Principal Component Analysis (PCA)

Algebra, SVD, and Geometric Interpretation

The Goal of PCA: An Intuitive View

Core Idea

Principal Component Analysis is a dimensionality reduction technique used to transform a high-dimensional dataset into a lower-dimensional one, while retaining as much of the original "information" as possible.

- PCA finds a new set of coordinates (or axes), called **Principal Components**.

The Goal of PCA: An Intuitive View

Core Idea

Principal Component Analysis is a dimensionality reduction technique used to transform a high-dimensional dataset into a lower-dimensional one, while retaining as much of the original "information" as possible.

- PCA finds a new set of coordinates (or axes), called **Principal Components**.
- These components are ordered such that the first component captures the largest possible variance in the data.

The Goal of PCA: An Intuitive View

Core Idea

Principal Component Analysis is a dimensionality reduction technique used to transform a high-dimensional dataset into a lower-dimensional one, while retaining as much of the original "information" as possible.

- PCA finds a new set of coordinates (or axes), called **Principal Components**.
- These components are ordered such that the first component captures the largest possible variance in the data.
- Each subsequent component is orthogonal to the previous ones and captures the largest possible remaining variance.

The Goal of PCA: An Intuitive View

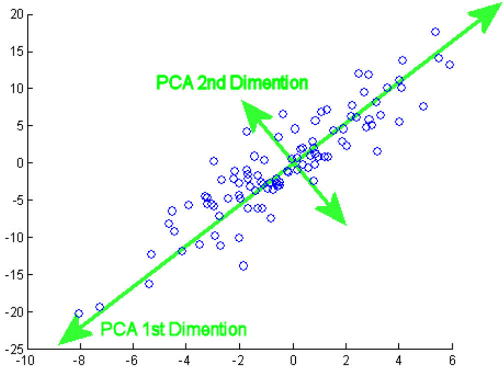


Figure: PCA finds the direction of maximum variance (PC1) and then subsequent orthogonal directions (PC2).

The Mathematics of Maximizing Variance

1. **Center the Data:** Start with a data matrix $X \in \mathbb{R}^{n \times p}$ (n samples, p features). First, make each feature have a mean of zero.

The Mathematics of Maximizing Variance

1. **Center the Data:** Start with a data matrix $X \in \mathbb{R}^{n \times p}$ (n samples, p features). First, make each feature have a mean of zero.
2. **Covariance Matrix:** The goal is to find directions that maximize variance. This is captured by the sample covariance matrix C:

$$C = \frac{1}{n-1} X^T X \quad \in \mathbb{R}^{p \times p}$$

The diagonal entries C_{ii} are the variances of each feature, and the off-diagonal entries C_{ij} are the covariances.

The Mathematics of Maximizing Variance

1. **Center the Data:** Start with a data matrix $X \in \mathbb{R}^{n \times p}$ (n samples, p features). First, make each feature have a mean of zero.
2. **Covariance Matrix:** The goal is to find directions that maximize variance. This is captured by the sample covariance matrix C:

$$C = \frac{1}{n-1} X^T X \quad \in \mathbb{R}^{p \times p}$$

   The diagonal entries C_{ii} are the variances of each feature, and the off-diagonal entries C_{ij} are the covariances.
3. **Eigendecomposition:** PCA solves this problem by finding the eigenvectors and eigenvalues of the covariance matrix.

$$C\mathrm{v}_j = \lambda_j \mathrm{v}_j$$

   - The eigenvectors v_j are the **Principal Components**. They give the directions of maximum variance.
   - The eigenvalues λ_j give the amount of variance captured by each principal component.

A More Stable Approach: Using SVD

Computing the covariance matrix $X^T X$ can be numerically unstable. A more robust method is to use the Singular Value Decomposition (SVD) of the centered data matrix X.

Recall SVD

Any matrix $X \in \mathbb{R}^{n \times p}$ can be decomposed as:

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

where U and V are orthogonal matrices and Σ is a diagonal matrix of singular values σ_i.

- $U \in \mathbb{R}^{n \times n}$: Left singular vectors.
- $\Sigma \in \mathbb{R}^{n \times p}$: Diagonal matrix of singular values.
- $V \in \mathbb{R}^{p \times p}$: Right singular vectors.

The SVD-PCA Connection

Let's substitute the SVD of X into the covariance matrix formula:

$$
\begin{aligned}
C = \frac{1}{n-1}X^TX &= \frac{1}{n-1}(U\Sigma V^T)^T(U\Sigma V^T) \\
&= \frac{1}{n-1}(V\Sigma^T U^T)(U\Sigma V^T) = \frac{1}{n-1}V\Sigma^T(U^TU)\Sigma V^T \\
&= \frac{1}{n-1}V\Sigma^T\Sigma V^T = V\left(\frac{\Sigma^2}{n-1}\right)V^T
\end{aligned}
$$

The Key Insight

This final expression, $C = V\Lambda V^T$, is exactly the eigendecomposition of the covariance matrix C.

- The columns of V (right singular vectors of X) are the principal components.
- The eigenvalues of C are related to the singular values of X by $\lambda_j = \frac{\sigma_j^2}{n-1}$.

A Crucial Distinction: What is being Minimized?

Although both PCA and Linear Regression (Least Squares) fit a line to data, their objectives and the errors they minimize are fundamentally different.

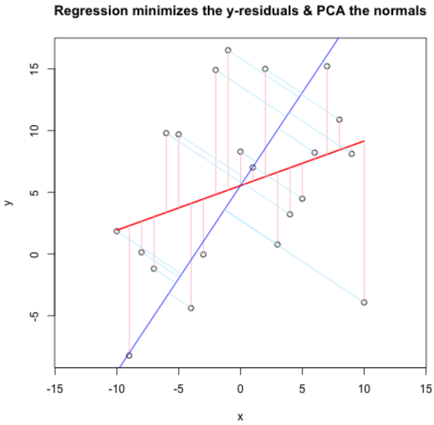
Principal Component Analysis

- **Goal:** Describe the data, find the best low-dimensional representation.
- **Error:** Minimizes the **orthogonal distance** from each point to the line (the principal component).
- **Symmetry:** Treats all variables equally; there is no distinction between dependent and independent variables.

Linear Regression

- **Goal:** Predict a specific variable (y) from others (x).
- **Error:** Minimizes the **vertical distance** from each point to the regression line.
- **Asymmetry:** Assumes x is a predictor and y is a response. Swapping them changes the result.

Visualizing the Difference



Key Takeaway

The line you get from PCA and the line you get from Linear Regression are generally **not the same**. They solve different problems.

Summary

- **PCA** is a powerful tool for dimensionality reduction that finds the directions of maximum variance in a dataset.

Summary

- **PCA** is a powerful tool for dimensionality reduction that finds the directions of maximum variance in a dataset.
- These directions, the **principal components**, are the eigenvectors of the data's covariance matrix.

Summary

- **PCA** is a powerful tool for dimensionality reduction that finds the directions of maximum variance in a dataset.
- These directions, the **principal components**, are the eigenvectors of the data's covariance matrix.
- A more stable way to find them is by computing the SVD of the centered data matrix; the principal components are the right singular vectors (V).

Summary

- **PCA** is a powerful tool for dimensionality reduction that finds the directions of maximum variance in a dataset.
- These directions, the **principal components**, are the eigenvectors of the data's covariance matrix.
- A more stable way to find them is by computing the SVD of the centered data matrix; the principal components are the right singular vectors (V).
- PCA is fundamentally different from Linear Regression, as it minimizes orthogonal projection errors, not vertical prediction errors.