

# Report

## Assignment on Dimensionality Reduction Techniques

Sankalp Rajendran

IIT2019173

[IPYNB LINK](#)

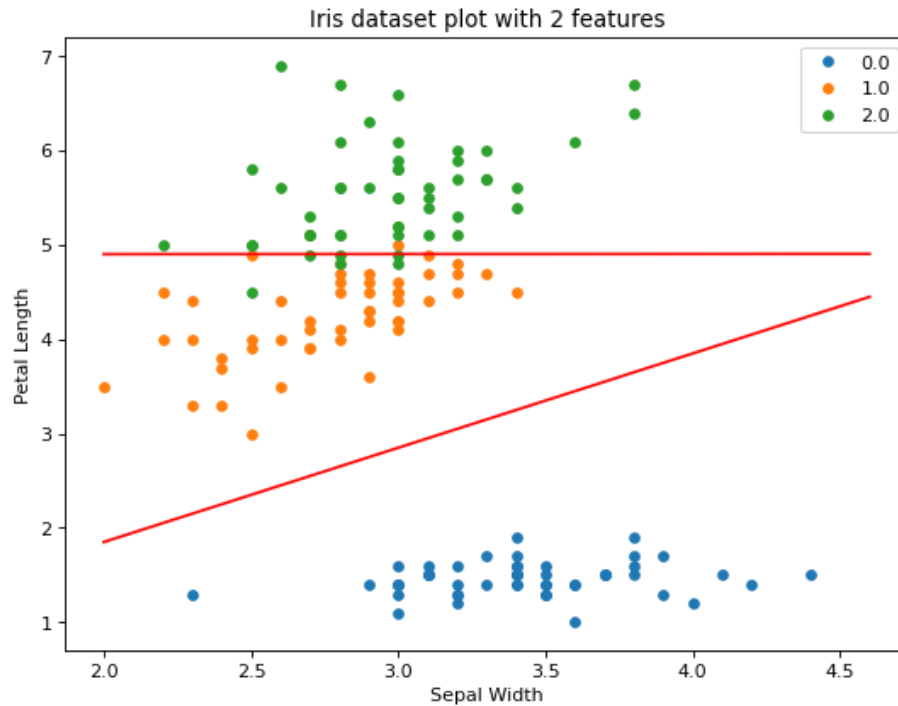
**Linear Discriminant Analysis:** LDA is a supervised machine learning method that is used to separate two groups/classes. The main idea of linear discriminant analysis (LDA) is to maximize the separability between the two groups by creating a new linear axis and projecting the data points on that axis, so that we can make the best decision to classify them.

**Principal Component Analysis:** PCA is an unsupervised machine learning method that is used for dimensionality reduction. The main idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of many variables correlated with each other. This is done by transforming the variables into a new set of variables which is a combination of variables or attributes from our original dataset in such a way that maximum variation is retained.

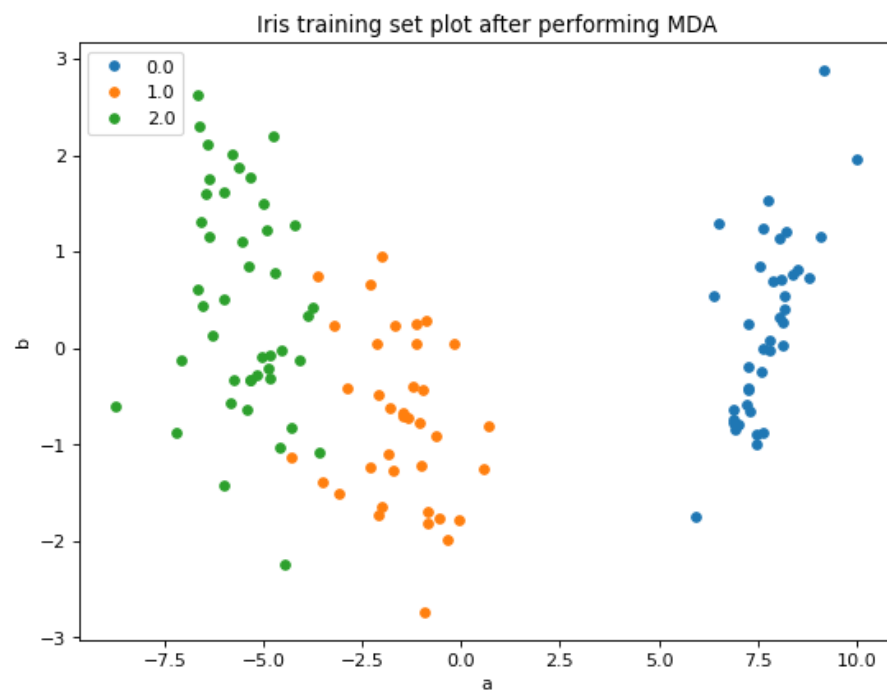
**T-Distributed Stochastic Neighbour Embedding:** T-SNE is an unsupervised machine learning method that is used to visualize the higher dimensional data in low dimensions. T-SNE is used for designing/implementation and can bring down any feature space into 2-D feature space.

## **Problem 1**

**1- Method 1-Projecting Iris dataset with 2 features and drawing lines to classify.**



## **2- Method 2- Applying MDA on Iris Dataset**



**3-** Method 2 is better than method 1 because when the data has more than 3 or 4 dimensions, then we can't use linear classifiers as it can't be visualised on a coordinate system but Multiple discriminant analysis overcomes this problem by eliminating the features which contribute very little in determining the label.

## **Problem 2**

**1-** Plotting on 2D plane after dimensionality reduction using 100 Training data points.

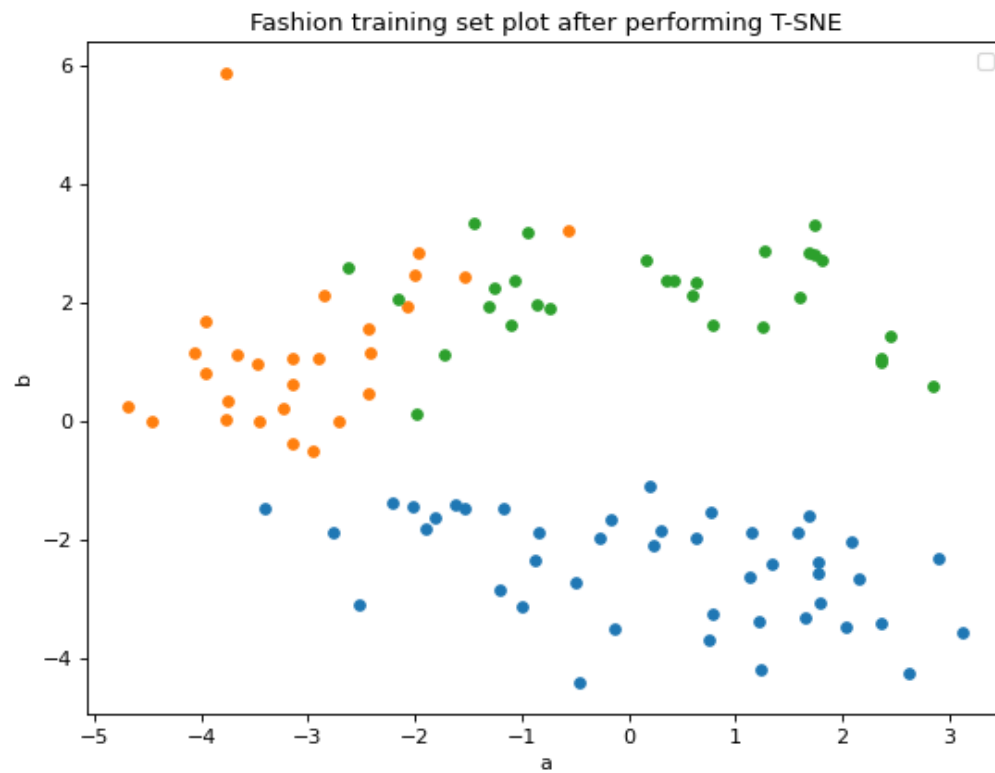
### **MDA**



## PCA



## T-SNE



## 2- Comparing accuracy of MDA, PCA and T-SNE on testing set.

I used a KNN classifier and trained it on the training dataset and used the predictor to predict on the testing dataset containing 50 points.

Method	n_neighbors	Accuracy
MDA	2	98%
MDA	3	94%
MDA	4	98%
PCA	2	88%
PCA	3	90%
PCA	4	88%
T-SNE(perplexity=50)	2	32%
T-SNE(perplexity=50)	3	38%
T-SNE(perplexity=50)	4	40%

**3-** According to me, MDA is the best Dimensionality Reduction Technique as it gives the best accuracy of above 90% on the testing set in Problem 2.