Principal Component Analysis (PCA): A Detailed Breakdown

Tanmay Chopra

November 21, 2025

1 What is PCA?

Principal Component Analysis (PCA) is a powerful **dimensionality reduction** technique used in statistics and machine learning. Its main goals are:

- Reduce the number of variables (features) in a dataset while retaining as much **information (variance)** as possible.
- Identify underlying patterns in high-dimensional data.
- Remove redundancy caused by correlated features.

In essence, PCA transforms your data into a new coordinate system where the axes (called **principal components**) capture the directions of maximum variance.

—

2 Intuition Behind PCA

Imagine a cloud of 3D data points. If the points lie mostly on a flat 2D plane within the 3D space, you don't need all three dimensions to describe their structure—two are sufficient. PCA finds that optimal plane and defines new axes along the directions of maximum variance.

- **Principal Components (PCs)** are linear combinations of the original features.
- **PC1** points in the direction of maximum variance.
- **PC2** is **orthogonal** (perpendicular) to PC1 and captures the next highest variance, and so on.
- PCA is an **unsupervised** technique, meaning it does not use label information.

—

3 Mathematical Formulation

Assume we have a dataset X with n samples and p features.

3.1 Step 1: Standardize the Data

PCA is sensitive to feature scale, so we must standardize the data. For each feature, we calculate the Z-score:

$$Z = \frac{X - \mu}{\sigma}$$

Where μ is the mean vector and σ is the standard deviation of each feature.

3.2 Step 2: Compute the Covariance Matrix

The covariance matrix C measures how features vary together.

$$C = \frac{1}{n-1} Z^T Z$$

C is a $p \times p$ symmetric matrix. The diagonal elements are the variance of each feature, and the off-diagonal elements are the covariance between features.

3.3 Step 3: Compute Eigenvectors and Eigenvalues

We solve the characteristic equation for the covariance matrix C:

$$Cv = \lambda v$$

Where:

- λ (**eigenvalue**) represents the variance captured by that component.
- v (**eigenvector**) represents the direction of the principal component.

We sort the eigenvectors by their corresponding λ values in descending order to find the components that capture the most variance.

3.4 Step 4: Select Principal Components (k)

Decide how many PCs, k, to keep.

Variance Explained Ratio

The proportion of total variance explained by the i-th component is:

$$\text{Variance explained by PC}_i = \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}$$

A common approach is to select k such that the cumulative variance explained is around 90%–95%.

Scree Plot

Alternatively, we plot the eigenvalues and look for the "elbow point" where the eigenvalues begin to flatten out.

3.5 Step 5: Transform the Data

Project the standardized data Z onto the top k selected eigenvectors (the projection matrix W):

$$Y = ZW$$

Where W is the $p \times k$ matrix of selected eigenvectors, and Y is the $n \times k$ reduced-dimensional representation of the data. Y now has uncorrelated features and a reduced dimension k.

—

4 Geometric Interpretation

The original axes are the features (dimensions). PCA finds new axes (principal components) such that the first axis aligns with the direction of maximum spread of the data, and subsequent axes are orthogonal (uncorrelated) while capturing the remaining variance. The data is effectively rotated to align with these new axes.

—

5 Properties of PCA

- **Orthogonality:** Principal Components are mutually orthogonal, which ensures they are **uncorrelated**.
- **Maximizes Variance:** The m-th PC captures the maximum possible variance that is orthogonal to the previous $m-1$ PCs.
- **Linear Combinations:** Each PC is a linear combination of the original features.
- **Lossy Compression:** Dimensionality reduction necessarily discards some information, leading to lossy compression.

—

6 Applications of PCA

- **Dimensionality Reduction:** Reducing the computational cost for subsequent machine learning models.
- **Data Visualization:** Enabling the visualization of high-dimensional data by projecting it onto the first two or three PCs.
- **Noise Reduction:** Components with very small eigenvalues often correspond to noise; discarding them can denoise the dataset.
- **Feature Extraction:** Creating new, uncorrelated features that are optimal in representing the data's variance.

—

7 Advantages and Limitations

Advantages:

- Reduces complexity and computation time.
- Removes multicollinearity by generating uncorrelated features.
- Helps visualize high-dimensional data.
- Can improve model generalization in noisy settings.

Limitations:

- **Linear Method:** Cannot capture non-linear relationships in the data.
- **Sensitive to Scaling:** Requires proper data standardization prior to execution.
- **Interpretability:** Principal Components are linear combinations of original features, which can make the new features difficult to interpret.
- **Loss of Information:** Inevitably loses some information when reducing dimensions.

—

8 Quick Example

Suppose a dataset has 3 features: **height**, **weight**, and **age**.

1. **Standardize** the features.
2. **Compute Covariance** matrix: Identify correlated variables (e.g., height & weight).
3. **Compute Eigen-pairs**: Find the directions and magnitudes of variance.
4. **Select** k: Keep the top 2 PCs (if they capture $\sim 95\%$ of variance).
5. **Transform Data**: The original data is mapped to a new 2D space, providing a new dataset with 2 uncorrelated features that optimally represent the original information.

In short: PCA finds the "best-fit" lower-dimensional space where your data can be represented with minimal information loss, prioritizing directions with the highest variance.