Rishabh Baid 2018A7PS0189H
Pragya Sinha 2017B1A3PS1749H
Akshat Bajpai 2018A7PS0498H

Assignment-1A:
Fischer's Linear Discriminant

Introduction:

Fisher's LDA method is one of the popular methods used in the binary and multi-classification problem. This involves reducing parameters to reduce the number of dimensions of a point from 'n' to 2. This way, we would have a linear distribution of points belonging to all the classes. By finding the mean of the set of points belonging to each class, we can find the discriminating point between any two classes. This finding of the discriminating point works under the assumption that each class's set of points follows a Gaussian Distribution. So, finding the discriminating point would essentially mean that we would be finding the point of intersection of the Normal distribution of both sets of points. When a new testing point is given, we first project it onto our line and then see if it is above or below the discriminating point(in the case of Binary classification). To test the training and validation accuracy, there would be no necessity for a test-train split in this algorithm.

Implementation:

In the beginning, we extracted the training examples provided to us in the form of Pandas Dataframe. We were then able to obtain the parameters and final result in two different arrays.

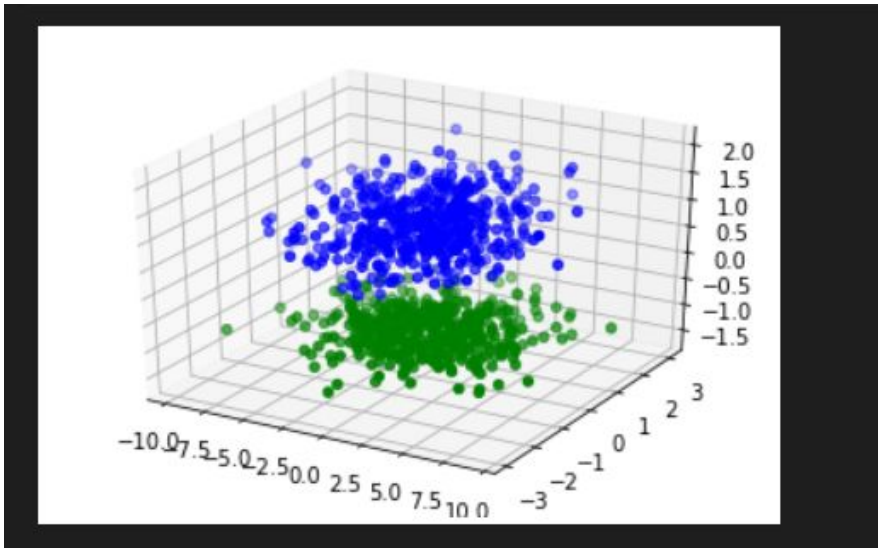
After extracting data from the Pandas Dataframe, we found the inverse of the sum of variances and differences of the mean. Once the plane was obtained, we derived a line perpendicular to the plane, which would be the one onto which all the points would be projected. Later, we find the mean of means of

Rishabh Baid 2018A7PS0189H
Pragya Sinha 2017B1A3PS1749H
Akshat Bajpai 2018A7PS0498H

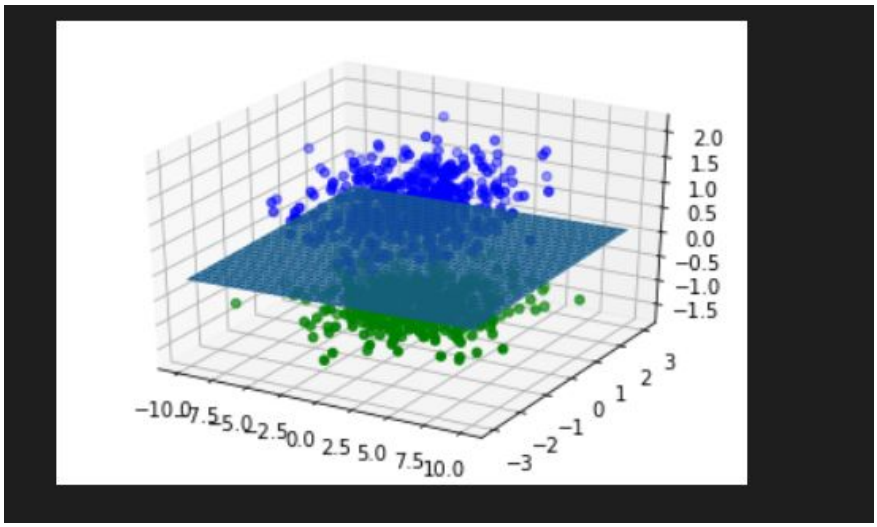
both distributions. This resulted in a discriminating point lying along the previously obtained line. We have assumed that the data points belonging to each class lying on the obtained line follow a Normal distribution.
We have assumed that each class's points follow a Normal Distribution, so finding the mean of means would essentially mean finding out the intersection point of the standard distribution graphs of both classes.

Output and Graphs:

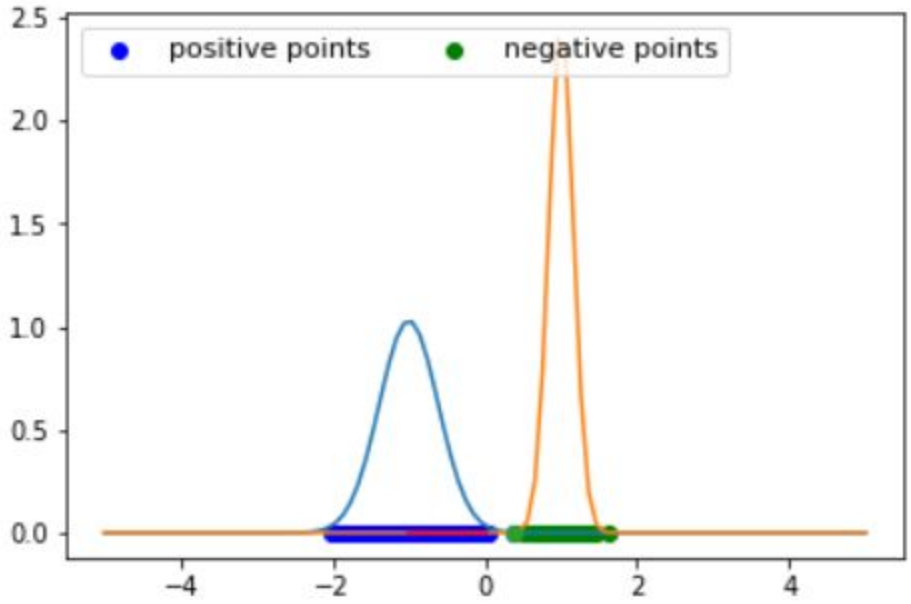
Cluster Plots



Rishabh Baid 2018A7PS0189H
Pragya Sinha 2017B1A3PS1749H
Akshat Bajpai 2018A7PS0498H



Fischer's Discriminant



Normal Distribution fitted over FLD