

ISCO630E

Analysis Report on Assignment 3

Name - Sashank Mishra

Roll Number - IIT2016515

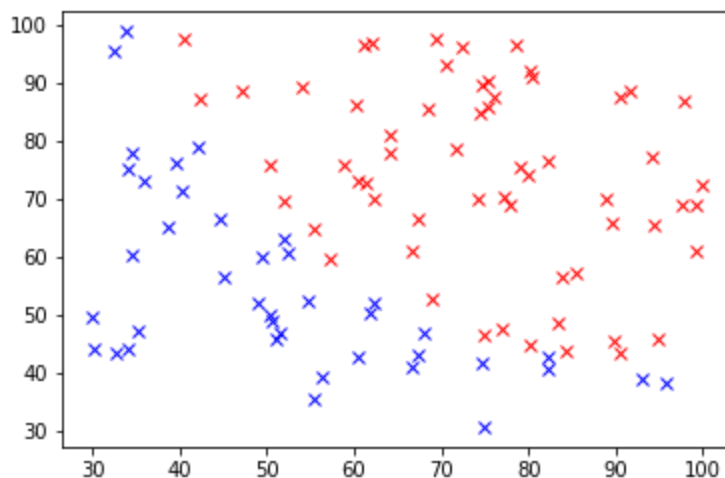
QUESTION 1

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	Learning Rate	Accuracy (in %)
10000	0.1	86.67
10000	0.01	83.33
10000	0.001	70.0
20000	0.1	90.0

Best accuracy achieved at $\alpha = 0.1$, and iterations = 20000(In attached code, the best accuracy achieved is 96.67%. This difference is due to the initial random initialization of parameters)

With Regularization

No. of iterations	Learning Rate	Lambda	Accuracy (in %)
10000	0.1	10	83.33
10000	0.1	100	76.67
10000	0.1	0.1	90.0
20000	0.1	10	86.67
20000	0.1	10	83.33
20000	0.1	100	76.67
100000	0.01	0.1	90.0
100000	0.001	0.1	70.0
200000	0.01	0.1	86.67
200000	0.001	0.1	83.33

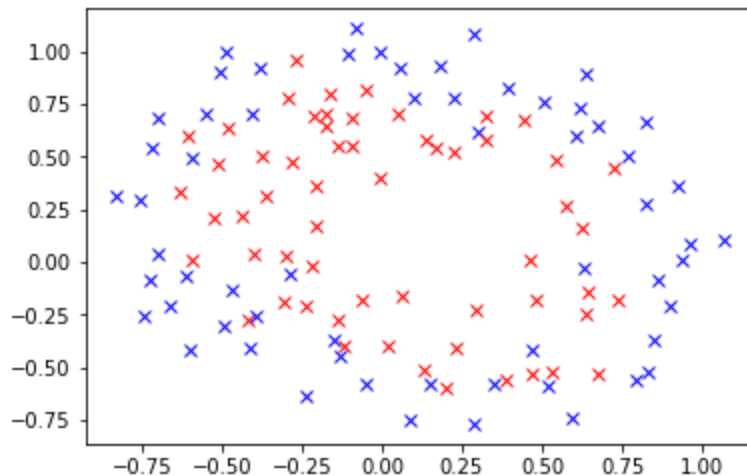
QUESTION 2

Dataset : Microchip Quality Results (in CSV format)
118 rows, 3 columns

test1	test2	result
0.54666	0.48757	1
0.82892	-0.52120	0
-0.13882	0.54605	1
-0.15035	-0.36769	0
0.59274	-0.74050	0

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.
- In later stage, 2 more features added, which are squared values of the original 2 columns. This has been done to add dependence of polynomial features in the model

Analysis :

Without Regularization ($\lambda = 0$)

No. of iterations = 20000, $\alpha = 0.05$, Accuracy = 47.22%

No. of iterations = 20000, $\alpha = 0.1$, Accuracy = 52.78%

With Regularization

No. of iterations	Learning Rate	Lambda	Accuracy (in %)
20000	0.1	0.1	55.56
20000	0.01	0.1	44.44
20000	0.001	0.1	41.67
20000	0.1	1	38.89
20000	0.1	10	47.22
20000	0.1	100	44.44
10000	0.1	0.1	38.89
100000	0.1	0.1	41.67

With Regularization and Polynomial Features added

No. of iterations	Learning Rate	Lambda	Accuracy (in %)
20000	0.1	0.1	69.44
20000	0.01	0.1	58.33
20000	0.001	0.1	68.89
20000	0.1	1	86.67
20000	0.1	10	69.44
20000	0.1	100	58.33
10000	0.1	0.1	58.33

100000	0.1	0.1	86.11
--------	-----	-----	-------

Therefore, better results are obtained on addition of polynomial features on this data, and using regularizer.

=====