ISCO630E

*Analysis Report on **Assignment 4***

Name - Sashank Mishra
Roll Number - IIT2016515

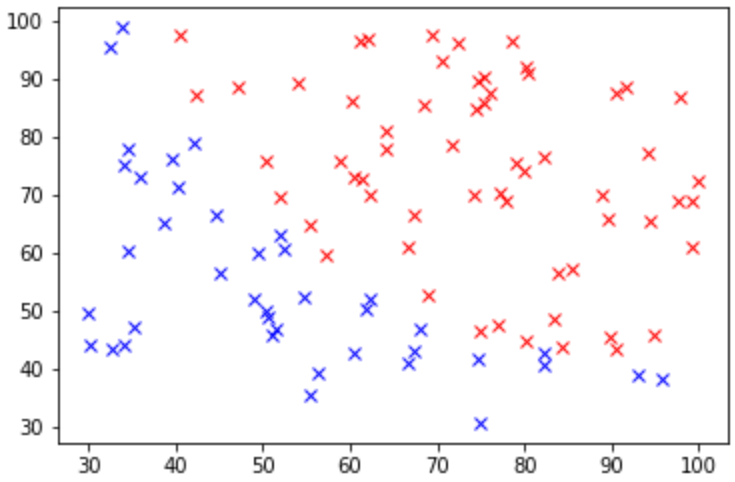
QUESTION 1

Dataset : Assignments Marks (in CSV format)
100 rows, 3 columns

marks1	marks2	selected
79.032736	75.344376	1
56.253818	39.261473	0
51.047752	45.822701	0
67.372028	42.838438	0
50.458160	75.809860	1

Data Preprocessing :

- Data splitted into features and prediction(X and Y, respectively)
- A Bias term added to feature set
- Data is visualized as follows:



- Data is normalized, and splitted into 70% training and 30% test set.

Analysis :

Without Regularization (lambda = 0)

No. of iterations	Accuracy (in %)
1000	93.33
100	93.33
50	86.67
10	66.67

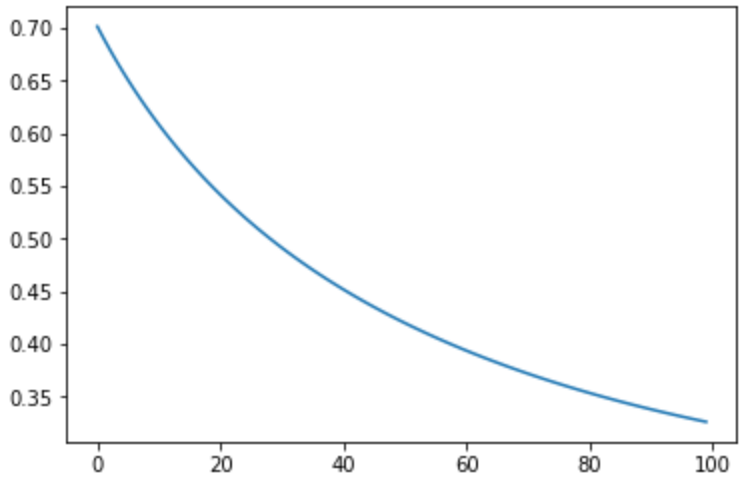
With Regularization

No. of iterations	Lambda	Accuracy (in %)
500	0.1	96.67
100	0.1	96.67
50	0.1	86.67
10	0.1	70
100	0.001	93.33
10	0.001	40
50	0.001	50

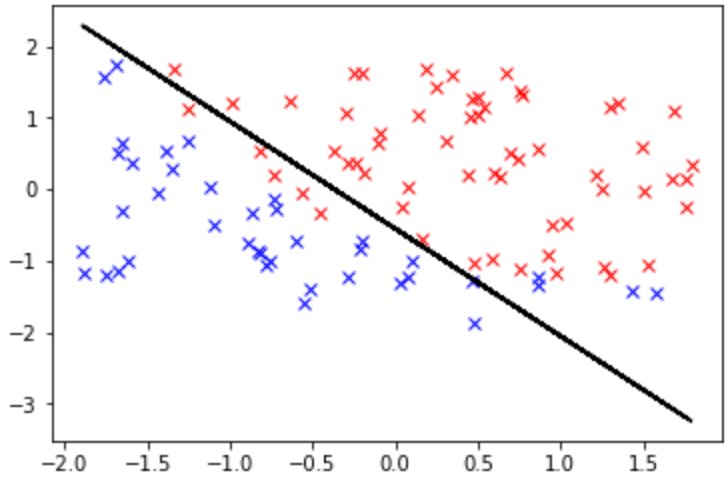
Best accuracy on test data, is achieved at lambda = 0.1, and running for 100 iterations.

The accuracy on test data = 96.67%
The accuracy on train data = 88.57%

The corresponding cost function is given as follows:



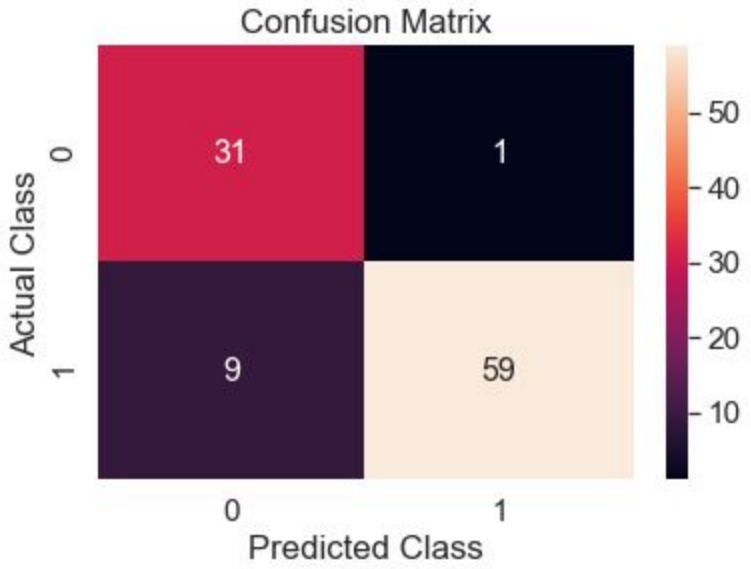
The data, with the predicted decision boundary, is given as follows:



Therefore, we obtain an almost perfect fit to the data, in lesser number of iterations than Gradient Descent.

The parameters obtained:-
W_0 = 0.63292359
W_1 = 1.71286238
W_2 = 1.14186915

We now obtain the Confusion Matrix on the complete dataset (test+train). The accuracy on the complete dataset is 90%.
The confusion matrix is then visualized as follows:



Other parameters calculated:
Precision - 98.33%
Recall - 86.7647%
F_1 Score - 0.921875