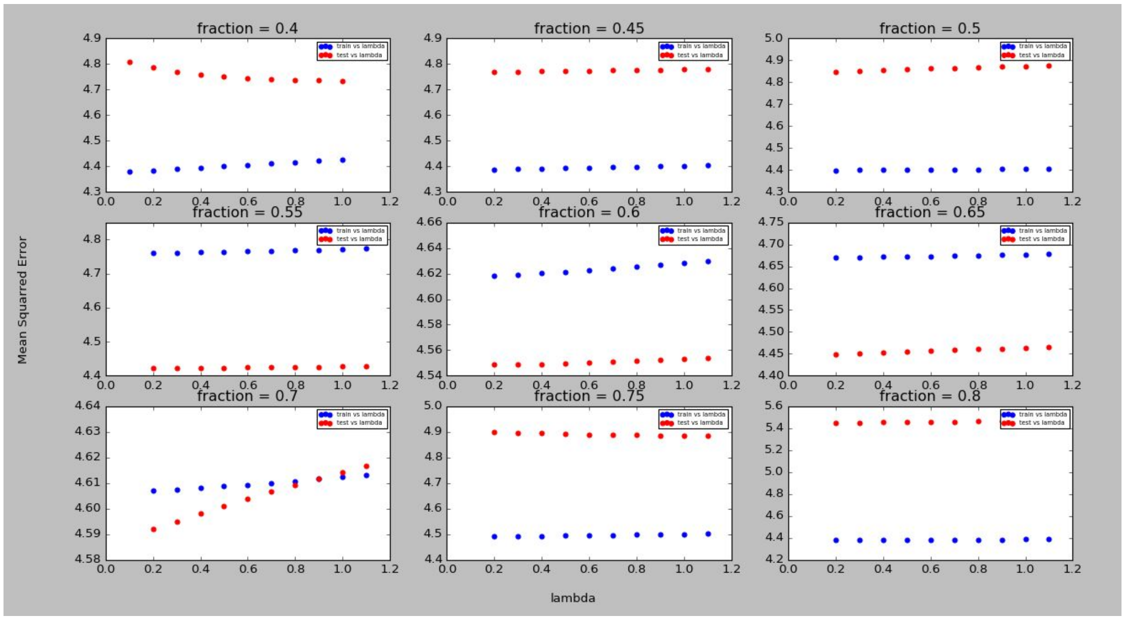
Observations during pre-processing & Analysis of data:

- During analysis of data, It was found that the height attribute has an almost linear relation with the age attribute (target Value).
- It was found that the size of the abalone i.e height,length and diameter, are the most significant attributes.
- It was found that the gender attribute is the least significant attributes because plotting its graph provides least linearity
- During Pre-processing the gender attributes were converted from character to integers i.e -1,0,1
- Additional columns were added to increase the dimensionality of the equation

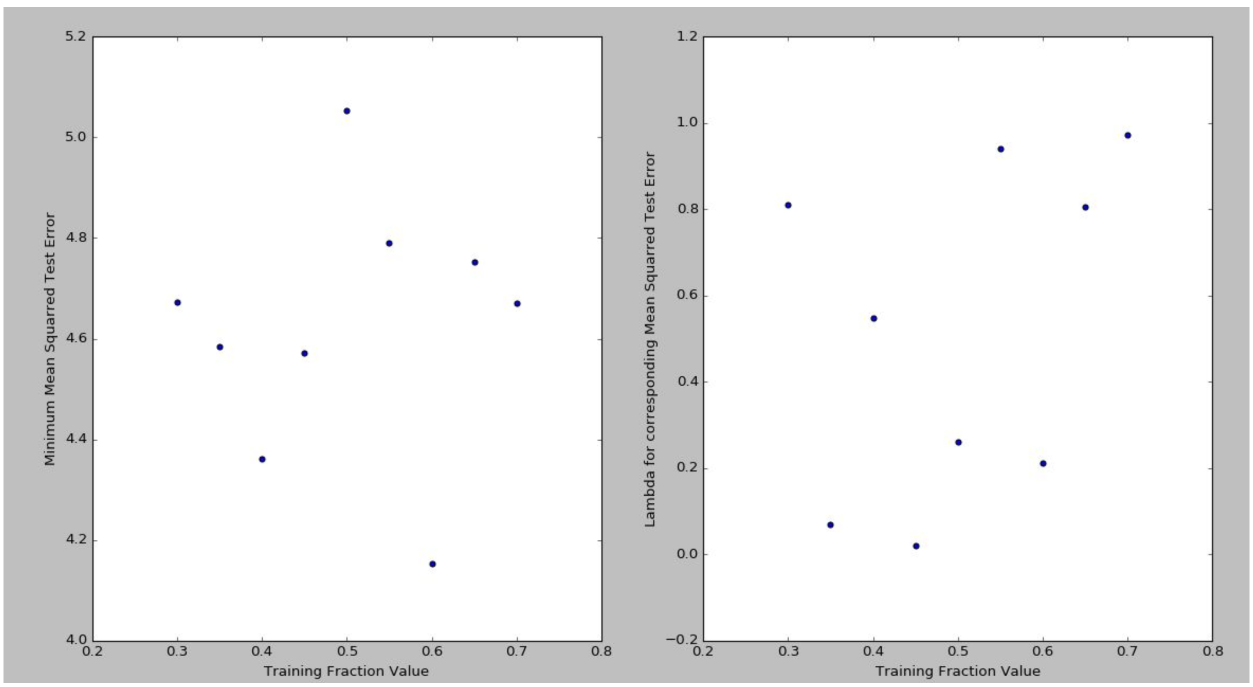
Questions :

1. The gender columns was replaced with its respective tuple representation and stored in q1output.txt
2. All the columns except the gender and the output column is standardized and result is stored in q2output.txt
3. Normal Equation was used to calculate the weight of the equation
4. Entire dataset was shuffles and partitioned into 2 parts, wherein 20% was considered as train data and 80% was considered as test data.
5. Mean Squared Error was calculated using the predicted values and actual values

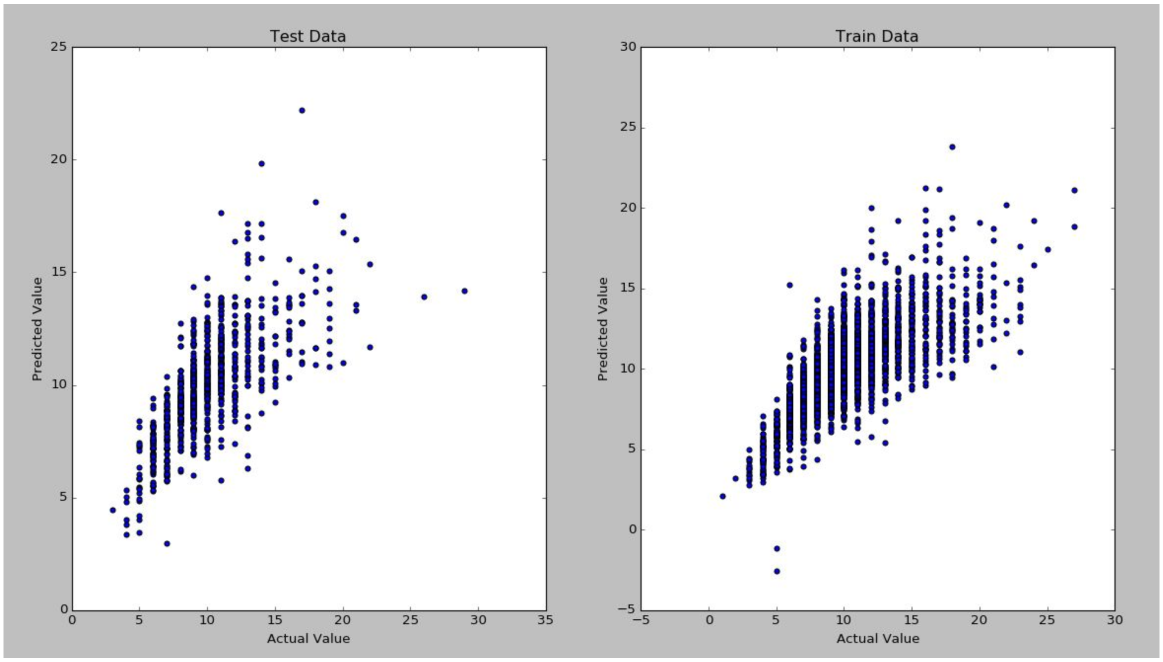
6. Since gender and some of the weight attributes are least significant attributes therefore they were eliminated and the MSE was calculated and compared. The size i.e length,height,diameter are the most significant attributes
7. Questions :
    a. Does the effect of λ on error change for different partitions of the data into training and testing ?
        i. Yes, as the λ increases it is found that the training error increases gradually but the test error decreases sometimes .
    b. How do we know if we have learned a good model?
        i. When the difference in the training error and test error doesn’t vary by a large extent.
        ii. A good model is one in which the predicted value when plot against the actual value lie close to the 45-degree line plotted through the origin
        iii. A bad model is one in which the training accuracy is maximum but the test accuracy is very low

8. 

As we can see in subplot(1,1) as the value of λ increases the train error (blue color) increases but the test error (red color) decreases

9. 

As we can see from figure, MSE is around 4.1 for the instance when the training fraction was 0.6 and λ value 0.2. So as the training fraction increases the MSE decreases when λ value is lies between range(0.1,0.2)

10. 

For the above figure the training data was choosed to be 0.8 and the λ value was 0.01. As we can see that all the points (except some outliers) are close to the 45-degree plot throught the line