# Principal Component Analysis (PCA) using SVD

By: Sujal Sinha(2311191)

Date: 26th November 2025

# What is PCA?

- A technique to find the directions of maximum variance in data.
  - Used for dimensionality reduction and pattern discovery.
  - Transforms data into new orthogonal axes called Principal Components.

# Why Use PCA?

- Removes redundancy by combining correlated features.
  - Simplifies datasets while preserving important patterns.
  - Useful in visualization, ML preprocessing, noise reduction.

# Core Mathematical Idea

Given centered data matrix Xc:
Compute covariance matrix $C = Xc^T Xc$.
Eigenvectors of C = PCA directions.
Eigenvalues of C = variance along each direction.

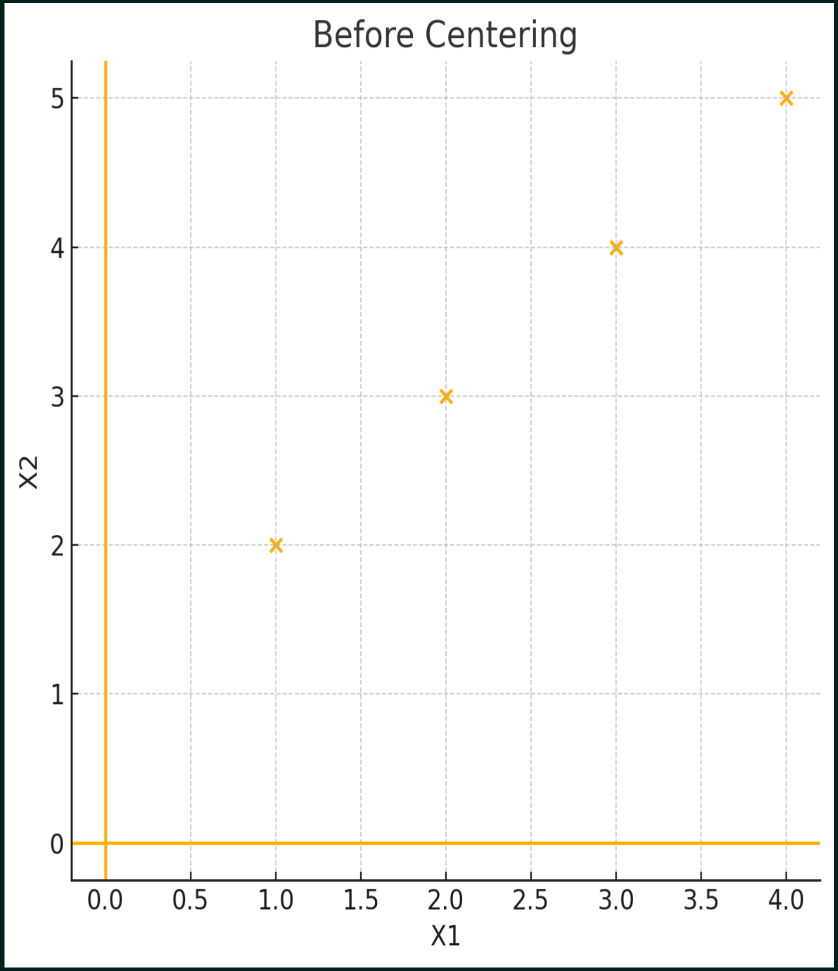
# Why and how do we center the Data?

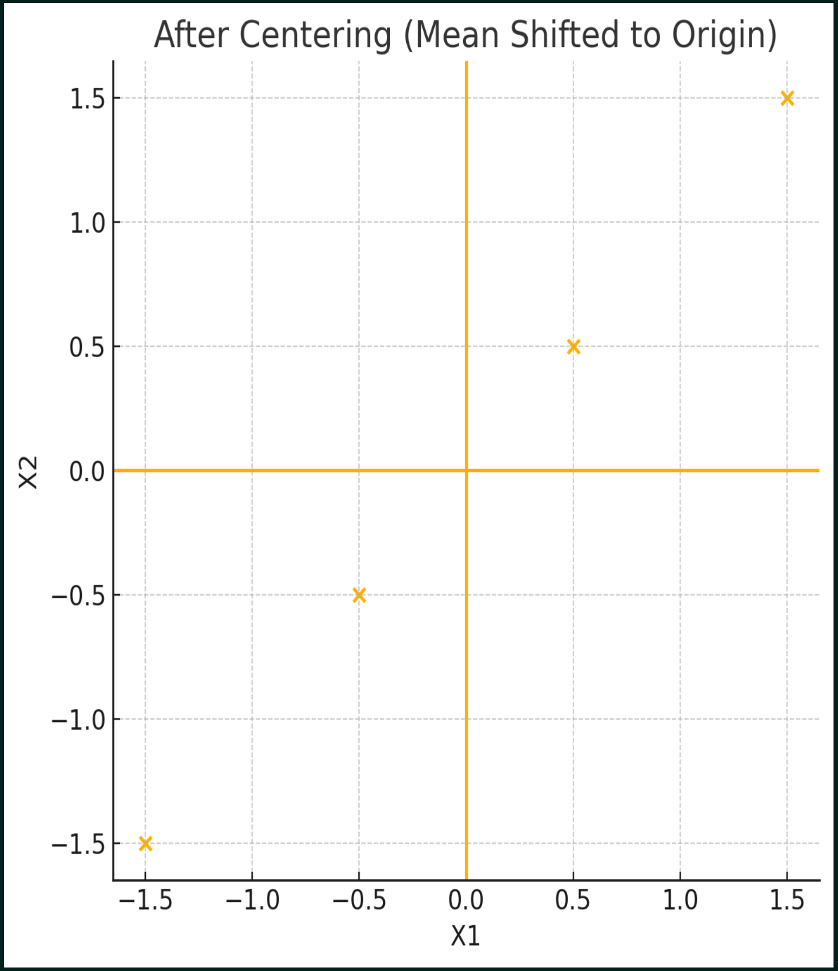
Means are subtracted so PCA analyzes shape, not location.

– Centering ensures it reflects true variability.

1. Compute mean of each column.
2. Subtract mean from each data point.

# Before and after centering our data



# SVD and PCA Connection

$X_c = U \Sigma V^T$ (Singular Value Decomposition).
Columns of V are principal directions.
Singular values $\sigma \rightarrow$ variance $= \sigma^2$.
SVD avoids numerical issues with $X^T X$.

# Understanding the Code: PCA using SVD

1. Compute column means.
2. Subtract means → Xc.
3. Use SVD(Xc) → returns singular values S and V matrix.
4. PCA variances = $S^2$.
5. Principal components = columns of V.

# Understanding the Code: Jacobi Eigenvalue

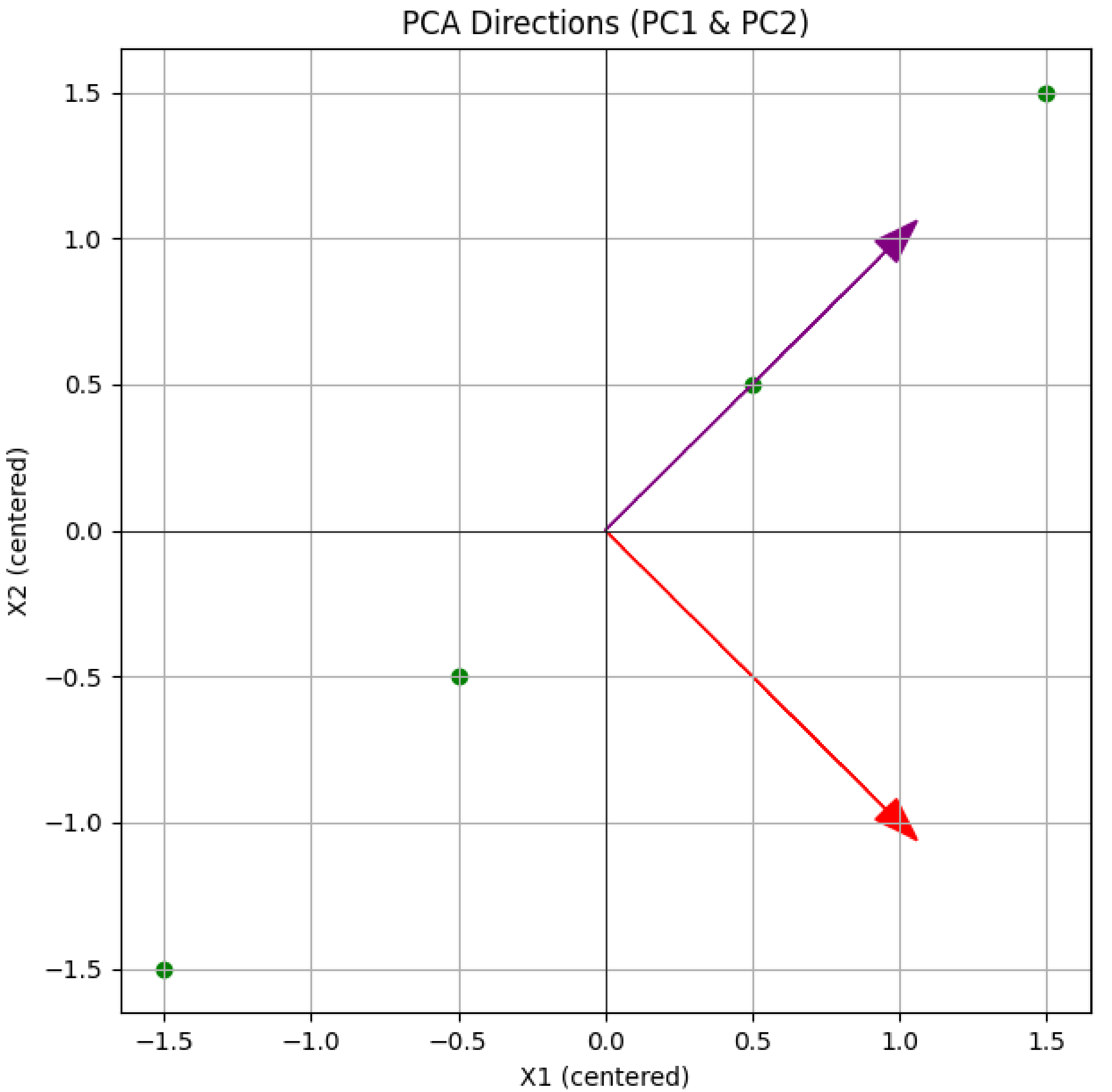
1. Iteratively zeroes out off-diagonal entries using rotations.
2. Finds angle θ so rotation eliminates M[p][q].
3. Repeats until matrix is diagonal.
4. Diagonal entries = eigenvalues
5. Rotation matrix = eigenvectors.

# Understanding the Code: SVD Function

1. Computes $A^T A$ using matrix multiplication.
2. Calls jacobi_eigen() to find eigenvalues/vectors.
3. Singular values = sqrt(eigenvalues).
4. Uses these V vectors as PCA directions.

# Final Output of PCA Code

1. Means: used to center any future data.
2. Principal components V: directions of maximum variance.
3. Variances $S^2$: tells how important each component is.
4. These are used to project or reduce dimensionality.



PCA Directions (PC1 & PC2)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
X2 (centered)
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
X1 (centered)

# Applications of PCA

1. Data compression.

2. Noise filtering.

3. Machine Learning preprocessing (scaling + PCA).

# Summary

PCA finds patterns by rotating data into meaningful axes.

Outputs: Means, Variances, PCA Directions.

Result: A powerful dimensionality reduction tool.

# References


- Ian T. Jolliffe, and Jorge Cadima. “Principal component analysis: a review and recent developments.” Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.

- Jolliffe, I.T. Principal Component Analysis. 2nd edition. Springer, 2002. DOI:10.1098/rsta.2015.0202.