Principal Component Analysis (PCA)

1. Introduction

Principal Component Analysis (PCA) is a dimensionality reduction technique used to transform data into a lower-dimensional space while retaining most of its variance. In this report, we analyze a dataset of 10 observations with two characteristics and perform PCA to extract the most significant component.

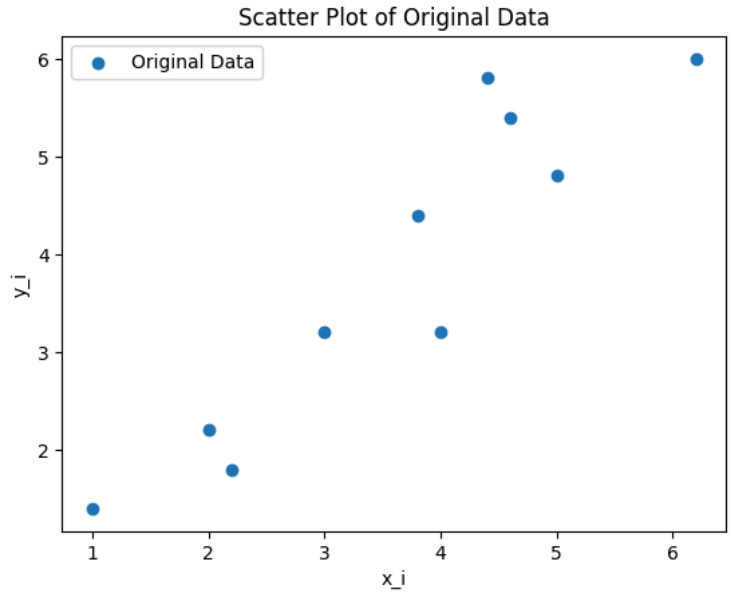
2. Dataset Description

The dataset consists of 10 observations with two characteristics (x_i and y_i). The goal is to apply PCA to find the principal components, transform the data, and reconstruct it back.

3. Steps Performed

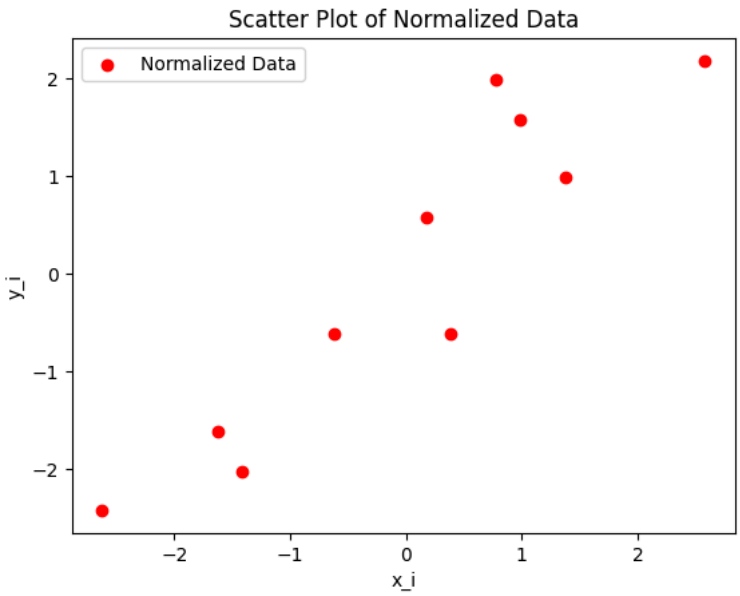
a. Scatter Plot of Original Data

A scatter plot generated to visualize the original dataset, showing the spread of data points.



b. Data Normalization

The dataset was normalized by subtracting the mean of each feature, centering it around the origin. A scatter plot was plotted to visualize the normalized data.



c. Computation of Covariance Matrix

The covariance matrix was computed to understand the relationship between the two features. The covariance matrix is:

[[2.658 - 2.495]
 [-2.495 2.816]]

d. Eigenvalues and Eigenvectors

Eigenvalues and eigenvectors were computed from the covariance matrix:

- **Eigenvalues:** `λ1 = 5.069, λ2 = 0.405`
- **Eigenvectors:**

[[0.710 -0.704]
 [0.704   0.710]]

The principal component corresponds to the eigenvector associated with the largest eigenvalue, which represents the direction of maximum variance.

e. Transformation Using Principal Component

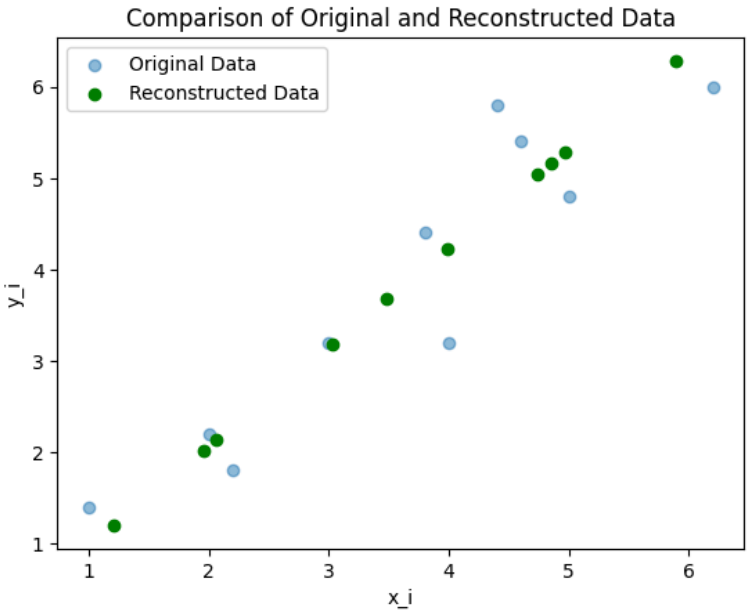
The data was projected onto the principal component to reduce dimensionality while preserving the most variance.

f. Reconstruction of Data

The transformed data was projected back to the original space using the principal component and the mean was added back to approximate the original dataset.

g. Comparison of Original and Reconstructed Data

A scatter plot was plotted to compare the reconstructed data with the original dataset, showing how well PCA captured the main structure of the data.



4. Conclusion

PCA successfully identified the principal direction of variance in the dataset. The reconstructed data closely approximates the original dataset, demonstrating the effectiveness of PCA in reducing dimensionality while retaining key information. This method can be extended to larger datasets for feature extraction and noise reduction.