

# Assignment on Dimensionality Reduction

IIT2019044  
Shreesh Swaraj  
Semester 5

10/10/2021

# Google Colab Link for Code

Problem 1 -

[https://colab.research.google.com/drive/1iw1QWLbqjZaeKTE6cndnJuhL0\\_QiumHz?usp=sharing](https://colab.research.google.com/drive/1iw1QWLbqjZaeKTE6cndnJuhL0_QiumHz?usp=sharing)

Problem 2 -

[https://colab.research.google.com/drive/1u7vqNxaibDILIX\\_Muowhr\\_DtbTvZpGh8?usp=sharing](https://colab.research.google.com/drive/1u7vqNxaibDILIX_Muowhr_DtbTvZpGh8?usp=sharing)

# Problem 1

**Method 1:**

**(i) From Iris flower data set take only two features (sepal width and petal length). Project the labelled data on a 2D graph. Try to classify this data by drawing linear boundaries intuitively.**

**(15 marks)**

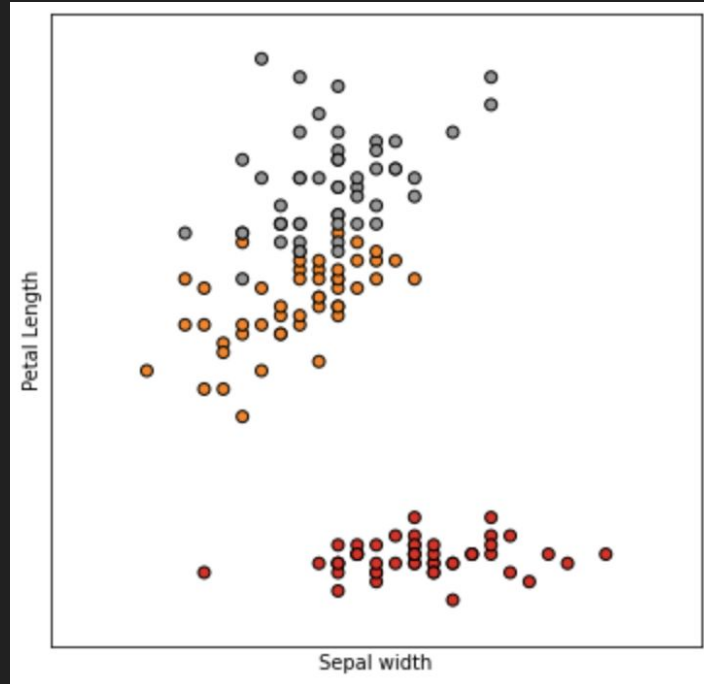
**Method 2:**

**(ii) Now take all the features and apply MDA on this Iris flower data set and plot the labelled data on 2D graph(15 marks)**

**(iii) Show the plots of above two methods and write in your own words why method2 works better than method1 (10 marks)**

# Problem 1 - Part 1

Plotting 'sepal width' vs 'petal length'



# Using Logistic Regression for 3 classes

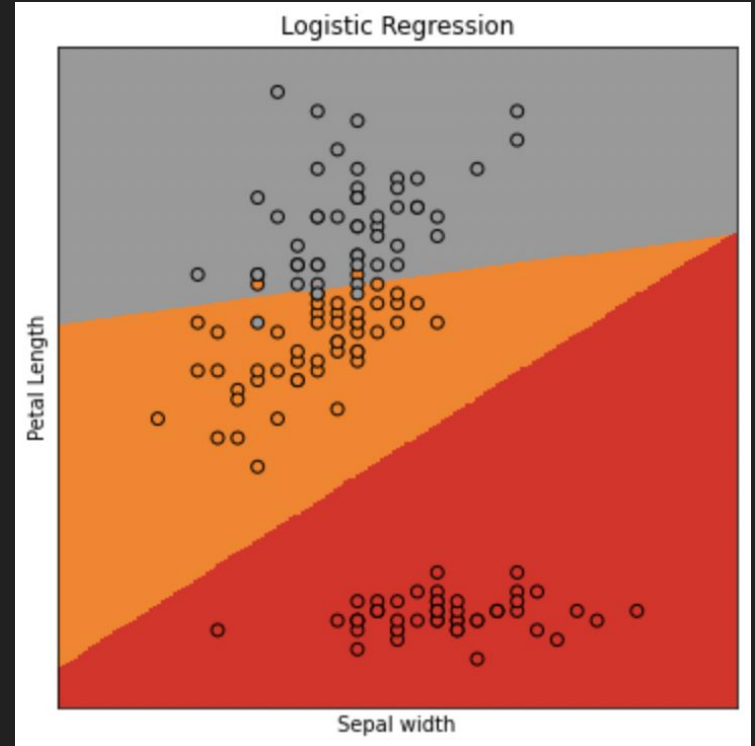
```
Accuracy = 0.9533333333333334
```

```
Confusion Matrix:
```

```
[[50  0  0]
```

```
 [ 0 47  3]
```

```
 [ 0  4 46]]
```



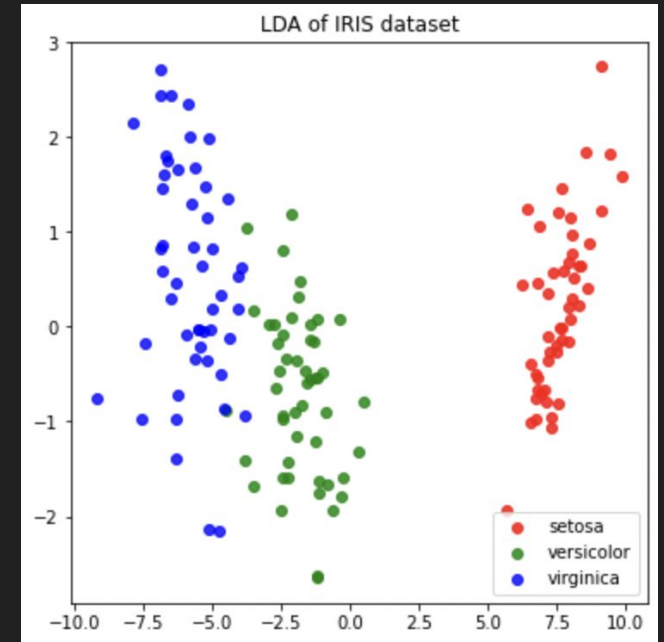
# Problem 1 - Part 2

## Using MDA on IRIS Dataset

Accuracy = 0.98

Confusion Matrix:

```
[[50  0  0]
 [ 0 48  2]
 [ 0  1 49]]
```



## Problem 1 - Part 3

We can see that the accuracy using dimensionality reduction technique of LDA (Linear Discriminant Analysis) is much better than using Logistic Regression for 3 classes.

# Problem 2

From MNIST Fashion data take 100 data points belonging to classes Sneaker, Pullover and Ankle boot.

(i) Project these data points on a 2D plane using data reduction techniques PCA, T-SNE and MDA.(25 marks)

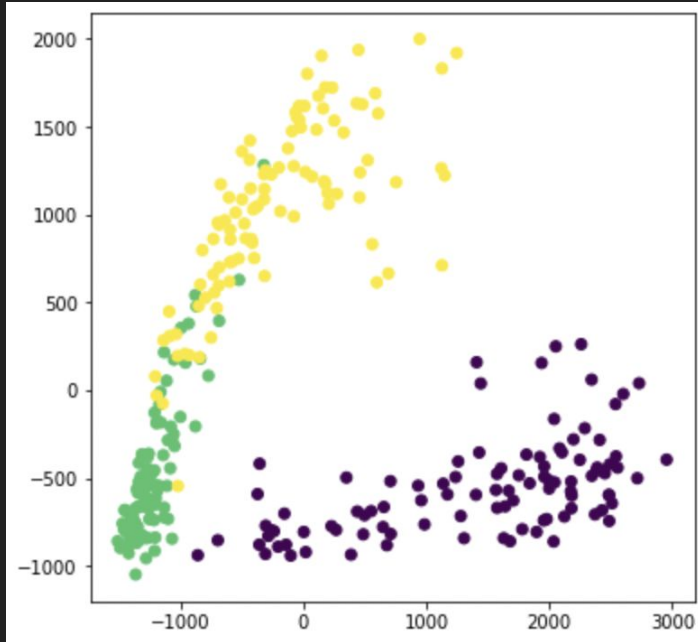
(ii) Take 50 data points belonging to classes Sneaker, Pullover and Ankle boot from the test data set and compare the performance of these techniques(PCA, T-SNE, MDA) in terms of accuracy.  
(25 marks)

(iii) According to your view what is the best dimensionality reduction technique for solving this problem and justify your answer.(10 marks)

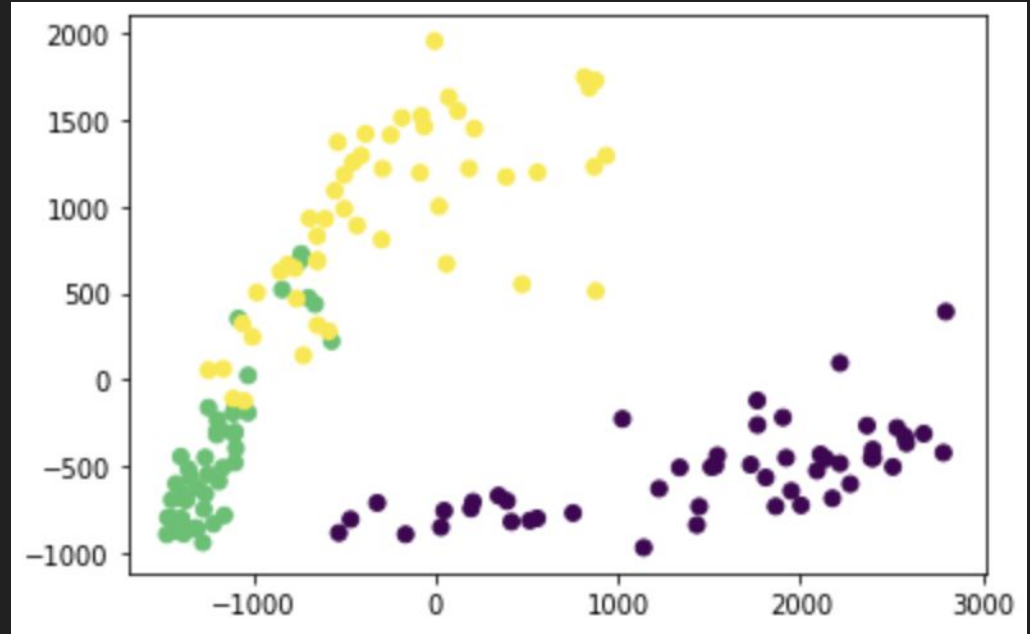


# Problem 2 - Part 1/Part 2

## PCA



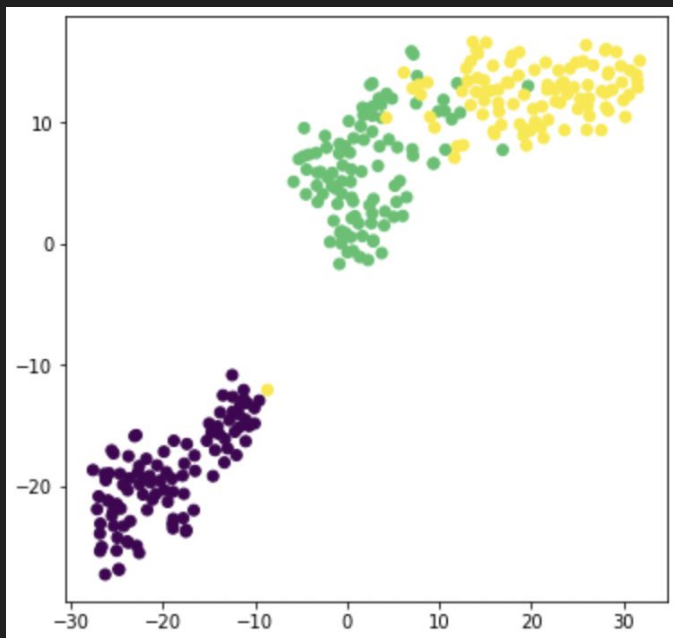
Plot on Training set



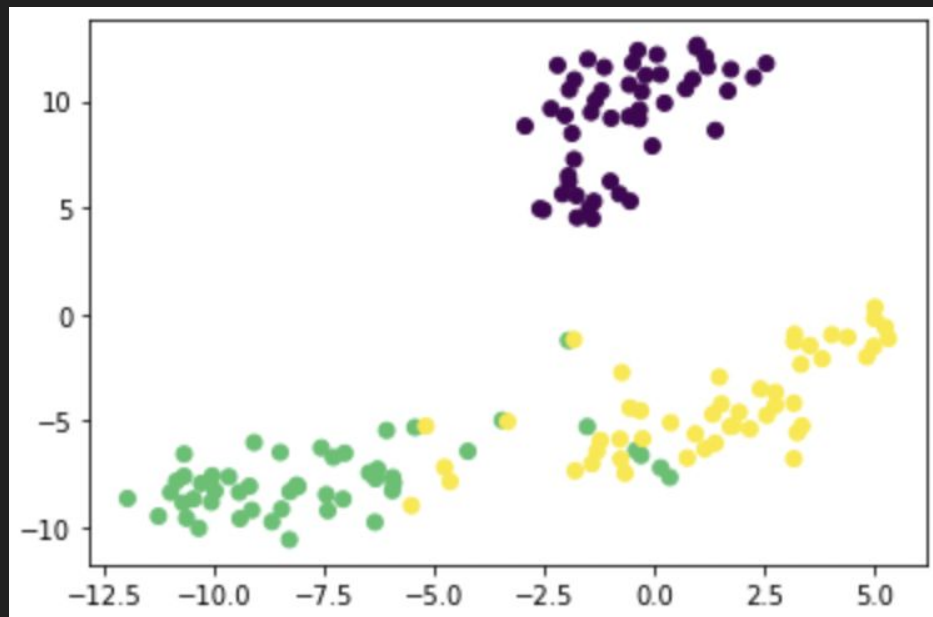
Plot on Testing set

Accuracy: 0.913333

# T-SNE

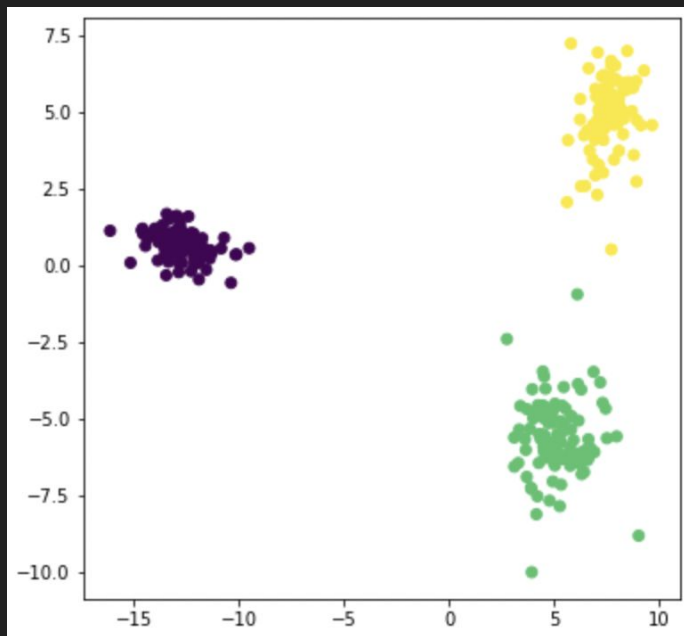


**Plot on Training set**

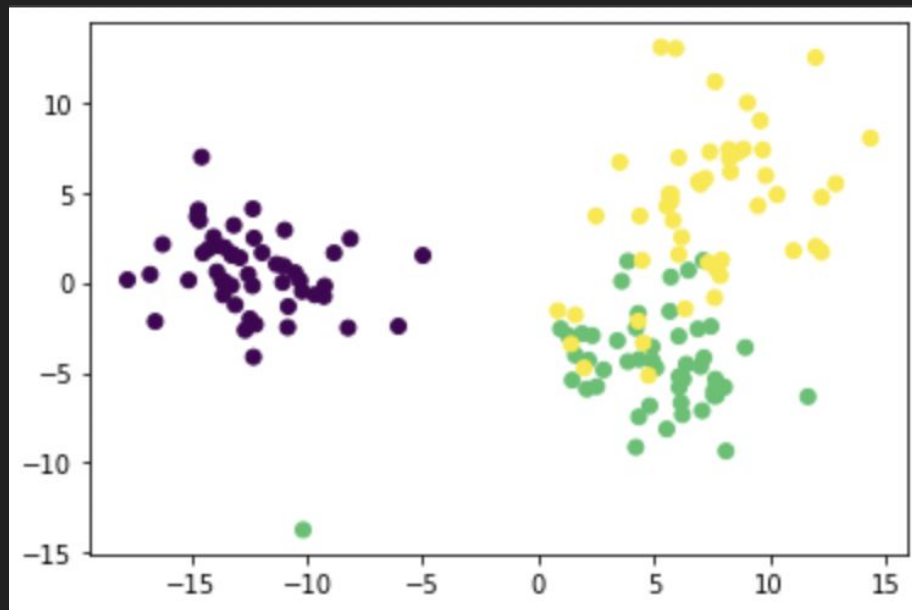


**Plot on Testing set**  
**Accuracy: 0.12668**

# MDA



Plot on Training set



Plot on Testing set  
Accuracy: 0.9

## Problem 2 - Part 3

**PCA works best among the three given dimensionality reduction techniques namely PCA, T-SNE and MDA.**

**T-SNE does not have any transform method. T-SNE is also an unsupervised method for dimensionality reduction/visualization, so it does not really work with a 'train' and 'test'.**

**We make use of `fit_transform` to make the desired transformation and then use classification techniques to get the desired classification results.**