**ML ASSIGNMENT 5**

**Name: Vathumilli Manasa**

**ID: 700745467**

**Question-1: Principal Component Analysis**

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

1. **Apply PCA on CC dataset**

**Table

Description automatically generated**

1. **Apply k-means algorithm on the PCA result and report your observation if the silhouette score has improved or not?**

**Text

Description automatically generated**

**Table

Description automatically generated**

1. **Perform Scaling+PCA+K-Means and report performance.**

**Text

Description automatically generated**

**Table

Description automatically generated**

**Graphical user interface, text, application

Description automatically generated**

**Table

Description automatically generated**

**Table

Description automatically generated**

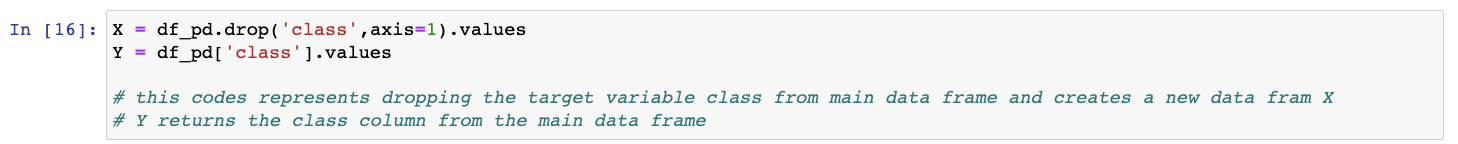
**Question-2: Use pd\_speech\_features.csv**

**Table

Description automatically generated**

**Graphical user interface, application

Description automatically generated**

****

1. **Perform Scaling**

**Text, chat or text message

Description automatically generated**

1. **Apply PCA (k=3)**

**Table

Description automatically generated**

****

**C. Use SVM to report performance**

**Text

Description automatically generated with low confidence**

**Question-3: Apply Linear Discriminant Analysis (LDA) on Iris.csv dataset to reduce dimensionality of data to k=2.**

**A classifier with a linear decision boundary, generated by fitting class conditional densities to the data and using Bayes’ rule.**

**A picture containing graphical user interface

Description automatically generated**

**Question-4: Briefly identify the difference between PCA and LDA**

PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) are both popular techniques in machine learning for dimensionality reduction. However, they have different purposes and methods:

Purpose: PCA is used for unsupervised learning and finds the directions of maximum variance in a dataset. It reduces the number of features by transforming the original dataset into a new coordinate system, where the features are uncorrelated and sorted by their variance. PCA is commonly used for data compression, visualization, and noise reduction. LDA, on the other hand, is used for supervised learning and aims to find the linear combinations of features that best separate the classes. It reduces the number of features by projecting the original dataset onto a lower-dimensional space while maximizing the class separability. LDA is commonly used for feature extraction, pattern recognition, and classification.

Method: PCA operates by finding the eigenvectors and eigenvalues of the covariance matrix of the data. The eigenvectors represent the directions of maximum variance, and the eigenvalues represent the amount of variance explained by each eigenvector. PCA selects the top k eigenvectors, where k is the desired dimensionality of the reduced dataset. LDA, on the other hand, maximizes the between-class scatter and minimizes the within-class scatter of the data. It involves finding the eigenvectors and eigenvalues of the product of two matrices: the between-class scatter matrix and the within-class scatter matrix. LDA selects the top k eigenvectors that correspond to the largest eigenvalues.

**Video Link:** <https://drive.google.com/file/d/1bO4PjeQqkdmY9Rp4aibTV7f1IY7PUJSB/view?usp=sharing>

**Git Hub Link:** <https://github.com/Manasav17/ML_Assignment5>

**Source code**: <http://localhost:8888/notebooks/Desktop/Ml/ML_Assignment_5.ipynb>