**ASSIGNMENT-5**

**NAME: VAISHNAVI MADIRI**

**700#: 700735078**

**COURSE: MACHINE LEARNING**

**1. Principal Component Analysis**

**a. Apply PCA on CC dataset.**

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

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

**OUTPUT:**

**Text

Description automatically generated with low confidence**

**A picture containing table

Description automatically generated**

**Cont:**

**Table

Description automatically generated**

**Cont:**

**A picture containing graphical user interface

Description automatically generated**

**Text

Description automatically generated**

**Table

Description automatically generated with medium confidence**

**Text

Description automatically generated**

1. **Apply PCA on CC dataset.**

* Reading CC.csv data and checking for any null values if any
* Applying iloc function to select the data
* Use data frame for tabular data.
* Head method for rows.

**Output:**

**Graphical user interface

Description automatically generated with medium confidence**

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

* Apply KMeans Algorithm for K=3.
* Use Predict function to predict cluster for each data point.
* Calculate Sihouette Score using metrics. Sihouette\_Score.

**Output:**

**Table

Description automatically generated**

****

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

* Use transform method to get scaled array.
* Apply fit\_transform on scaled array.
* Use dataframe to get tabular data.
* Use head to get rows.

**Output:**

**Graphical user interface, table

Description automatically generated with medium confidence**

**Graphical user interface, text, application, email

Description automatically generated**

**Table

Description automatically generated**

****

**Graphical user interface, text, application, chat or text message

Description automatically generated**

**Table

Description automatically generated**

****

**2. Use pd\_speech\_features.csv**

**a. Perform Scaling**

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

**c. Use SVM to report performance**

**Output:**

**Table

Description automatically generated**

**Cont:**

**A picture containing text

Description automatically generated**

**Cont:**

**A picture containing text

Description automatically generated**

**Table

Description automatically generated**

**A. Perform Scaling**

* Use fit transform method on x to scale

**Output:**

**Graphical user interface, text, application

Description automatically generated**

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

* Apply fit\_transform on scaled array.
* Use dataframe to get tabular data.
* Use head to get rows.

**Output:**

**Graphical user interface, text, application

Description automatically generated**

**C. Use SVM to report performance**

* Apply fit function on x\_train and y\_train.
* Use predict method to find predicted.
* Use accuracy\_score method to calculate accuracy score.
* Use Sihouette\_Score method to calculate Sihouette Score.

**Output:**

**Graphical user interface, text, application

Description automatically generated**

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

* Calculate x train and x test values using fit\_transform and transform.
* Calculate the x\_train and x test values for k =2 using LDA by using x train and x test.

**Output:**

**Graphical user interface, text, application, email

Description automatically generated**

**Graphical user interface, text, application, email

Description automatically generated**

**4. Briefly identify the difference between PCA and LDA**

Both LDA and PCA rely on linear transformations and aim to maximize

the variance in a lower dimension. PCA is an unsupervised learning algorithm

while LDA is a supervised learning algorithm. This means that PCA finds directions of maximum variance regardless of class labels while LDA finds directions of maximum class separability.

**PCA:**

It reduces the features into a smaller subset of orthogonal variables,

called principal components – linear combinations of the original variables. The

first component captures the largest variability of the data, while the second

captures the second largest, and so on.

**LDA:**

LDA finds the linear discriminants in order to maximize the variance

between the different categories while minimizing the variance within the class.