Principal Component Analysis – Intuition & Mathematics

This note explains the intuition and mathematics behind Principal Component Analysis (PCA). Both geometric interpretation and the linear algebra underlying PCA are discussed. A simple example is included to clarify the concepts.

What is PCA?

Principal Component Analysis (PCA) is a statistical method used for simplifying a dataset by reducing its number of dimensions (features), while preserving as much variance (information) as possible.

This is achieved by transforming the original features into new uncorrelated features, called principal components. These components are ordered based on the amount of variance in the data that they capture.

Mathematical Steps Behind PCA

Step 1: Standardization of the Data

PCA is sensitive to the scale of data. Therefore, features must be standardized to have a mean of 0 and a standard deviation of 1.

```
X_std = (X - X.mean()) / X.std()
```

Step 2: Calculation of the Covariance Matrix

The covariance matrix is used to capture the relationships between features.

For a dataset with `n` samples and `p` features:

$$\mathrm{Cov}(X) = \frac{1}{n-1} X^T X$$

If the dataset has 2 features, the covariance matrix takes the following form:

$$\begin{bmatrix} \mathrm{Var}(x_1) & \mathrm{Cov}(x_1, x_2) \\ \mathrm{Cov}(x_2, x_1) & \mathrm{Var}(x_2) \end{bmatrix}$$

Step 3: Computation of Eigenvalues and Eigenvectors

The characteristic equation is solved to find the eigenvalues (λ) and eigenvectors (v):

det (A – λ *I*) = 0

Where:

- A is the covariance matrix
- *I* is the identity matrix
- The eigenvalues λ indicate the amount of variance in the direction of their corresponding eigenvectors
- The eigenvectors define the directions of the principal components

---

Step 4: Sorting and Selection of the Top-k Components

The eigenvectors are sorted by their corresponding eigenvalues in descending order. The top `k` eigenvectors are retained for dimensionality reduction. These eigenvectors are used as the principal components.

Step 5: Transformation of the Data

The original standardized data is projected onto the new PCA subspace:

$$X_{\text{reduced}} = X_{\text{std}} \cdot W$$

Where:

- `W` is the matrix composed of the `k` selected eigenvectors
- `X_reduced` is the transformed dataset in the reduced dimensional space

---

Geometric Interpretation

- The eigenvectors define the directions of the new coordinate axes
- The eigenvalues represent the variance along each axis
- The original coordinate system is rotated by PCA to align with these new axes

Example: If the following vector is obtained: $v = \begin{bmatrix} 0.6779 \\ 0.7352 \end{bmatrix}$

This vector defines a line in 2D space along which the data varies the most. The projections of all points on this line form the first principal component.

Sample 2D Example

Given a covariance matrix:

The eigenvalues are obtained by solving: $\det(A - \lambda I) = (2 - \lambda)^2 - 1 = 0 \Rightarrow \lambda = 3, 1$

The corresponding eigenvectors are found by solving: $(A - \lambda I)v = 0$

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

These eigenvectors serve as the new coordinate axes for PCA.

Advantages of PCA

- Dimensionality is reduced
- Multicollinearity is eliminated
- Model performance is improved
- High-dimensional data can be visualized

When PCA Should Not Be Used

PCA may not be suitable in the following situations:

- When the original features are meaningful and their interpretability is important, as PCA transforms features into linear combinations that may lose semantic clarity.
- When the dataset exhibits strong non-linear relationships, since PCA assumes linearity.
- When PCA components do not align with class boundaries in supervised learning tasks, leading to potential degradation in model performance.
- When the goal is to retain interpretability of individual features, because PCA makes it difficult to trace back to the original features.

Summary

Concept	Description
Covariance Matrix	Represents feature relationships
Eigenvectors	Define directions of maximum variance
Eigenvalues	Indicate variance magnitude in each direction
Principal Components	Selected top-k eigenvectors
PCA Projection	Data is rotated and projected

References

- CampusX: Principle Component Analysis (PCA), YouTube
- 3Blue1Brown: PCA Animation, YouTube
- Scikit-learn: PCA Documentation
- Khan Academy: Linear Algebra Review
