Department of Computer Science and Engineering,
Anna University, Chennai- 600025
CS23005 – Business Analytics Lab
Practical
July – November 2026
Year/Sem : IV / VII

Exercise: 07	Dimensionality Reduction using PCA, KPCA & SVD	21 – Aug - 2026

Observation (5 Marks)

1. What is dimensionality reduction?
2. What is PCA?
3. Why is standardization required before PCA?
4. What is the difference between PCA and KPCA?
5. What is SVD?
6. A dataset contains 10 features. Apply PCA to reduce it to 3 components. Write the Python statement required.
7. Which dimensionality reduction technique is suitable for capturing non-linear relationships?
8. If the first two principal components explain 85% of the variance, what does this indicate?
9. What are the three matrices obtained from SVD?
10. A business dataset contains 8 numerical features. You want to reduce the dataset to 3 important features while retaining maximum variance.
    a) Which dimensionality reduction technique would you choose?
    b) Write the Python statement to perform this reduction using the selected technique.

Execution (20 Marks)

1. A retail company collects several customer-related attributes such as:

    Annual Income
    Spending Score
    Purchase Frequency
    Average Purchase Value
    Website Visits
    Customer Age

Some of these variables may contain correlated information. The company wants to reduce the number of dimensions while retaining the important information for further analysis.

Data Preprocessing

Load the dataset using Pandas.
Display the first five records.
Check for missing values.
Select the numerical attributes.
Standardize the data using StandardScaler.

PCA

Apply Principal Component Analysis (PCA) to reduce the dataset from 5 dimensions to 2 dimensions.

Apply PCA with n_components=2.
Transform the standardized data.
Display the two principal components.
Calculate the explained variance ratio.
Plot the transformed data using a scatter plot.

Kernel PCA (KPCA)

Apply Kernel PCA to reduce the dataset to two dimensions.

Use the RBF kernel.

Apply Kernel PCA with two components.
Use the RBF kernel.
Transform the standardized dataset.
Display the transformed data.
Plot the two kernel principal components.

SVD

Apply Singular Value Decomposition (SVD) to reduce the dimensionality of the dataset.

Perform SVD on the standardized dataset.
Select the first two singular vectors.
Generate a two-dimensional representation.
Display the reduced dataset.
Calculate the proportion of variance represented by the selected singular values.

Compare PCA, KPCA and SVD

Create a comparison table:

Compare the results based on:

1. Dimensionality reduction
2. Variance retained
3. Computational complexity
4. Linear/non-linear transformation
5. Suitability for business data

Visualization

Generate three separate scatter plots:

Plot 1 - PCA – First Two Principal Components
Plot 2 - KPCA – First Two Kernel Components
Plot 3 - SVD – First Two Components