|  |  |
| --- | --- |
| Subject: Machine Learning Lab | Course ID: CSL-604 |
| Semester: VI | Course: AI & DS |
| Laboratory: 406-B | Name of teacher: Prof. Seema Pawar |
| Name of Student: Ayush Gupta | Roll No: VU2S2324001 |

**EXPERIMENT NO. 9**

**Aim:**

To implement Principal Component Analysis (PCA).

**Theory:**

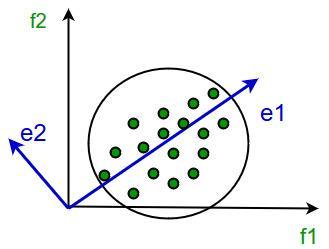
Dimensionality reduction is the process of reducing the number of features (or dimensions) in a dataset while retaining as much information as possible. It transforms high-dimensional data into a lower-dimensional space that still preserves the essence of the original data. This helps in reducing model complexity, improving algorithm performance, and making data visualization easier.

There are two major approaches to dimensionality reduction:

1. **Feature Selection:** Selecting the most relevant features from the original dataset (e.g., filter, wrapper, and embedded methods). The goal is to reduce the dimensionality of the dataset while retaining the most important features.
2. **Feature Extraction:** Creating new features that best represent the original data in a lower-dimensional space (e.g., PCA, LDA, t-SNE). The goal is to create a set of features that captures the essence of the original data in a lower-dimensional space.

**Principal Component Analysis (PCA):**

PCA is a popular feature extraction technique introduced by Karl Pearson. It projects data onto a lower-dimensional space while retaining the maximum variance. It helps in removing redundancy and improving computational efficiency in machine learning models.



It involves the following steps:

* Construct the covariance matrix of the data.
* Compute the eigenvectors of this matrix.
* Eigenvectors corresponding to the largest eigenvalues are used to reconstruct a large fraction of variance of the original data.

**Steps for PCA Algorithm:**

1. **Standardize the Data:** Ensure all features have a mean of 0 and standard deviation of 1.
2. **Compute the Covariance Matrix:** Identify relationships between different features.
3. **Compute Eigenvectors and Eigenvalues:** Determine the principal components that capture the highest variance.
4. **Select the Top Principal Components:** Choose the eigenvectors with the largest eigenvalues.
5. **Transform the Data:** Project the original data onto the selected principal components.

**Advantages of PCA:**

* **Reduces Dimensionality:** Helps in handling high-dimensional data by reducing redundant features.
* **Improves Computational Efficiency:** Reducing the number of features speeds up model training and inference.
* **Removes Correlation:** PCA transforms correlated features into uncorrelated principal components.

**Disadvantages of PCA:**

* **Loss of Information:** Some variance from the original dataset is lost, affecting model accuracy.
* **Interpretability Issues:** Principal components are not always easy to interpret in terms of original features.
* **Assumes Linearity:** PCA assumes that maximum variance captures the most important information, which may not be true for non-linear data.

**Learning Objectives:**

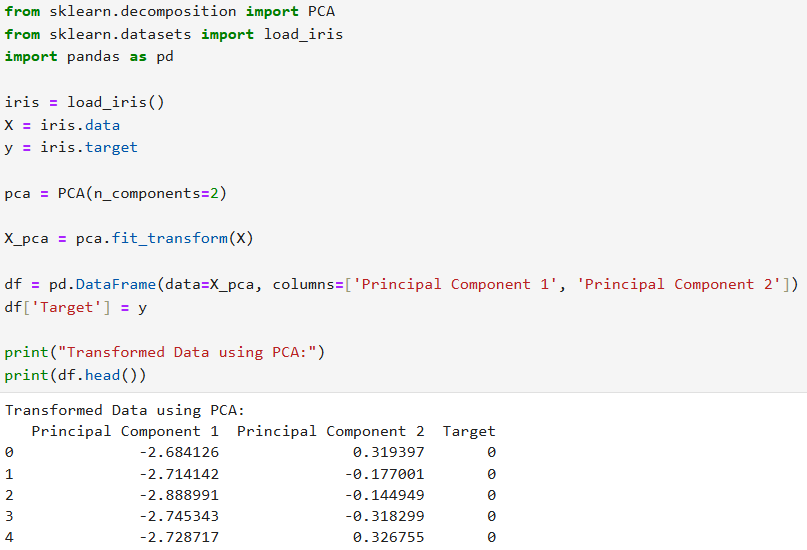
* To understand and implement Principal Component Analysis (PCA).
* To reduce the dimensionality of a dataset while retaining important information.

**Conclusion:**

Thus, the Principal Component Analysis (PCA) algorithm is studied and successfully implemented to reduce dimensionality while preserving the variance in the dataset.



**Program and Output:**

****