Assignment 03

Question 1:

Using the data set of two examination results design a predictor using logistic regression for predicting whether a student can get an admission in the institution. Use regularizer to further tune the parameters. Use 70 % data for training and rest 30% data for testing your predictor and calculate the efficiency of the predictor/hypothesis.

Fig. 1 plots the entire datasets where the different color points represents the different labels

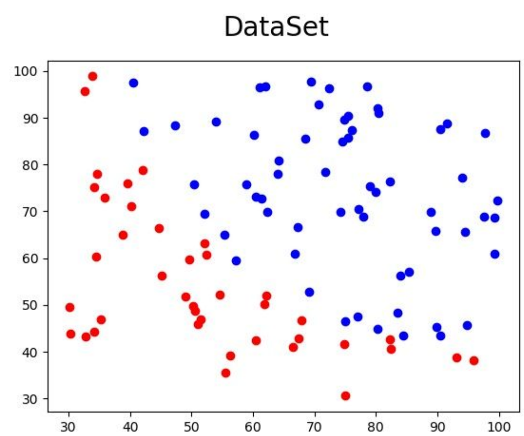

Fig. 1 Exam Results DataSet

Clearly, It is evident that normal line equation will predict better as We can draw some line such that data will different labels will be on different sides with respect to line (approx).

So, I trained on 70% of data and tests the hypothesis on remaining 30% of data. Fig 2 clearly shows that value of cost decrease with the increase in number of epochs.

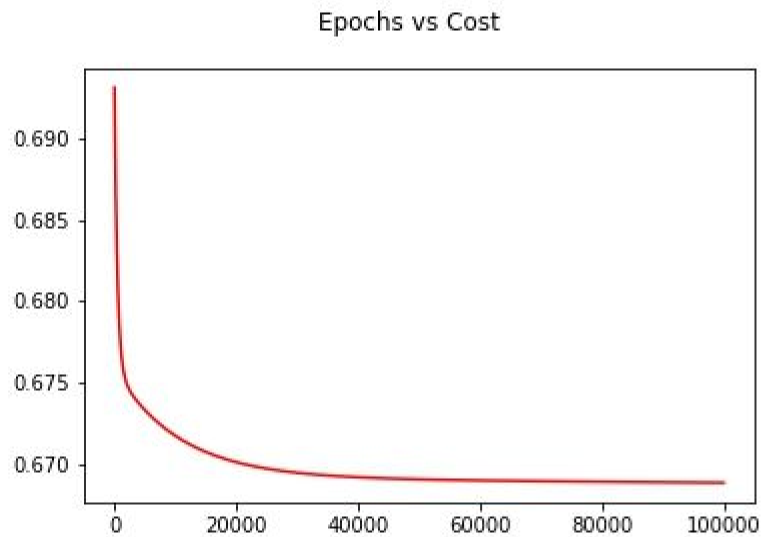

Fig. 2 Epohs vs Cost, trained over 10^5 epochs at $\alpha = 5 \times 10^{-6}$

I predicted the value on test dataset. Accuracy of this predictor is 80%. To further tune the parameters, I used regularization.

Question 2:

Using the data set of two quality test results of a microchip product, design a predictor using logistic regression which will predict the acceptance or rejection of the microchip given the two test results. Use regularizer to further tune the parameters. Use 70 % data for training and rest 30% data for testing your predictor and calculate the efficiency of the predictor/hypothesis.

Fig. 3 plots the entire datasets where the different color points represents the different labels

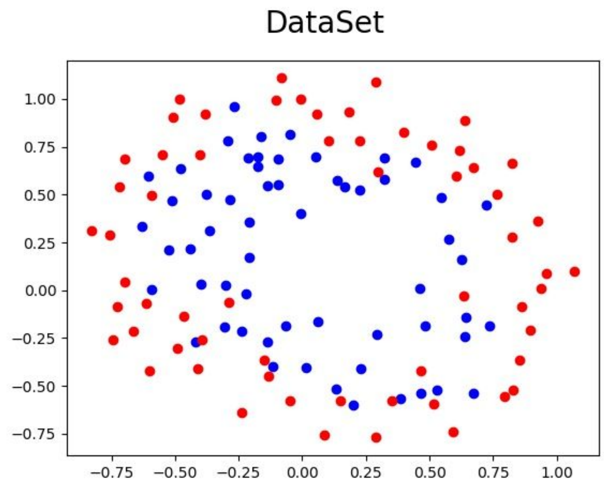

Fig. 3 Micro chip DataSet

Clearly, It is evident that normal line equation can't predict better as We can't draw a line such that data will different labels will be on different sides with respect to line (approx). So, I preferred 2nd order polynomial equation as represented below,

$$W^TX = \theta_1 + \theta_2\, x_1^2 + \theta_3\, x_2^2 + \theta_4\, x_1 x_2$$

So, I trained on 70% of data and tests the hypothesis on remaining 30% of data.Fig 4 clearly shows that value of cost decrease with the increase in number of epochs.

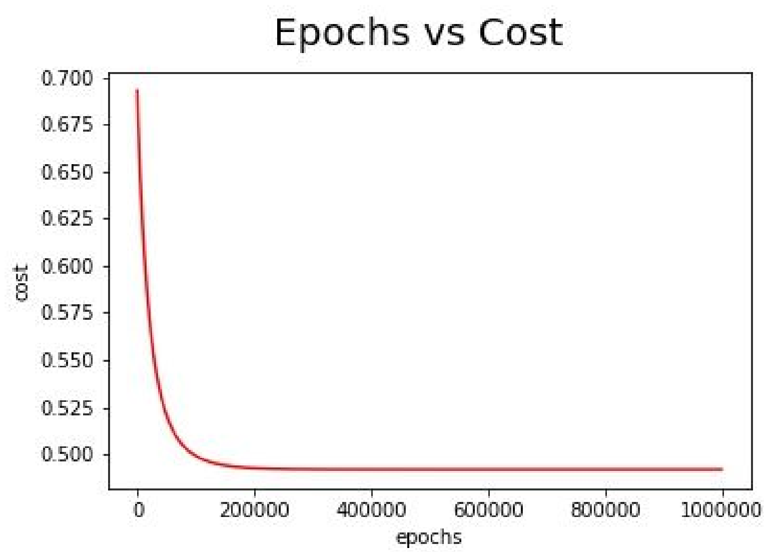

Fig. 4 Epochs vs Cost, trained over 10^6 epochs at $\alpha = 0.03$

I predicted the value on test dataset. Accuracy of this predictor is 77%. Accuracy may change if we compile it again, as I am shuffling the data in order to maintain the ratio of labels in train and test datasets. To further tune the parameters, I used regularization.

The below figure clearly shows the effect the change in value of cost function with respect to λ. This clearly shows that predictor performs better when λ = 0 and after reaching some value, it says same.

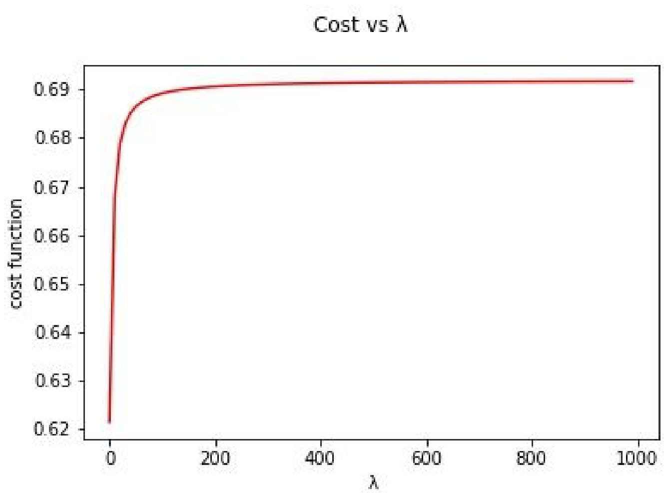

Fig. 5 λ vs Cost, trained over 10^4 epochs at α = 0.03

The plot between Accuracy and λ also looks like the above graph itself.

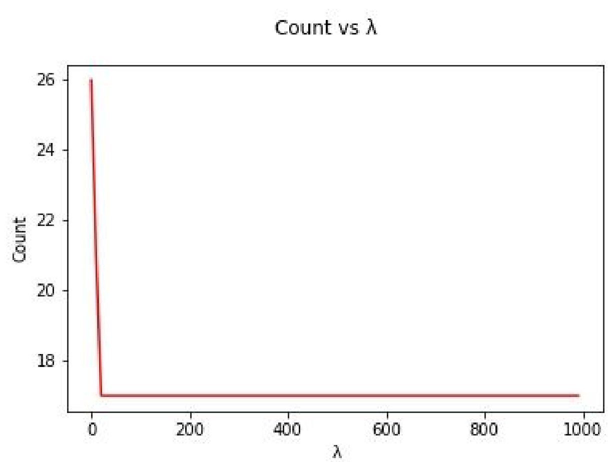

Fig. 6 Count vs λ where, count : number of correct predictions