Machine learning Assignment 5

Video link:

<https://drive.google.com/file/d/1Me8M1eeMtY4a0GbruASMFwLBwUw5KxAf/view?usp=sharing>

1. Principal Component Analysis

a. Apply PCA on the CC dataset.

b. Apply the 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.

Graphical user interface, text, application

Description automatically generated

A picture containing table

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

2. Use pd\_speech\_features.csv

a. Perform Scaling

b. Apply PCA (k=3)

c. Use SVM to report performance

Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface

Description automatically generated

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

Graphical user interface, text

Description automatically generated with medium confidence

Graphical user interface

Description automatically generated with medium confidence

Chart, scatter chart

Description automatically generated

4. Briefly identify the difference between PCA and LDA

PCA:

* PCA is an unsupervised learning algorithm
* PCA finds directions of maximum variance regardless of class labels
* PCA explicitly attempts to model the difference between the classes of data
* PCA performs better in cases where the number of samples per class are less

LDA:

* LDA is a supervised learning algorithm
* LDA finds directions of maximum class separability
* LDA on the other hand does not consider any difference in class
* LDA works better with large datasets having multiple classes