An Overview of Principal Component Analysis: Theory and Applications

Blaine Swieder

Department of Computer Science
Tennessee Technological University
bcswieder42@tntech.edu

December 10, 2024

Introduction

- Principal Component Analysis (PCA) is machine learning technique that is used for dimensionality reduction and feature extraction.
- PCA transforms data into orthogonal components to capture maximum variance.
- **Applications:** Machine learning, image processing, bioinformatics, finance, etc.

Goals of this presentation:

- Explain theoretical foundations.
- Describe computational methodologies.
- Discuss some applications within the literature.

Theoretical Foundations

Mathematical Basis:

- Computes covariance matrix: $\mathrm{Cov}(X) = \frac{1}{n-1} X^\top X$.
- Solves eigenvalue equation: $\mathrm{Cov}(X)v = \lambda v$.
- Can also use Singular Value Decomposition (SVD): $X = U\Sigma V^\top$.

Limitations:

- PCA assumes linearity and Gaussian-distributed data.
- PCA is sensitive to scaling and outliers.

Computational Methodology

Data Preprocessing:

- Used to center data: $X_{centered} = X - \mu$.
- Used to Standardize data: $X_{standardized} = \frac{X-\mu}{\sigma}$.

PCA Steps:

1. Compute covariance matrix.
2. Perform eigenvalue decomposition.
3. Sort eigenvalues and select top k components.
4. Transform data: $X_{reduced} = X_{centered} V_k$.

Visualization Example

Dataset: Wine Dataset

- There are 13 chemical features of wine types in the dataset.
- PCA projects data onto the first two components, retaining 55.4% variance.

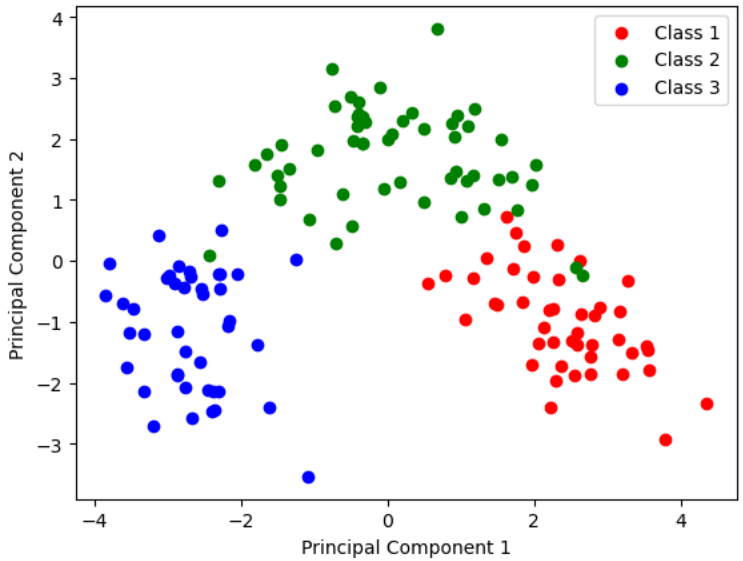

Figure: Scatterplot of wine classes in 2D space.

Applications of PCA

Examples Across Fields:

- **Image Processing:** PCA enhances object detection under rainy conditions (APCA-Net).
- **Bioinformatics:** Identifies significant patterns in gene expression data.
- **Finance:** Reduces data dimensions for portfolio and risk management.
- **Air Quality Prediction:** Improves forecasting with PCA-IFGA.
- **Translation Evaluation:** Assesses translation quality with KPCA.

Case Study: APCA-Net

Challenge: Rain impairs image-based object detection.

Solution:

- Uses PCA for rain removal as a preprocessing step.
- Incorporates Preprocessing Refinement Module (PRM) to automate hyperparameter selection.

Results:

- Improved detection accuracy under adverse weather.

PCA Limitations and Solutions

Limitations:

- PCA assumes linearity, limiting its ability to model nonlinear relationships.
- PCA is sensitive to outliers, which can distort results.

Solutions:

- **Kernel PCA:** Extends PCA to nonlinear data using the kernel trick.
- **Sparse PCA:** Improves robustness to outliers.

Conclusion

- PCA simplifies high-dimensional data while preserving essential patterns.
- Widely applicable across diverse fields, from image processing to finance.
- Future advancements in computational efficiency and hybrid methods will allow for more uses for PCA.

Takeaway: PCA is crucial for modern data analysis, machine learning, etc.

Questions

Questions?

Contact: bcswieder42@tntech.edu

References I

Y. Zhang, Y. Wang, and X. Liu, "Principal Component Analysis-Improved Fuzzy Genetic Algorithm," *Journal of Computational Intelligence and Applications*, vol. 12, no. 3, pp. 456–465, 2023.

Z. Chen and Q. Li, "Feature Selection Based on Improved Principal Component Analysis," *Journal of Data Mining and Knowledge Discovery*, vol. 45, no. 7, pp. 1123–1137, 2021.

H. Sun and Y. Zhao, "Cyclostationary Signal Sensing Algorithm Based on Principal Component Analysis and AdaBoost," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2345–2358, 2022.

References II


L. Yang and J. Wang, "APCA-Net: Adaptive Object Detection in Rainy Weather Based on Principal Component Analysis," *International Journal of Computer Vision and Pattern Recognition*, vol. 8, no. 4, pp. 98–110, 2022.


X. Liu, S. Chen, and R. Zhao, "Quality Classification and Evaluation of Human-Machine Composite Translations of Scientific Text Based on KPCA," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 5, pp. 2049–2059, 2021.


J. Shlens, "A Tutorial on Principal Component Analysis," *arXiv preprint*, arXiv:1404.1100, 2014.


I. T. Jolliffe and J. Cadima, *Principal Component Analysis, Second Edition*. New York: Springer, 2016.

References III

C. Bishop, "Principal Component Analysis: A Natural Approach to Data Exploration," in *Pattern Recognition and Machine Learning*, 2nd ed. Springer, 2006, pp. 348–350.

H. Zou, T. Hastie, and R. Tibshirani, "A Selective Overview of Sparse Principal Component Analysis," *Proceedings of the IEEE*, vol. 101, no. 3, pp. 775–792, 2013.

G. GeeksforGeeks, "ML — Introduction to Kernel PCA," *GeeksforGeeks*. [Online]. Available: https://www.geeksforgeeks.org/ml-introduction-to-kernel-pca/. [Accessed: 30-Nov-2024].

References IV

G. GeeksforGeeks, "Principal Component Analysis with Python," *GeeksforGeeks*. [Online]. Available: https://www.geeksforgeeks.org/principal-component-analysis-with-python/. [Accessed: 30-Nov-2024].