

# COMPARATIVE ANALYSIS

## ASSIGNMENT 5 ML

**Kartic Choubey - IIT2018181**

Q1 .

In our dataset , we have two features  $x_1$  and  $x_2$  . As it is given in the question that we have to add 3 more features to our dataset, so we will add  $x_1^2$  ,  $x_2^2$  &  $x_1x_2$  . But before doing this , we will divide the dataset into 70% training and 30% testing . So after dividing the dataset and adding the features, we have to scale the features as there are negative features in the dataset. So I used min max scaling to scale the features so that all our features will lie in the range  $[0,1]$ . After applying the GDA for the testing data , we get our predicted y values .

For the performance Analysis, I used confusion matrix. My matrix comes out like this:

Testing Data = 36	Predicted : No	Predicted: Yes
Actual : No	18	4
Actual : Yes	0	14

By analysing confusion matrix, we can conclude that:

1. There are two predicting classes as our matrix is  $2 \times 2$ . Those Classes are "Yes" & "No"
2. The Classifier made a total of  $18 + 4 + 0 + 14 = 36$  predictions.
3. Out of those 91 cases,the classifier predicted "Yes"  $4 + 14 = 18$  times and "No"  $18 + 0 = 18$  times.
4. In reality, 14 patients in the sample have a heart disease and 22 do not.
5. True Negative(TN) = 18  
False Positive(FP) = 4  
False Negative(FN) = 0  
True Positive(TP) = 14
6. Accuracy =  $(TP + TN)/\text{Total} = (18 + 14)/36 = \mathbf{0.88}$

Q2.

For this , I applied the Box Muller Transformation to the given dataset to create a new dataset having gaussian distribution within the range of the given dataset. Then I applied the GDA to get my predicted Y values.

For the performance Analysis, I used confusion matrix. My matrix comes out like this:

Testing Data = 36	Predicted : No	Predicted: Yes
Actual : No	11	3
Actual : Yes	10	12

By analysing confusion matrix, we can conclude that:

7. There are two predicting classes as our matrix is 2\*2. Those Classes are "Yes" & "No"
8. The Classifier made a total of  $11 + 3 + 10 + 12 = 36$  predictions.
9. Out of those 91 cases,the classifier predicted "Yes"  $3 + 12 = 15$  times and "No"  $11 + 10 = 21$  times.
10. In reality, 22 patients in the sample have a heart disease and 14 do not.
11. True Negative(TN) = 11  
False Positive(FP) = 3  
False Negative(FN) = 10  
True Positive(TP) = 12
12. Accuracy =  $(TP + TN)/\text{Total} = (11 + 12)/36 = \mathbf{0.638}$ .

## RESULT:

Accuracy in Q1 comes out to be **88.88%** and **63.88%** in Q2. As the data is already normally distributed and box muller transformation works on the uniform continuous data, that's why our model doesn't give accurate results for Q2. As the dataset is normally distributed, it gives good accuracy for Q1. Also the data is very small . Box muller transformation is used for vast data.