# Algorithms for Data Science

## Unsupervised Learning: Eigen Decomposition

Johns Hopkins
Whiting School of Engineering

# Eigen Decomposition Overview

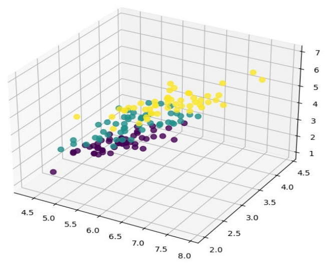
**Eigen Decomposition involves breaking a matrix into its eigenvalues and eigenvectors, representing variance and direction in data.**

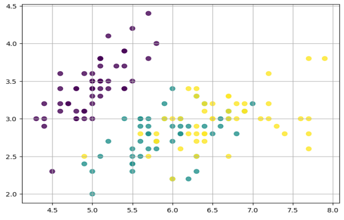
## Purpose

1. Dimensionality Reduction
2. Variance Expansion
3. Factor Analysis

## Applications

Data Compression and Feature Extraction



# Eigen Decomposition: Mathematical Formulation

## Eigen Decomposition:

$$\Sigma = V\Lambda V^T$$

Where:

- $\Sigma$: Covariance matrix
- $V$: Matrix of eigenvectors
- $\Lambda$: Diagonal matrix of eigenvalues

## Variance Expansion:

$$\Sigma = V\Lambda V^T$$

- Variance captured by each principal component is proportional to its eigenvalue.

## Factor Analysis: Rotated Components

$$Z_{\mathrm{rot}} = ZR$$

- Where $R$ is a rotation matrix (e.g. Varimax)

# Eigen Decomposition Algorithm Analysis

1. **Compute** the covariance matrix. → Covariance computation: $O(N^2 \times d^2)$

2. **Perform** eigen decomposition to get eigenvectors and eigenvalues. → Eigen decomposition: $O(d^3)$

3. **Project** data onto eigenvectors. → Data projection: $O(N \times d^2)$

Total Complexity: $O(N \times d^2 + d^3)$

# Variance and Dimensionality Reduction

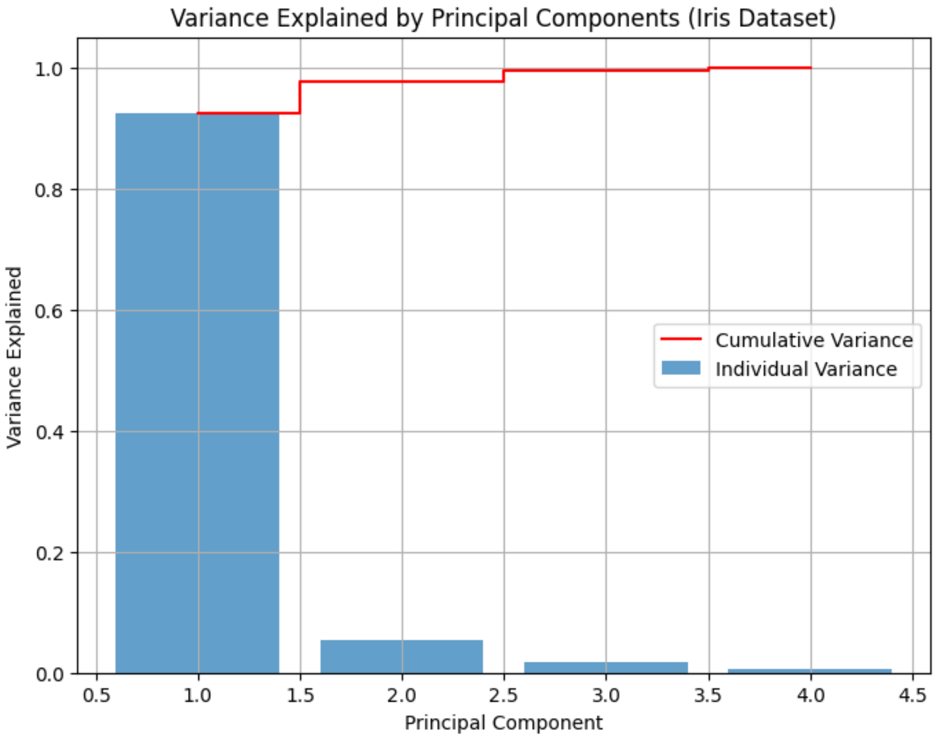
- Variance Explained by Eigenvalues:

$$Variance\ Explained\ (\%) = \frac{\lambda_i}{\sum_{j=1}^{d} \lambda_j} \times 100$$

- Where $\lambda_i$ is the eigenvalue for component $i$.

- Dimensionality Reduction:
  - Retain components explaining a high cumulative variance (e.g. 90%)
  - Reduce $d$ dimension to $k$ dimensions, where $k$ retains the desired variance.

# Variance and Dimensionality Reduction



# Eigen Decomposition: Correctness

**Theorem: Any symmetric covariance matrix has eigenvalues and orthogonal eigenvectors.**

- **Proof:**
  - Symmetry ensures: $\Sigma = \Sigma^T$.
  - Spectral theorem guarantees diagonalizability.
  - Eigen decomposition ensures: $\Sigma = V \Lambda V^T$

**Applications**

1. Dimensionality Reduction
2. Exploratory Data Analysis
3. Noise Filtering

JOHNS HOPKINS

WHITING SCHOOL
of ENGINEERING