Regularized Linear Regression and Logistic Classification

Aditya Tiwari

[*] Computer Science and Engineering
[**] IIIrd Year Undergraduate

Linear Regression: Given data about snails,we have to predict its age using linear ridge egression

Logistic Classification:Given two attributes of a credit card application,we have to decide whether to accept or reject the application using Logistic Classification

I. Linear Ridge Regression

REPRESENTATION:

The data set has 10 attributes out of which the first is the gender of the snail and the rest are some attributes which are to be used. I use three columns [F M I] to represent the gender of the snail where each column is 1 or 0 depending upon data where :

F~female
M~Male
I~Infant
and then I append the the rest of attributes.There are a total of 4177 datas each of 11 columns.

Experiment Observations and Findings

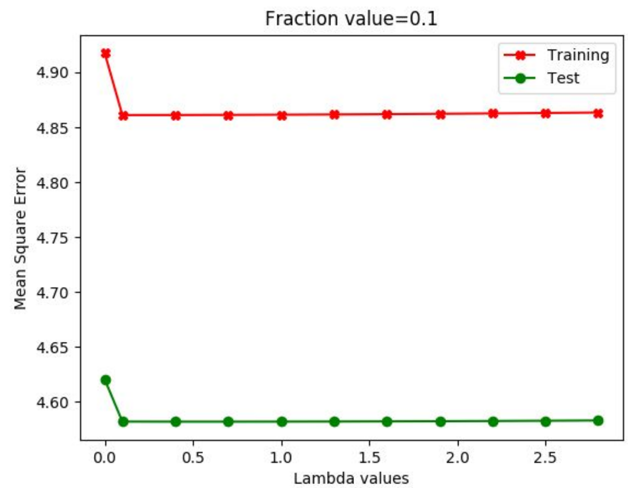
Does the effect of λ on error change for size of the training set?

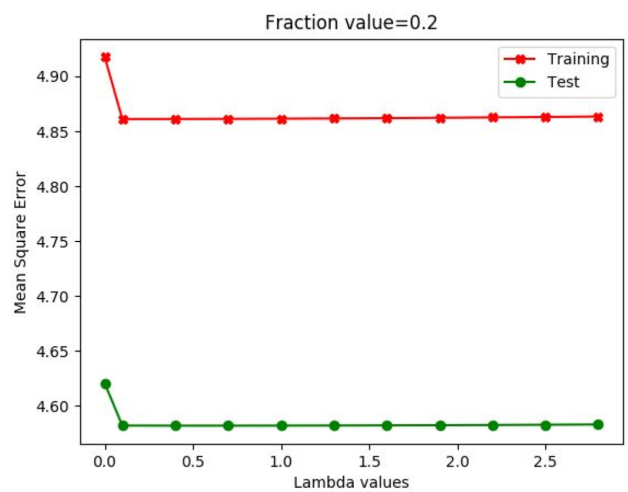
I first observe the change of **λ** in error in test error and training error for different values of fraction

The graph show that with no penalizing factor i.e **λ=0** Training Error and Test Error is more which means the model was initially overfitting the dataset for each value of fraction

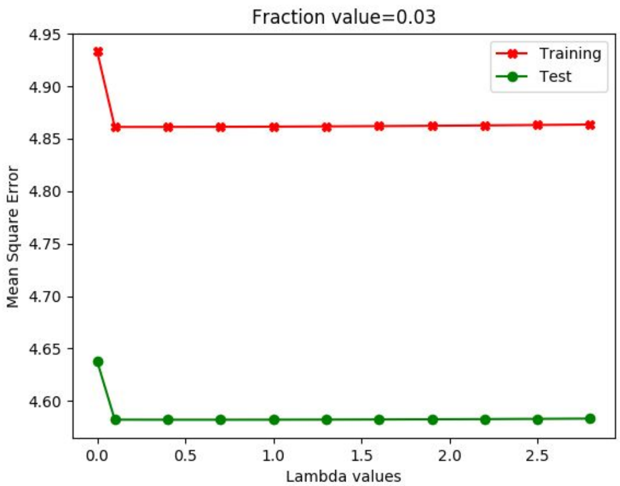
I have tested for the following combinations of **λ** and fractions:

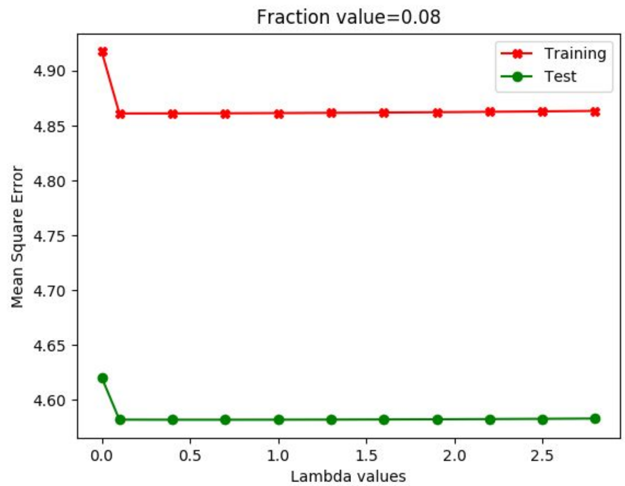
λ=[0.0, 0.1, 0.4, 0.7, 1.0, 1.3, 1.6, 1.9, 2.2, 2.5, 2.8]
fraction=[0.03, 0.05, 0.08, 0.1, 0.15, 0.2, 0.5,1.0]

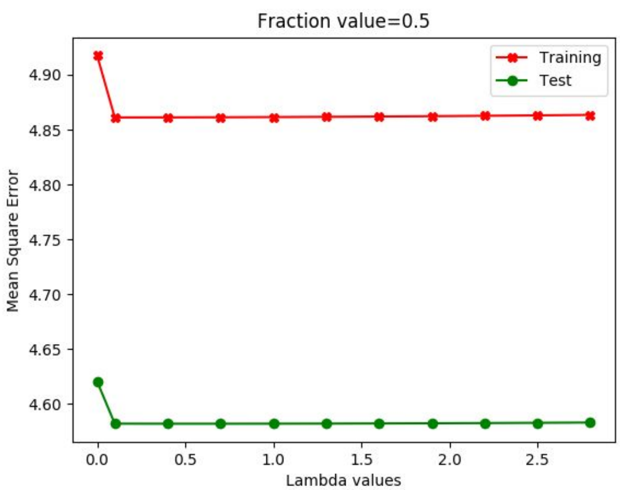


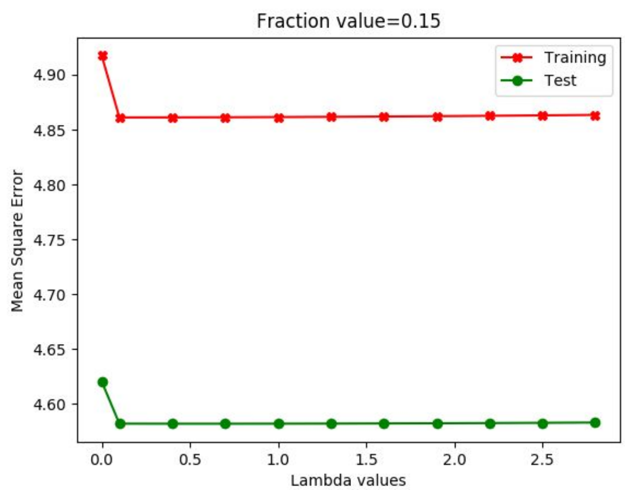


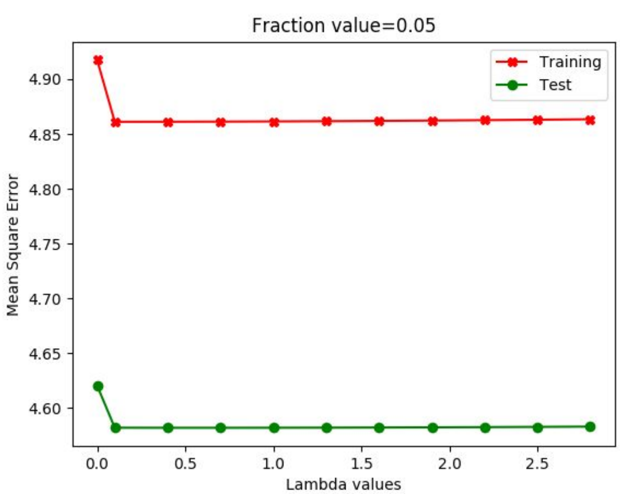
The Error in Test or Training initially decreases with increasing **λ** even with small value of **λ.**Which means that the model is very sensitive to penalizing and hence maybe decreasing the **λ** might even decrease the error.The Error later increases very slowly with increase in **λ**,which means we are over penalizing the model and regressor is losing essential information for fitting the model.

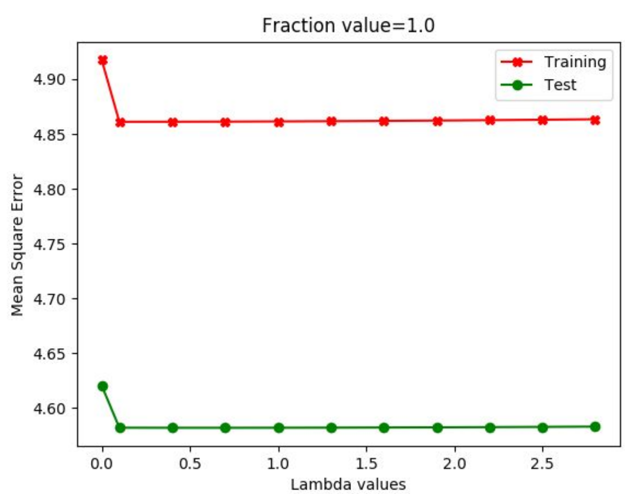










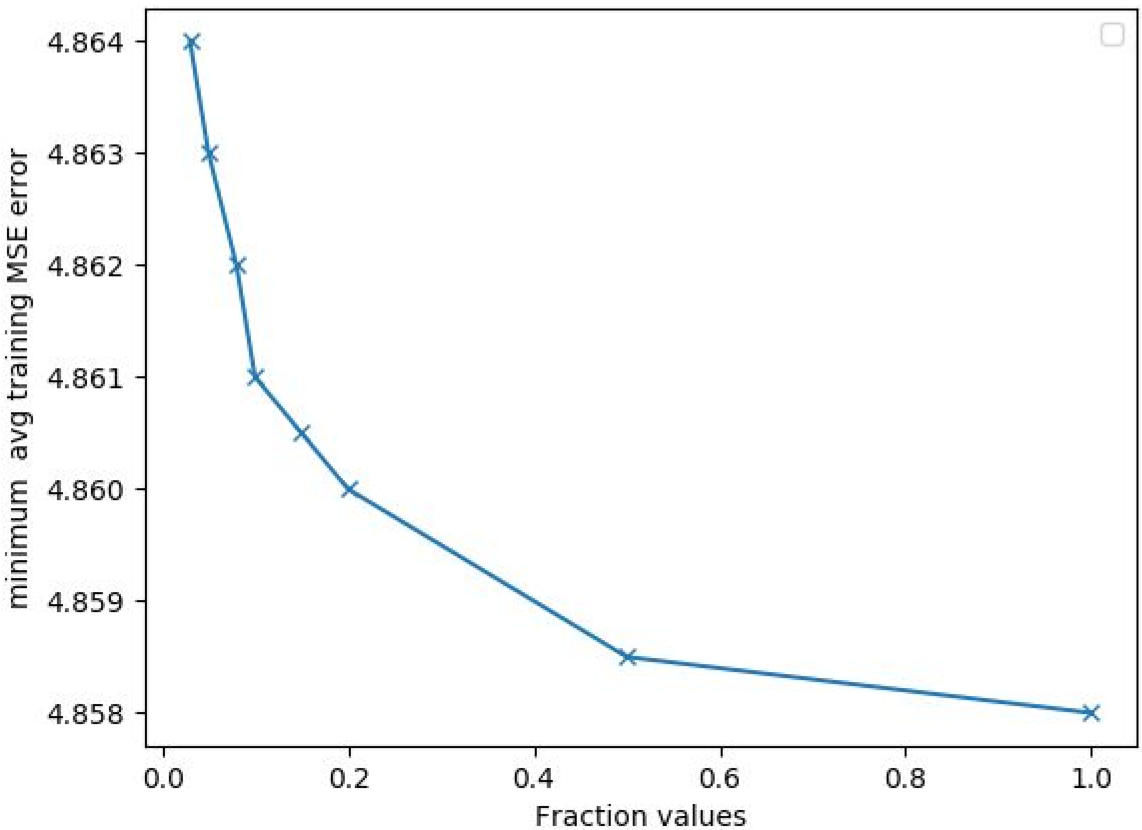


So hence the Increasing the **λ** first decreases the error and later is almost constant.

How do we know if we have learned a good model?

If the model is good then it should not only minimize the Mean square Error but also should minimize absolute loss as well.Means when graph drawn between actual and predicted values should be of Y=X

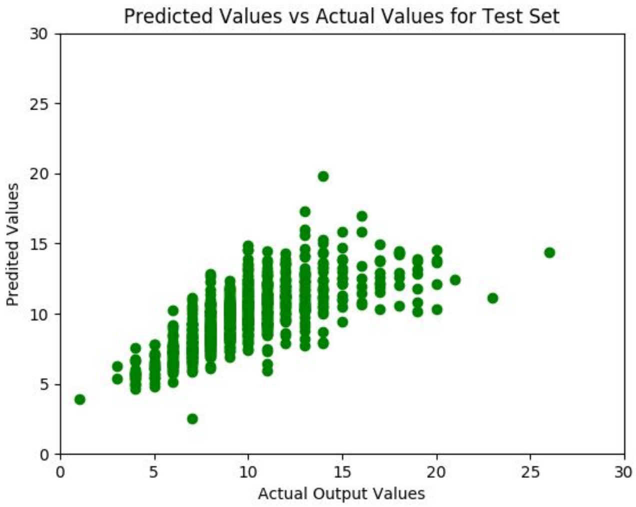
More insights from Graph:

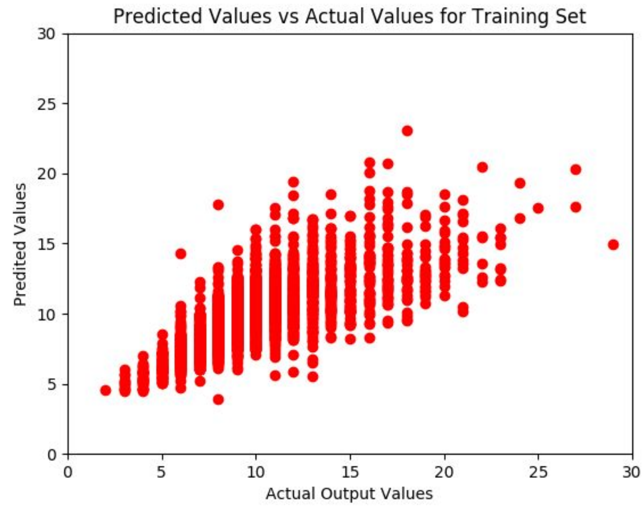


We can see the Minimum avg Training Mean square Error Reduces with Increasing Fraction of Training sample which is explainable as increasing the sample would give the model more data to regress and come up with better Hypothesis than with less sample with overlooking the over-fitting. Although the difference in decrease is very minor.

Predicted Vs Actual Output

The graph comes somewhat balancing Y=X .The Deviation from Y=X is what I think contributing so much to Mean Square Error because it is squared when adding. Absolute Error Function

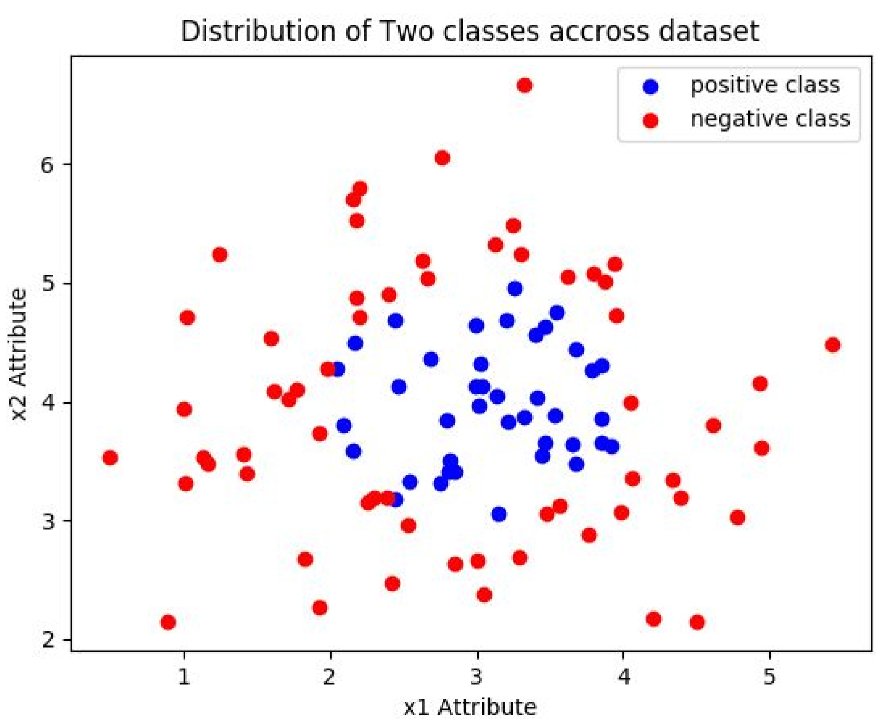




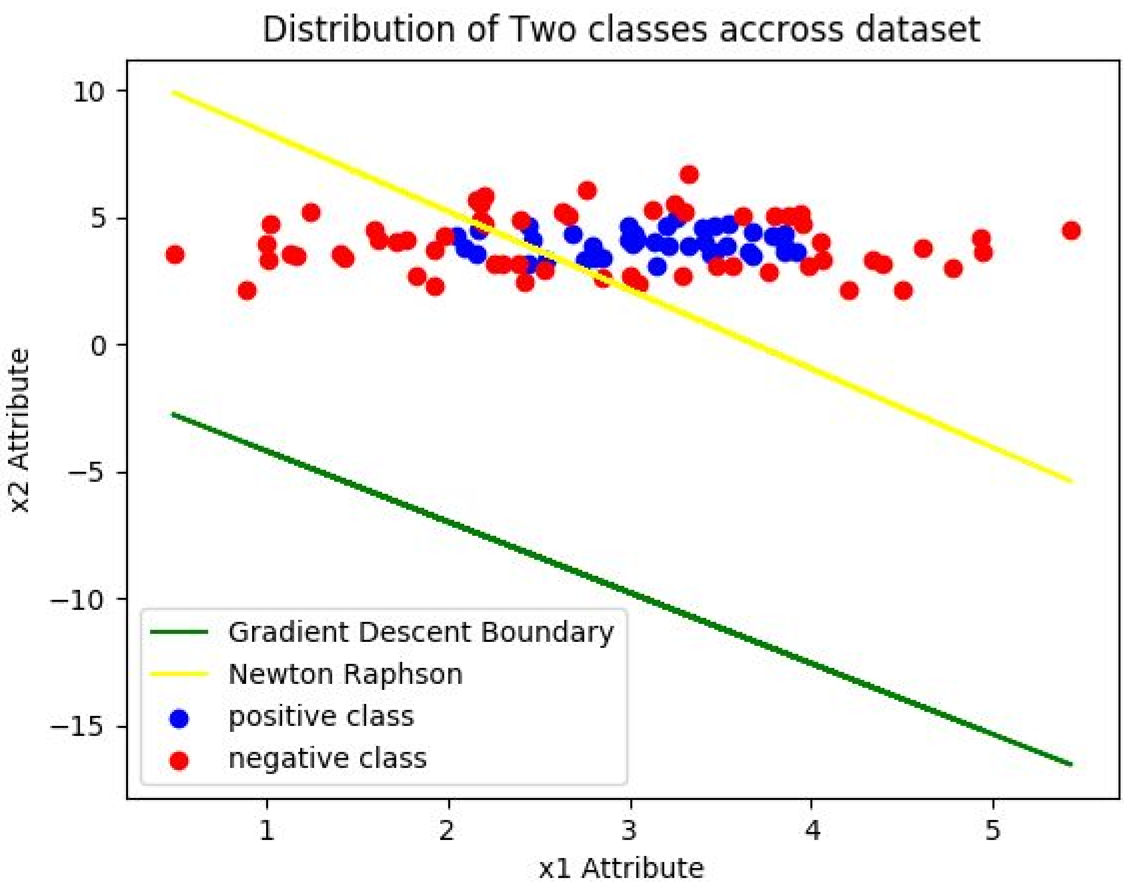
II. Regularized Logistic REgression

Data Representation: The dataset given has 3 values, 2 attributes on which we have to predict the acceptance or rejection and third is the actual that we should have done.

Data Distribution: The data is not linearly separable i.e a linear line cannot separate two class where blue represents the acceptance of application and red rejection



The boundary drawn using Gradient Descent and Newton Raphson Method are linear as:



Newton Raphson takes much more time to find the weights.

The accuracy is on test set is with parameters[λ=1,Number of iterations=10000]:

1. Gradient Descent:58.75
2. Newton Raphson:61.25

Since the data seems such that a Hypothesis of Circle could separate it,I’d say the best Hypothesis is of degree 2.