FISHER'S LINEAR DISCRIMINANT ANALYSIS REPORT

INTRODUCTION :

In this problem,we tried to understand and develop a Fisher's discriminant model for classifying the positive and negative data points from the given dataset.

BASIC THEORY:

In Fisher's discriminant analysis, we project all the data points to a one dimensional vector. We try to find that vector W which will best separate the positive and negative data points on it.

We can do this by maximising the separation between the positive and negative points on it and minimising the standard deviation.

We will do this by maximising the separation of their respective means. Let m1 represent the sum of the means of all the positive points and m2 represent the sum of the means of all the negative points.

So our problem is max((m1-m2)). We also need to minimise the sum of the standard deviation of the positive and the negative training examples. Let s1 represent the standard deviation of the projected positive training examples and s2 represent the sum of the standard deviations of the negative training examples.

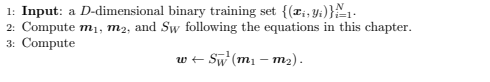
Our problem is min(s1^2 + s2^2). Hence the overall problem can be represented as

max((m1-m2)/(s1^2 + s2^2))

Solving this problem will give us our optimal W vector to which the points can be projected to.

METHODOLOGY:

Our implementation has following steps:



Where w is direction of the line on which we project points to classify

The data points are transformed into one dimension by using

.xn)

vector W is found through W = np.dot(Swinv,d1.mean()-d2.mean())

Results :

dataset1

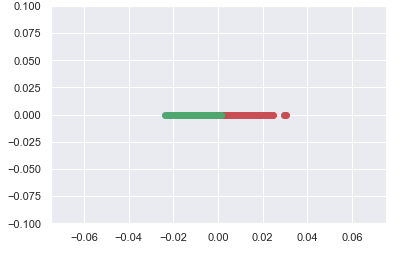
Direction of the vector on which points are projected

W = [0.00039786 0.01203286].

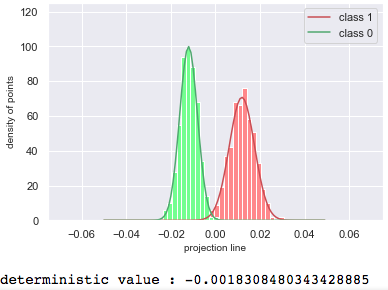
The accuracy obtained on the test set was = 99.3

The F score obtained on the test set was = 0.9929929929929929

The data points after being transformed into the one dimensional space have been plotted below with the red colour being the positive class and green colour being the negative class.



Normal distribution of each class has been plotted below.



Threshold value obtained(intersection of the two normal curves) was = -0.0018308480343428885

dataset2

Direction of the vector on which points are projected

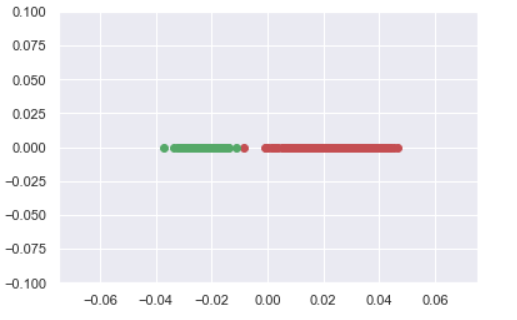
W = [-0.00014845 -0.0004129 0.02263587]

.

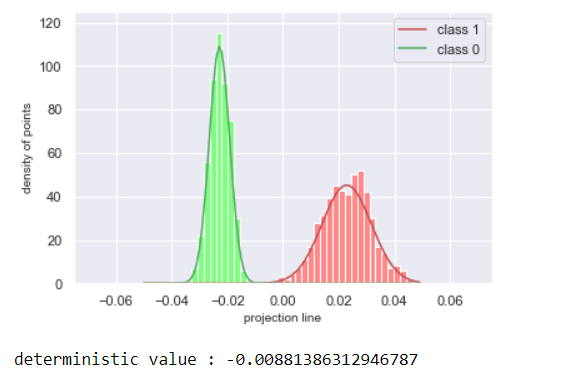
The accuracy obtained on the test set was = 100

The F score obtained on the test set was =1

The data points after being transformed into the one dimensional space have been plotted below with the red colour being the positive class and green colour being the negative class.



Normal distribution of each class has been plotted below.



Threshold value obtained(intersection of the two normal curves) was = -0.00881386312946787