**Comparison of LDA and PCA 2D projection on Iris dataset**

**Dataset: Iris Dataset**

**Preamble**: Principal Component Analysis (PCA) applied identifies the combination of

attributes (principal components, or directions in the feature space) that account for the most

variance in the data. Here we plot the different samples on the first 2 principal components.

Linear Discriminant Analysis (LDA) tries to identify attributes that account for the most

variance between classes. In particular, LDA, in contrast to PCA, is a supervised method, using known class labels.

**Problem:** In this section we will implement PCA and LDA with the help of Python's Scikit-

Learn library and compare both the techniques. We will follow the classic machine learning

pipeline where we will first import libraries and dataset, perform exploratory data analysis and preprocessing, perform PCA to find out optimal number of features, and finally train our models, make predictions and evaluate accuracies.

**Exercise I**: Study and Execute the given PCAvsLDA plot script. Write your observations.

**Exercise II:** Complete the following template for LDA to predict and evaluate accuracies using only optimal number of features calculated. Compare the accuracy with PCA.

# import suitable libraries

# import Dataset

#Preprocessing(Hint: Divide the dataset into a feature set and corresponding labels.)

(Also divide dataset into training and test set)

#Normalize feature set.( Hint: StandardScaler() function from Scikit-learn Library)

#Perform LDA

( Hint: Initialize the LDA class by passing the number of components to the constructor.

Call the fit( ) and then transform methods by passing the feature set to these methods. )

**#**Train the model

#Predict the label and Evaluate accuracy

References:

1. https://www.kaggle.com/arshid/iris-flower-dataset

2. https://scikit-learn.org/stable/modules/generated/sklearn.discriminant\_analysis.LinearDiscriminantAnalysis.html