Machine Learning
Assignment 1

Team Details
Tanmay Kulkarni 2015B3A70647H
Aditya Srikanth 2016A7PS0091H
Ch Vishal 2015B5A70605H

How to Run

1. Make sure the dataset csv files are in a folder named “ML - Assignment 1 - Datasets” inside the directory where the main.py script is.
2. Make sure all the dependencies (numpy, pandas, matplotlib, scipy) are installed.
3. On a terminal, type “python3 main.py”
4. Upon the prompt, enter “1” for the Perceptron algorithm and enter “2” for the Linear Discriminant Analysis.
5. Upon the next prompt enter “i” for using ith dataset in the algorithm
6. If everything is right, the plots or the animation will appear, along with the results in the terminal.

Note: The blue points represent class ‘0’ and the red points represent class ‘1. For the linear discriminant analysis, the projections on the w vector are in the respective class colours, the threshold point is in green, the vector w is yellow and the gaussians are in black.

Note: Results of the linear discriminant analysis on each of the datasets is saved as a .jpeg file. The files are named “lda i” for run on the ith dataset. It is also available to view while running the code.

Note: Results of the perceptron algorithm execution on each of the datasets is saved as a gif and an mp4 file. The files are saved as “perceptron i” for run on the ith dataset. It is also available to view while running the algorithm.

Fisher Linear Discriminant Analysis

Observations

Both the gaussians of classes 1 and 2 meet close to the w vector in datasets 1 and 3, this is because the classes are separable. Dataset 2 is not fully separable using a linear boundary therefore the gaussians meet slightly far away.

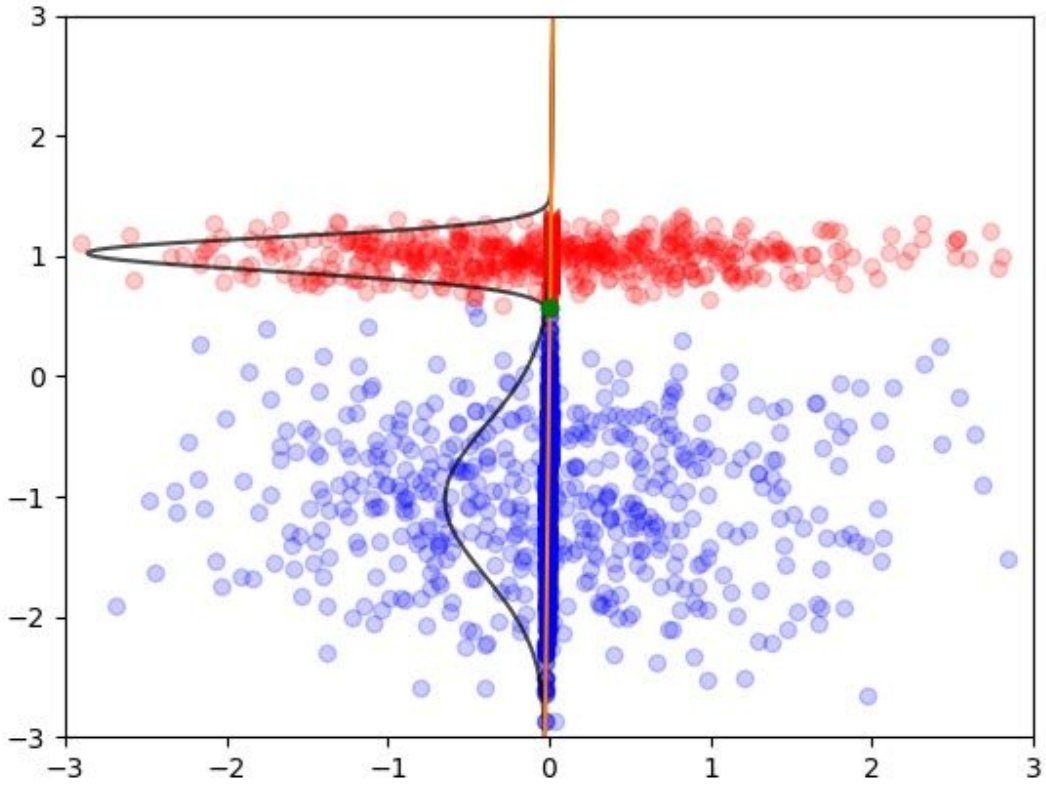
LDA gives a point estimate rather than a probabilistic interpretation which could give more knowledge for the data points that tend to be near the boundary, that is, we can get the degree to which it belongs to a particular class.

Results

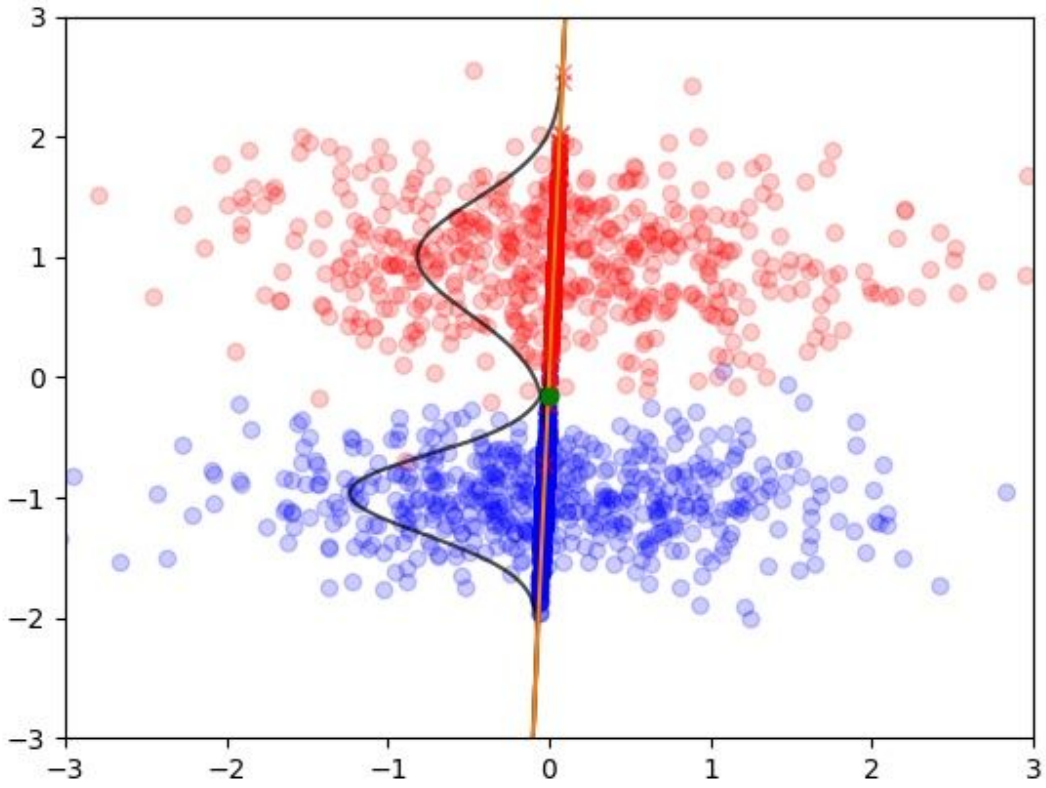
Attribute\Dataset	Dataset 1	Dataset 2	Dataset 3
Weights vector	[0.00818276 0.99996652]	[0.03314521 0.99945055]	[-0.01765168 0.9998442]
Separation point	[0.00466015 0.56948913]	[-0.00505303 -0.1523674]	[0.00692927 -0.39249443]

Plots

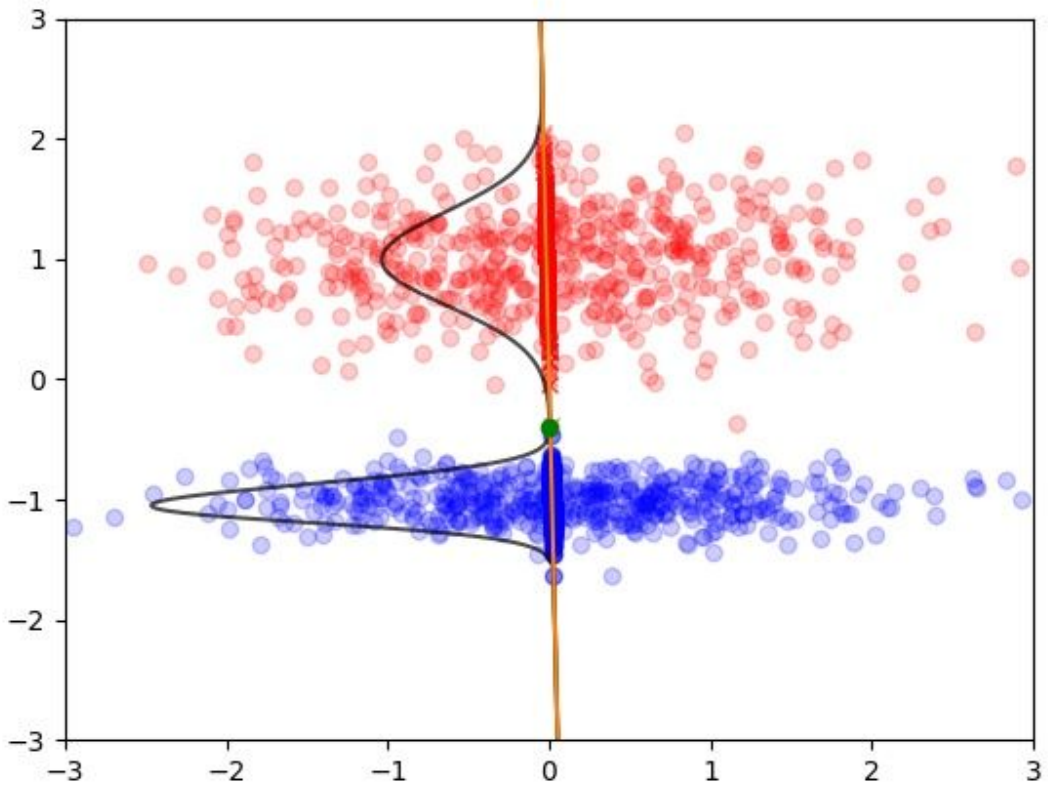
Dataset 1



Dataset 2



Dataset 3



Perceptron Algorithm

Observation

The algorithm is able to correctly draw a dividing line because the classes are truly separated in Dataset 1 & 3. However as we can see, dataset 2 is not fully separable with a linear boundary and hence the algorithm misclassifies a few points on either class and the error doesn't drop to 0.

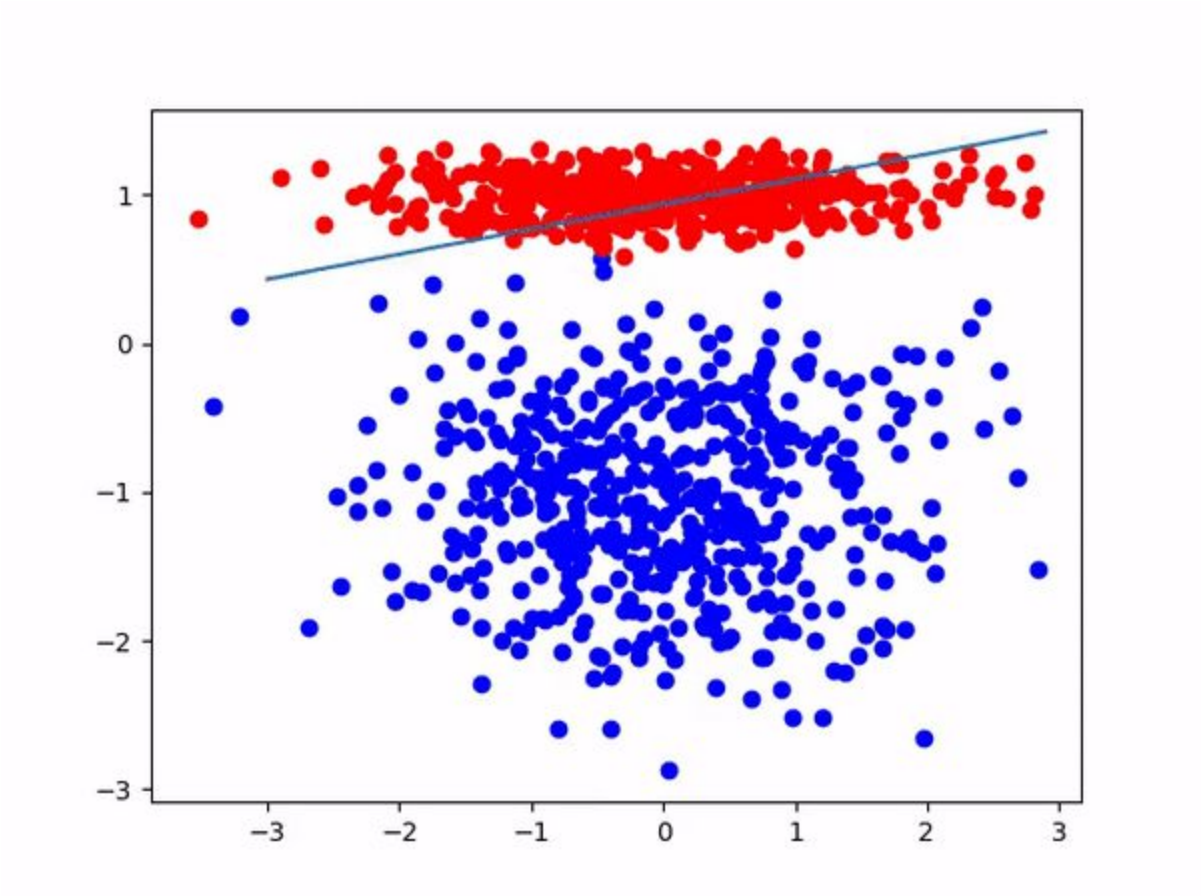
Perceptron has the tendency to overfit the data, which can be seen in the case of dataset 3 in which, there seems to be one outlier point which may have been generated due to noise. One possible suggestion is to regulariase the weights so as to reduce their freedom and make them more capable of dealing with noise

Results

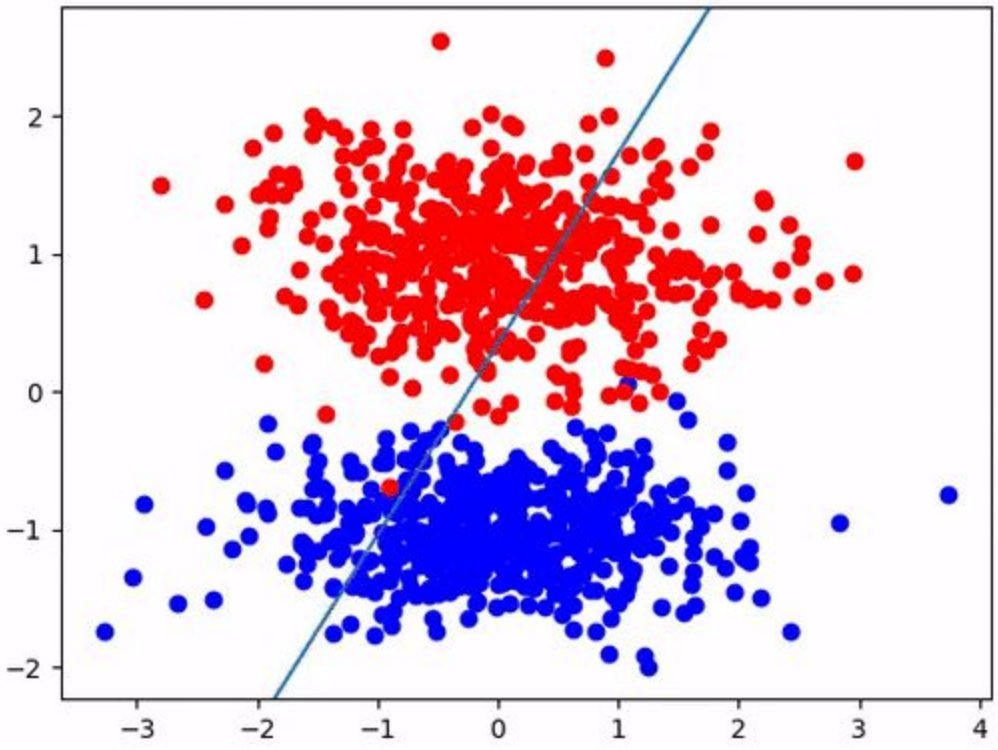
Attribute\Dataset	Dataset 1	Dataset 2	Dataset 3
Weights vector	[-26.57305231, 4.33533195, 50.37237729]	[3.26963636, -1.6149921, 18.34529555]	[6.32722623, 6.90127902, 38.35524323]

Plots

Dataset 1



Dataset 2



Dataset 3