Principal Component Analysis (PCA)

1. Data Representation

Assume we have a dataset with

- n samples
- d features

The data matrix is:

$$\boxed{X \in R^{n \times d}}$$

Each row is a data point, each column is a feature.

---

2. Mean Centering (Important Step)

PCA always works on centered data.

Compute mean of each feature:

$$\boxed{\mu = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

Center the data

$$\boxed{X_{centered} = X - \mu}$$

Why?

PCA finds directions of maximum variance around the origin.

---

3. Covariance Matrix

The covariance matrix measures how features vary together.

$$\Sigma = \frac{1}{n-1} X_{centered}^T X_{centered}$$

$$\Sigma \in R^{d \times d}$$

- Diagonal → variance of features
- Off-diagonal → correlation between features

4. Eigen Decomposition

We decompose the covariance matrix:

$$\Sigma v = \lambda v$$

Where:

- λ = eigenvalue
- v = eigenvector

Each eigenvector represents a direction

Each eigenvalue represents variance along that direction.

5. Sorting Eigenvalues

Eigenvalues are sorted in descending order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$$

- Largest eigenvalue → Principal Component 1
- Next → PC2, PC3, ...

6. Principal Components

Principal Components are the eigenvectors of the Covariance matrix.

$$PC_k = v_k$$

They form an orthonormal basis:

$$v_i^T v_j = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

7. Dimensionality Reduction

To reduce from d dimensions to k:

$$W = [v_1, v_2, \ldots, v_k]$$

$$W \in R^{d \times k}$$

8. Projection (Core PCA Equation)

Project data onto new subspace:

$$Z = X_{centered} W$$

Where:

- $Z \in R^{n \times k}$
- New features are uncorrelated
- Maximum variance is preserved

---

9. Variance Explained

Variance captured by each PC:

$$\boxed{\text{Explained Variance Ratio} = \frac{\lambda_k}{\sum_{i=1}^{d} \lambda_i}}$$

This tells how much information is kept.

---

10. Reconstruction

Original data approximation:

$$\boxed{\hat{X} = ZW^T + \mu}$$

More PCs → better reconstruction

Fewer PCs → more information loss

---

11. Optimization View (Deep Insight)

PCA solves:

$$\boxed{\max_{\|w\|=1} \operatorname{Var}(Xw)}$$

This leads directly to the eigenvector with largest eigenvalue.

Key Takeaways

- PCA is unsupervised
- Uses covariance + eigen decomposition
- ~~Principle~~ Principal components are orthogonal
- Preserves maximum variance
- Reduces redundancy + noise.