Fischer Discriminant

Design decisions/Approach taken/Algorithm flow:

- First , the data was split into 2 sets having data points in the ratio 80:20 for training and testing data respectively
- The mean of data points was calculated for each of class 0 and class 1.
- The mean vector and covariance matrix (S_W) were evaluated.

  The covariance matrix was evaluated using the formula given below:

$$(\sum_{1}^{n} (x_n - m_1)^2)/n_1 + (\sum_{1}^{n} (x_n - m_2)^2)/n_2$$

- The weight vector was evaluated by the dot product of the inverse of the covariance matrix and the mean vector . This was done such that the weight vector that we get is a unit vector.
- Then, the test data was projected on the weight vector to get the data in one dimension. The scatter plot showed the data in the original space.

- After projecting we found the discriminant point, by plotting the normal curves of both the classes. The point was calculated by knowing the point at which both the curves intersected.
- Other parameters such as accuracy, F-score, precision, and recall were also calculated.

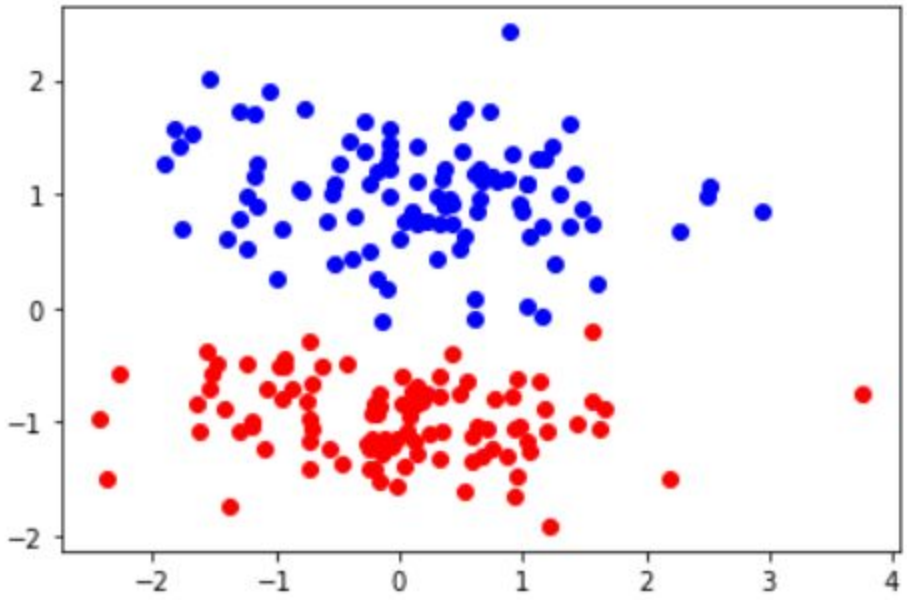
Results for data set 1 and data set 2(Obtained for **test data**)

	Data set 1	**Data set 2**
Accuracy	1	0.995
Precision	1	1
Recall	1	0.99
F-score	1	0.995
Weight vector	array([-0.01955343, -0.98384438])	array([0.005938 , 0.01997482, -0.99277946])
Discriminant point	0.157	0.377

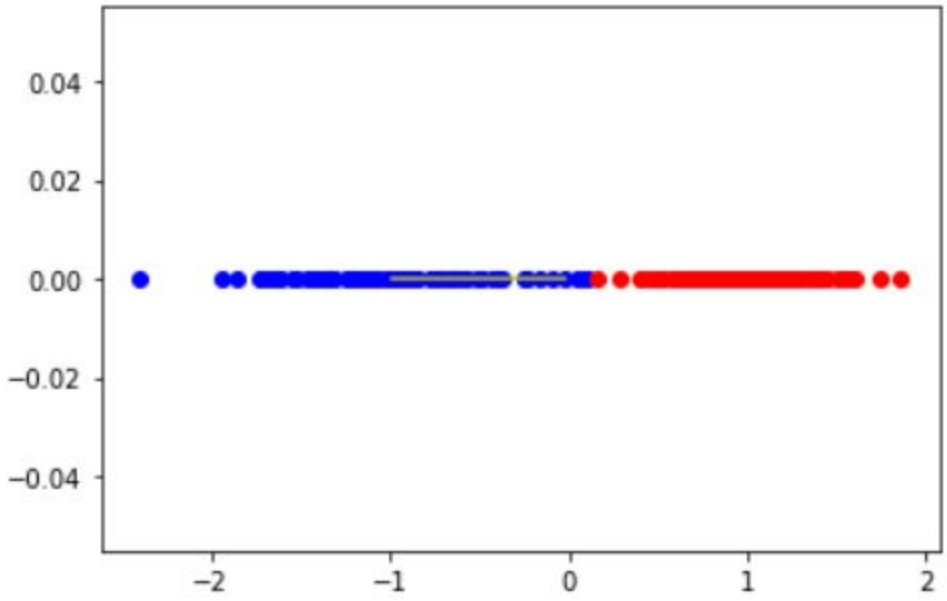
Visualizations(Obtained for test data only)

1. For data set 1 (Fischer in 2D)

- Scatter plot(Visualization in original dimension)

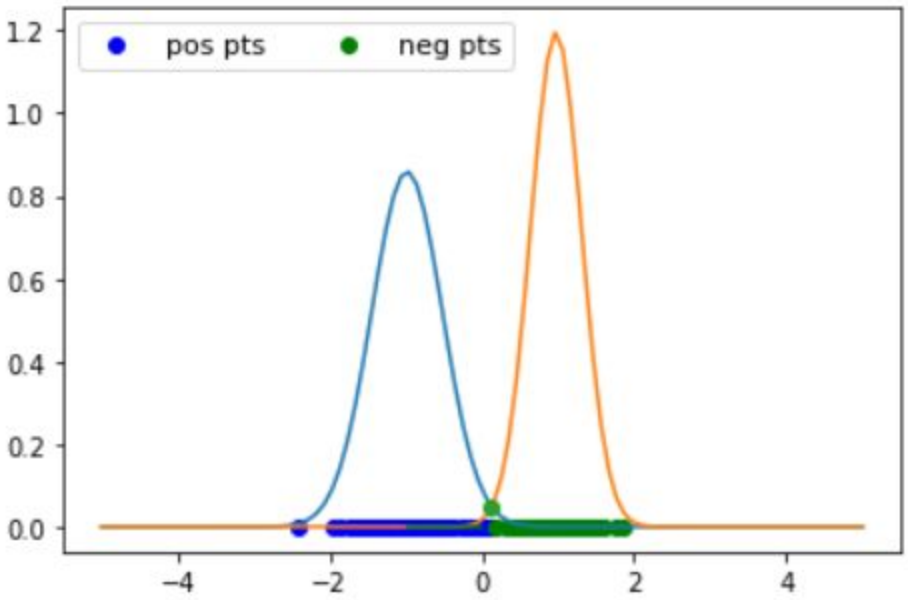


- Visualization of data in one dimension



This is the visualization of data after projecting them on one dimension (ie. the weight vector).

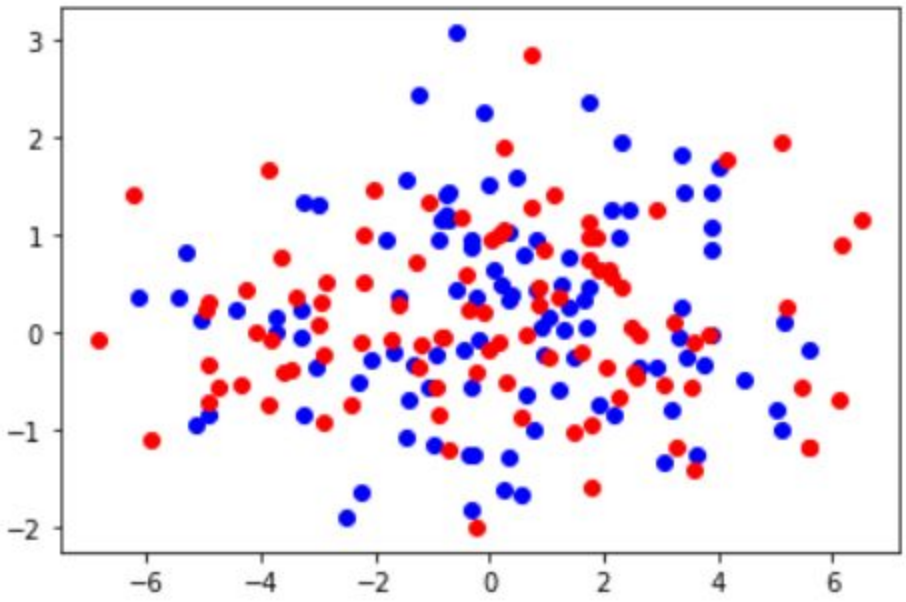
- Plot of normal curves



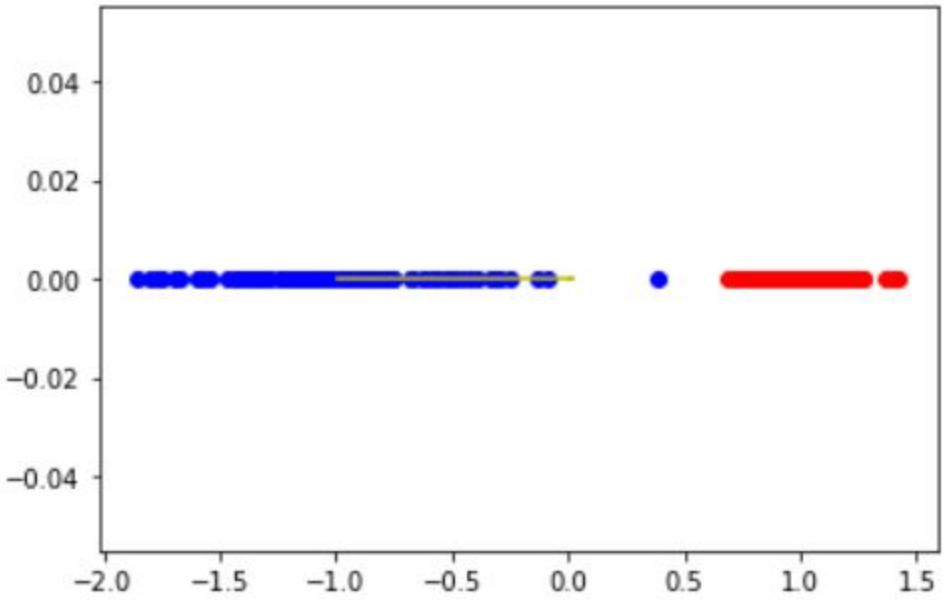
This is the normal plot of the data having both the classes. Here the x coordinate of the point of intersection of both the normal curves is the discriminant point. The blue points belong to the negative class and the green points belong to the positive class.

2. For data set 2 (Fischer in 3D)

- Scatter plot(Visualization in original dimension)

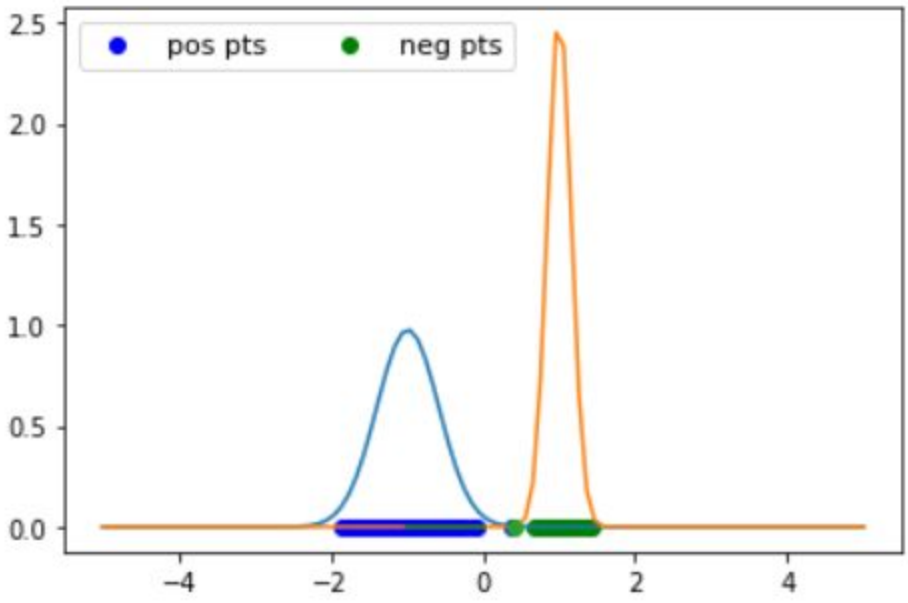


- Visualization in one dimension



This is the visualization of data after projecting it on the weight vector.

- Plot of normal curves



This is the normal plot of the data having both the classes. Here the x coordinate of the point of intersection of both the normal curves is the discriminant point. Since, the positive and negative class points are linearly separable(as can be observed from the above graph), the Fischer Discriminant correctly classifies the points to their respective classes. The blue points belong to the negative class and the green points belong to the positive class.

Conclusions

From the visualization plots, it can be clearly seen that the two classes of points, that is, the positive and negative points are linearly separable. So the Fischer Discriminant algorithm for the above two datasets holds good approximation and the discriminant point that we got is able to properly classify the points. This is the reason we are getting perfect results with very high accuracy and very high F-score.