# Linear Algebra Perspective of Principal Component Analysis

# What is PCA?

Principal Component Analysis (PCA) is a fundamental **dimensionality-reduction technique** grounded in linear algebra.

## Goal of PCA

- Identify directions of maximal variance in the data
- Express the data in a new orthogonal coordinate system
- Retain the most informative components

# Data Representation

We consider a dataset of:

- $m$ observations
- $n$ features

The data matrix is

$$X \in \mathbb{R}^{m \times n}$$

- Rows: observations
- Columns: features
- Each observation is a point in $\mathbb{R}^n$

**Assumption**

The data are **centered**: each feature has zero mean.

# Idea Behind PCA

PCA seeks directions along which the data vary the most.

- First principal component: direction of maximal variance
- Second principal component:
  - maximal remaining variance
  - orthogonal to the first
- And so on

### Outcome

- Orthogonal directions with decreasing variance
- A low-dimensional subspace approximating the data

# Covariance Matrix

The sample covariance matrix of the features is

$$C = \frac{1}{m} X^\top X \in \mathbb{R}^{n \times n}$$

- Symmetric
- Positive semidefinite

# Variance Maximization

For a unit vector $v \in \mathbb{R}^n$, the variance of the projection is

$$\mathrm{Var}(Xv) = v^\top C v$$

**Optimization Problem**

$$\max_{\|v\|=1} v^\top C v$$

This leads to the eigenvalue problem:

$$Cv = \lambda v$$

# Principal Directions

- Eigenvectors of $X^\top X$:
    - principal directions in feature space
- Eigenvalues:
    - variance explained along each direction

# Spectral Decomposition

Since $C = X^\top X$ is symmetric:

$$C = V \Lambda V^\top$$

- $V = [v_1, \ldots, v_n]$: orthonormal eigenvectors
- $\Lambda = \mathrm{diag}(\lambda_1 \geq \lambda_2 \geq \cdots \geq 0)$

# Dimensionality Reduction

Let

$$V_k = [v_1, \ldots, v_k]$$

### Projection onto Principal Subspace

$$Z = XV_k$$

- Each row of $Z$ is an observation in PCA coordinates
- The subspace captures maximal variance among all $k$-dimensional subspaces

# Dual PCA Formulation

When $n \gg m$, it is computationally advantageous to consider

$$XX^\top \in \mathbb{R}^{m \times m}$$

**Key Property**

$X^\top X$ and $XX^\top$ share the same nonzero eigenvalues.

# Feature Space vs Observation Space

If

$$X^\top X v = \lambda v$$

then

$$u = \frac{1}{\sqrt{\lambda}} X v \quad \Rightarrow \quad X X^\top u = \lambda u$$

- Eigenvectors of $X^\top X$: feature space directions
- Eigenvectors of $XX^\top$: observation space directions

# PCA via SVD

The SVD of the centered data matrix is

$$X = U\Sigma V^\top$$

Then:

$$X^\top X = V\Sigma^2 V^\top, \qquad XX^\top = U\Sigma^2 U^\top$$

# Interpretation via SVD

- Right singular vectors $V$:
    - principal directions
- Singular values:

$$\sigma_i^2 = \lambda_i$$

- Left singular vectors $U$:
    - PCA coordinates of observations

**Key Insight**

PCA is a direct output of the SVD.

# Feature Mapping

Each observation $x \in \mathbb{R}^n$ is mapped to

$$z = V_k^\top x \in \mathbb{R}^k$$

- Preserves maximal variance
- Removes correlations (in principal axis space)
- Produces a compact representation

# Summary

## Principal Component Analysis

- Build the covariance matrix
- Solve an eigenvalue problem
- Use eigenvectors as a new orthonormal basis
- Project data onto leading eigen-directions

| PCA = Eigenvalue problem and Projection |
|---|