

SCO Assignment 3 - Analysis Report

By IIT2016085, Luv

Overview

Use Logistic Regression to classify various datasets

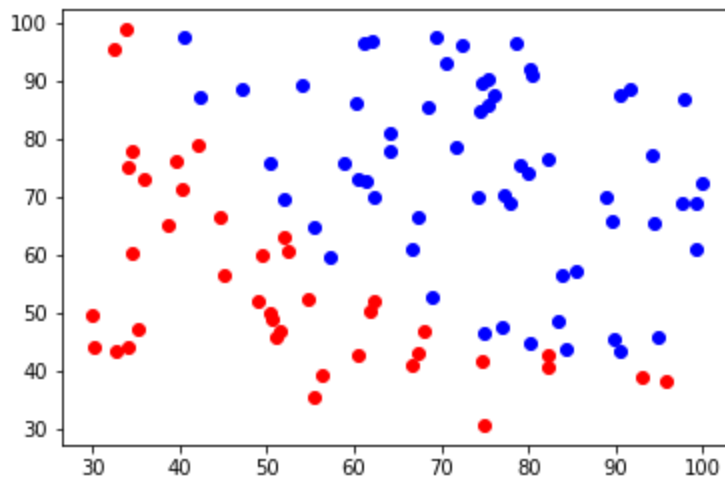
Goals

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.
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.

Analysis

1. Examination Results

Graph below shows the nature of the dataset.

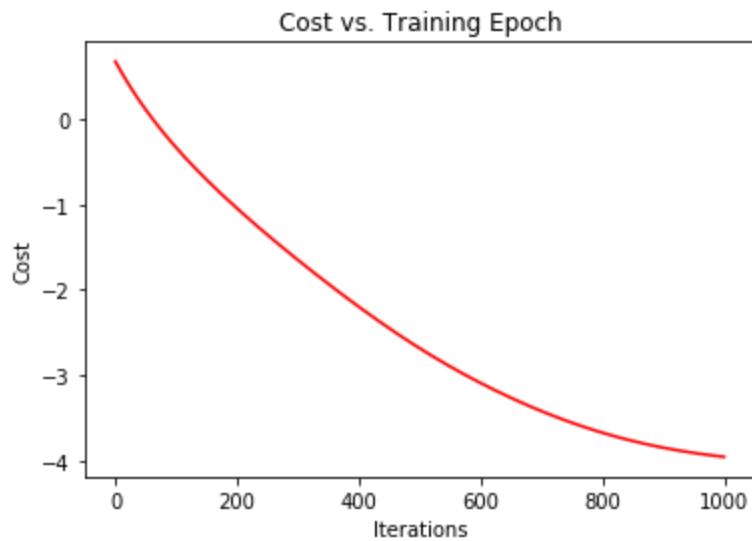


As can be seen using dataset the classification can be obtained by using only two given features Hence in total we use 3 features. Theta matrix is generated using Gradient Descent method and logistic regression with below results.

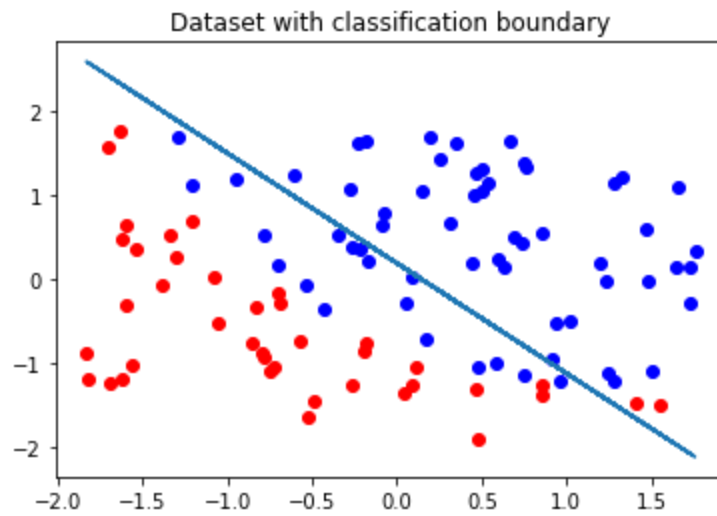
Accuracy : 82.75 %

Error : 17.25 %

Cost vs iteration curve for Gradient Descent Algorithm is shown below

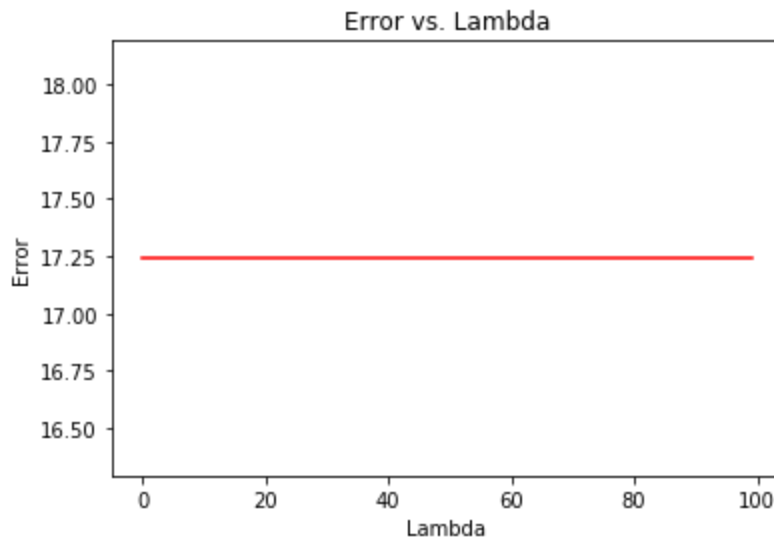


Below is the graph showing dataset with classification



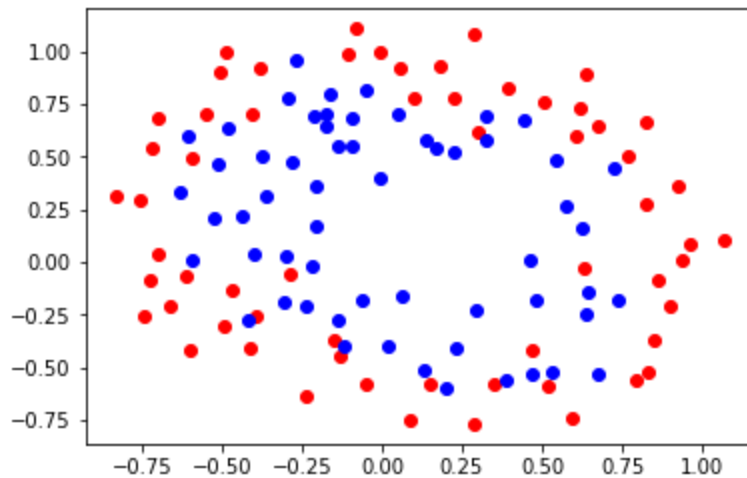
Use of Regulariser

The use of regulariser does changes the cost of the function but the change in cost is not substantial enough to produce any change in error percentage. Hence the error remains constant with using various regularizers.



Microchip Datasets

Graph below shows the nature of the dataset.

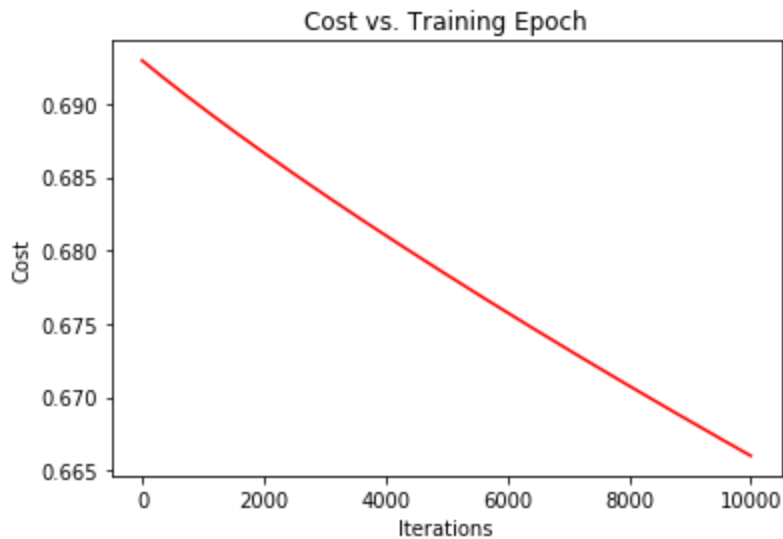


As can be seen using dataset the classification cannot be obtained by just a straight line as dataset needs a circular classifier. There we have to use more features instead of just two given features. Hence we also add squares of the both given features to the feature set. Hence in total we use 5 features for classification. Theta matrix is generated using Gradient Descent method and logistic regression. The results produced vary a lot depending on the training testing split and shuffling of dataset. Different shuffles of datasets produce different error/ accuracy percentages. **The accuracy percentage varies from 38 - 70 %**. Below are the accuracy results and graphs for a particular shuffled formation of dataset which produces good results.

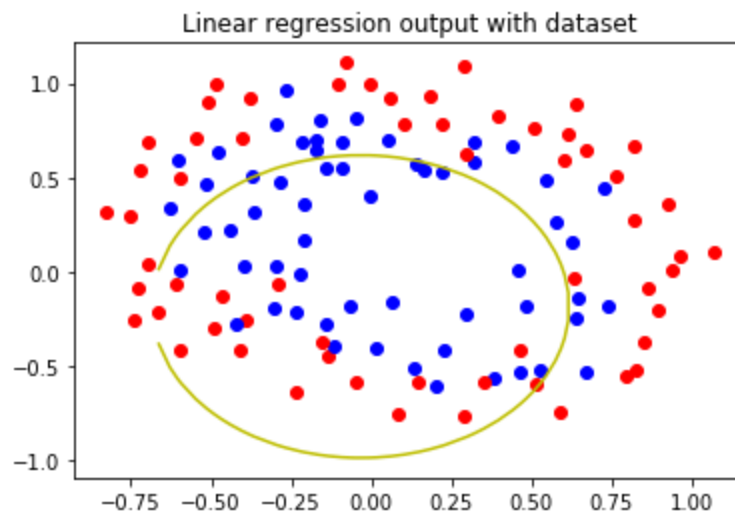
Accuracy : 68.57 %

Error : 31.42 %

Cost vs iteration curve for Gradient Descent Algorithm is shown below



Below is the graph showing dataset with classification



Use of Regulariser

The use of regulariser does changes the cost of the function, the change in cost in some cases also affects overall error percentage but percentage hasn't been seen less than that of without regulariser in my experimentation.