Fall 2022 5710 Machine Learning: Assignment 5

Alekhya Nalla

700734522

GitHublink:

Video Link:

Programming elements: Principal Component Analysis

In class programming:

1. Principal Component Analysis
   1. Apply PCA on CC dataset

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

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

Chart

Description automatically generated with medium confidence

Graphical user interface, text, application, email

Description automatically generated

The silhouette score improved.

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

Text

Description automatically generated

A picture containing graphical user interface

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. Use pd\_speech\_features.csv
   1. Perform Scaling

Graphical user interface, text, application, email

Description automatically generated

* 1. Apply PCA (k=3)

Graphical user interface, text, application, email

Description automatically generated

* 1. . Use SVM to report performance

Graphical user interface, text, application, email

Description automatically generated

Text, table

Description automatically generated with medium confidence

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

Table

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Chart, scatter chart

Description automatically generated

1. Briefly identify the difference between PCA and LDA

PCA is an unsupervised learning algorithm while LDA is a supervised learning algorithm. Consequently,

A principal component analysis (PCA) does not take into consideration class labels but emphasizes finding the principal components with the greatest variance. Thus it is an unsupervised algorithm. By contrast, LDA is a supervised algorithm that aims to identify linear discriminants that represent the axes that maximize separation between classes.

LDA performs better on multi-class classification tasks than PCA. However, PCA performs better when the sample size is comparatively small. In image classification, comparisons of classification accuracy could be made.

Both LDA and PCA are used in the case of dimensional reduction. Generally, PCA is followed by LDA.