Future Institute of Technology

NAME : **MD SUFIAN**

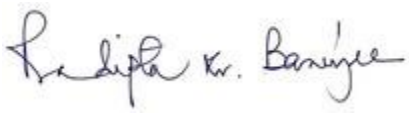
DEPARTMENT : **CSE AI&ML**

ROLL NO : **34230820012**

SEMESTER : **6TH**

TITLE OF REPORT : **PRINCIPAL COMPONENT ANALYSIS**

DATE OF SUBMISSION : **03/03/2023**



EXAMINATION SIGNATURE WITH DATE: …… …….

Contents

Abstract

Principal component analysis of a data matrix extracts the dominant patterns in the matrix in terms of a complementary set of score and loading plots. It is the responsibility of the data analyst to formulate the scientific issue at hand in terms of PC projections, PLS regressions, etc. Ask yourself, or the investigator, why the data matrix was collected, and for what purpose the experiments and measurements were made. Specify before the analysis what kinds of patterns you would expect and what you would find exciting.

1. What is Principal Component Analysis?

The Principal Component Analysis is a popular unsupervised learning technique for reducing the dimensionality of data. It increases interpretability yet, at the same time, it minimizes information loss. It helps to find the most significant features in a dataset and makes the data easy for plotting in 2D and 3D. PCA helps in finding a sequence of linear combinations of variables.

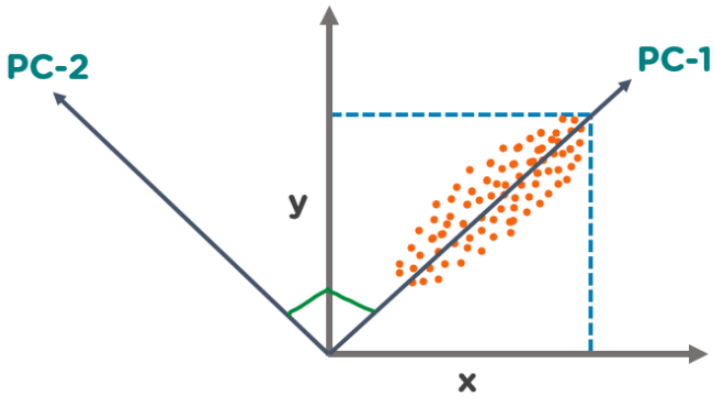


Fig. 1

In the above figure, we have several points plotted on a 2-D plane. There are two principal components. PC1 is the primary principal component that explains the maximum variance in the data. PC2 is another principal component that is orthogonal to PC1.

2. What is a Principal Component?

The Principal Components are a straight line that captures most of the variance of the data. They have a direction and magnitude. Principal components are orthogonal projections (perpendicular) of data onto lower-dimensional space.

Now that you have understood the basics of PCA, let's look at the next topic on PCA in Machine Learning.

Fig. 2

3. Applications of PCA in Machine Learning

- PCA is used to visualize multidimensional data.
- It is used to reduce the number of dimensions in healthcare data.
- PCA can help resize an image.
- It can be used in finance to analyze stock data and forecast returns.
- PCA helps to find patterns in the high-dimensional datasets.

4. How does Principal Component Analysis Work?

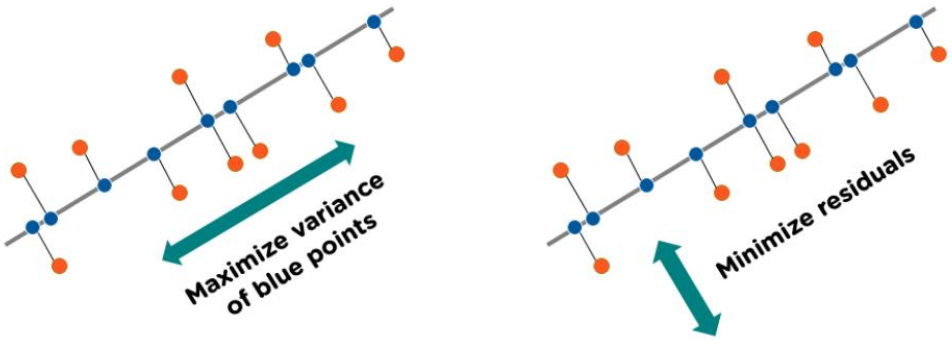


Fig. 3

1. Normalize the data

Standardize the data before performing PCA. This will ensure that each feature has a mean = 0 and variance = 1.

$$Z = \frac{x - \mu}{\sigma}$$

2. Build the covariance matrix

Construct a square matrix to express the correlation between two or more features in a multidimensional dataset.

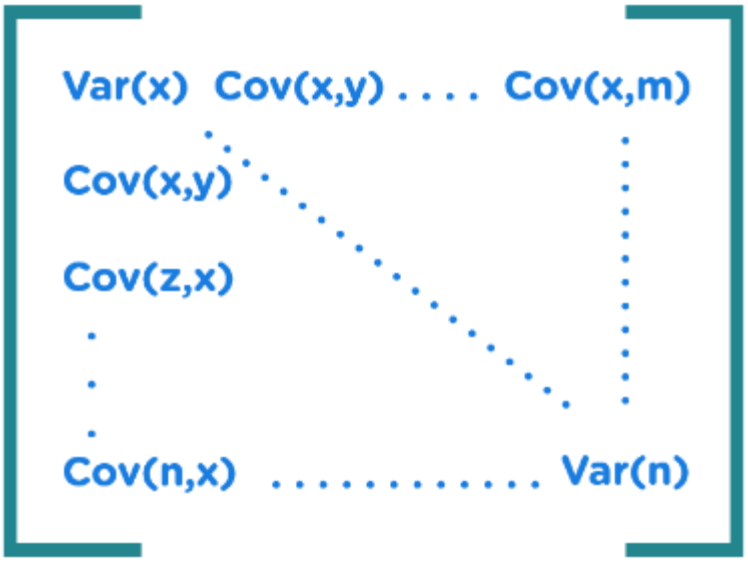


Fig. 4

3. Find the Eigenvectors and Eigenvalues

Calculate the eigenvectors/unit vectors and eigenvalues. Eigenvalues are scalars by which we multiply the eigenvector of the covariance matrix.

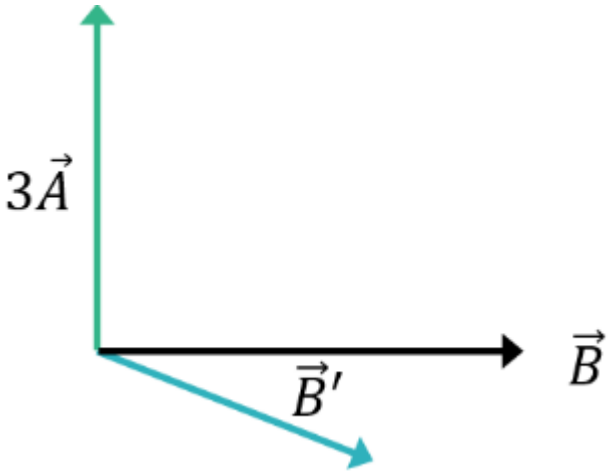


Fig. 5

4. Sort the eigenvectors in highest to lowest order and select the number of principal components.

5. Conclusion

The principal component analysis is a widely used unsupervised learning method to perform dimensionality reduction. We hope that this article helped you understand what PCA is and the applications of PCA. You looked at the applications of PCA and how it works.

Do you have any questions related to this article on PCA in Machine Learning? If yes, then please feel free to put them in the comments sections. Our team will be happy to solve your queries. Finally, we performed a