Principal Component Analysis is a powerful technique in statistics and data science that helps simplify high-dimensional data. In many situations like analyzing survey results, genetic data, or image pixels, we deal with datasets that have dozens, hundreds, or even thousands of variables. Making sense of such complex data can be overwhelming. This is where PCA becomes useful. It identifies the directions in the data, called principal components, which stands for the most important patterns or structures hidden within the dataset by close examination. PCA allows us to reduce the number of variables we need to consider.

PCA starts by computing the covariance matrix of the dataset, which captures how each variable relates to every other variable. Then, using tools like eigenvalues and eigenvectors, we extract new axes—directions in which the data varies the most. The first principal component represents the direction of maximum variance, the second is orthogonal to the first and captures the next most variance, and so on. This transformation not only helps reduce dimensions but also improves the performance of machine learning models.

One of the most fascinating aspects of PCA is how broadly it's used. PCA can reduce the number of pixels needed to represent an image while still maintaining a recognizable version. It also helps researchers find patterns in massive gene expression datasets. Moreover, PCA is used to analyze stock returns and identify common market movements.

We plan to explore the mathematical foundations of PCA, examining why it works. We will also implement PCA using programming tools like Python and apply it to real-world datasets across different domains. This project will deepen our understanding of linear algebra and data analysis.

Team mate:

Zhuoyang Lyu (Leo)：My major is statistics, which relevant to analyzing data, so I am interested in methods used to analyze complex data which huge amount of variables.

Charlie Zhang: My major is statistics, and I am very interested in using computers in analyzing data. So this why I chose this topic, not only acknowledge ways to analyze complex data, but also learn knowledge about computer science as well.

Yinan Wu: Majoring in statistics and minoring in data science provide a strong foundation for me in data analysis and statistical modeling, which are essential components for exploring PCA effectively.