Introduction to SVD

Singular Value Decomposition (SVD) decomposes a matrix A into:

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

where:

- U: Orthogonal matrix ($m \times m$) with left singular vectors.
- Σ: Diagonal matrix ($m \times n$) with singular values ($\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$).
- $V^\top$: Orthogonal matrix ($n \times n$) with right singular vectors.

Example

Find a singular value decomposition of

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

$$A^T A = \begin{bmatrix} 9 & -9 \\ -9 & 9 \end{bmatrix}.$$

The eigenvalues of $A^T A$ are 18 and 0 with corresponding unit eigenvectors

$$v_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \; v_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

Hence,

$$V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Example

To construct U, first construct Av_1 and Av_2 :

$$Av_1 = \begin{bmatrix} 2/\sqrt{2} \\ -4/\sqrt{2} \\ 4/\sqrt{2} \end{bmatrix}, \quad Av_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

$$u_1 = \frac{1}{3\sqrt{2}} Av_1 = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \end{bmatrix}$$

In order to write U, we need to extend the set $\{u_1\}$ to orthonormal basis for $\mathbb{R}^3$.HOW?

$$U = \begin{bmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{45} \\ 2/3 & 0 & 5/\sqrt{45} \end{bmatrix}.$$

- Find Orthogonal Compliment
- Use Gram Schmidth

The Four Fundamental Subspaces

The four fundamental subspaces of A are:

- **Column Space** ($\mathcal{C}(A)$): Spanned by the first r columns of U.
- **Row Space** ($\mathcal{C}(A^\top)$): Spanned by the first r columns of V.
- **Null Space** ($\mathcal{N}(A)$): Spanned by the last $n - r$ columns of V.
- **Left Null Space** ($\mathcal{N}(A^\top)$): Spanned by the last $m - r$ columns of U.

Key Property:

$$r = \text{rank}(A) = \text{Number of non-zero singular values.}$$

Relationships via SVD

$$
\begin{aligned}
\mathcal{C}(A) &= \text{Span}\{u_1, u_2, \ldots, u_r\} \quad \text{(Columns of } U\text{)} \\
\mathcal{C}(A^\top) &= \text{Span}\{v_1, v_2, \ldots, v_r\} \quad \text{(Columns of } V\text{)} \\
\mathcal{N}(A) &= \text{Span}\{v_{r+1}, \ldots, v_n\} \quad \text{(Zero singular values in } \Sigma\text{)} \\
\mathcal{N}(A^\top) &= \text{Span}\{u_{r+1}, \ldots, u_m\} \quad \text{(Zero singular values in } \Sigma\text{)}
\end{aligned}
$$

Geometric Interpretation

- U: Rotates the input space into the row and left null spaces.
- Σ: Scales along the principal axes (singular values).
- V: Rotates the input space into the column and null spaces.

Example: Four Fundamental Subspaces

Consider $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$:

- SVD: $U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, $V = I_2$.
- Column Space: $\mathcal{C}(A) = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$.
- Row Space: $\mathcal{C}(A^\top) = \text{Span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$.
- Null Space: $\mathcal{N}(A) = \{0\}$.
- Left Null Space: $\mathcal{N}(A^\top) = \text{Span}\left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$.

Summary

SVD and Fundamental Subspaces:

- The SVD explicitly reveals the structure of all four fundamental subspaces.
- The rank r of the matrix determines the dimensionality of the column and row spaces.
- The zero singular values define the null spaces.

Applications:

- Data science (e.g., PCA).
- Signal processing (e.g., noise reduction).
- Numerical stability in solving linear systems.

Theorem

Let A be an $m \times n$rix with singular vales $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$. Let $u_1,\ u_2,\ ,...,u_r$ be left singular vectors and let $v_1,\ v_2,\ ,...,v_r$ be right singular vectors of A corresponding to these singular vales. The,

$$A = \sigma_1 u_1 v_1^T + ... + \sigma_r u_r v_r^T.$$

Remark

Let A be a real $m \times n$ matrix, and let $A = U\Sigma V^T$ is any SVD for A. Then,

$$A^+(leftinverse) = V\Sigma^+ U^T.$$

where

$$\Sigma^+ = \begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}_{n\times m}.$$

Principal Component Analysis (PCA)

Sara Aziz

November 26, 2024

Motivation for PCA

- **High-Dimensional Data Challenges:**
  - Many datasets contain a large number of features.
  - Hard to visualize and interpret high-dimensional data.
  - Computationally expensive to process.
- **Correlation Among Features:**
  - Features may be highly correlated, leading to redundancy.
  - PCA identifies the most informative combinations of features.

Reasons to Study PCA

- **Dimensionality Reduction:**
  - Reduces the number of features while preserving variance.
  - Makes data easier to analyze and visualize.
- **Improves Machine Learning Models:**
  - Reduces overfitting by eliminating redundant features.
  - Speeds up training and inference in high-dimensional spaces.
- **Noise Reduction:**
  - Filters out less significant features, focusing on the core patterns.
- **Applications Across Domains:**
  - Image compression, gene expression analysis, recommendation systems, etc.
  - Widely used in data visualization and exploratory analysis.

Key Insight of PCA

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components.

- PCA finds new axes (principal components) that maximize variance.
- These axes are uncorrelated (orthogonal) and ordered by the amount of variance they explain.
- Focuses on patterns in data and simplifies analysis.

Takeaway: PCA is an essential tool for understanding and working with high-dimensional data.

Objective of PCA

- Find a new set of orthogonal axes (principal components).
- Maximize variance along the first axis (PC1).
- Subsequent axes maximize remaining variance while being orthogonal.

Goal: Reduce dimensionality while retaining the most variance.

Step 1: Data Representation

- Dataset of n observations with d features:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}$$

- Center the data:(Standardised Matrix)

$$X_{\text{centered}} = X - \text{Mean}(X)$$

Step 2: Compute the Covariance Matrix

- Covariance measures relationships between features:

$$C = \frac{1}{n-1} X_{\text{centered}}^{\top} X_{\text{centered}}$$

- Properties of the covariance matrix:
    - Symmetric: $C^{\top} = C$.
    - Size: $d \times d$ (features by features).

Step 3: Eigenvalue Problem

- Solve the eigenvalue equation for C:

$$\mathrm{Cv} = \lambda \mathrm{v}$$

- λ: Eigenvalues (variance explained by components).
- v: Eigenvectors (directions of principal components).
- Eigenvectors are orthogonal (uncorrelated).

Step 4: Maximizing Variance

- Find the direction v_1 that maximizes variance:

$$\max_{v} v^\top C v \quad \text{subject to } \|v\| = 1$$

- Solution: The eigenvector corresponding to the largest eigenvalue λ_1.
- Subsequent principal components correspond to smaller eigenvalues.

Step 5: Projection of Data

- Project the centered data onto the top k eigenvectors:

$$Z = X_{\text{centered}} V_k$$

- Z: Transformed data in the reduced k-dimensional space.
- V_k: Matrix of the top k eigenvectors.

Result: A lower-dimensional representation of the data.

Key Insights from PCA

- **Eigenvectors:** Define new axes (principal components).
- **Eigenvalues:** Represent variance along each principal component.
- **Orthogonality:** Ensures no redundancy among components.
- **Dimensionality Reduction:** Keep components with the highest eigenvalues.

Scenario: Customer Spending Analysis

Objective: Understand spending patterns of customers using PCA.

- Features:
  - Groceries (X_1)
  - Clothing (X_2)
- Data for three customers:

$$X = \begin{bmatrix} 2 & 1 \\ 4 & 3 \\ 6 & 5 \end{bmatrix}$$

Step 1: Standardize the Data

- Calculate the mean for each feature:

$$\text{Mean of } X_1 = 4, \quad \text{Mean of } X_2 = 3$$

- Center the data by subtracting the mean:

$$\mathrm{X}_{\text{centered}} = \begin{bmatrix} -2 & -2 \\ 0 & 0 \\ 2 & 2 \end{bmatrix}$$

Step 2: Covariance Matrix

- Compute the covariance matrix:

$$C = \frac{1}{n-1} X_{\text{centered}}^{\top} X_{\text{centered}}$$

- Result:

$$C = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

Step 3: Eigenvalues and Eigenvectors

- Eigenvalues:

$$\lambda_1 = 8, \quad \lambda_2 = 0$$

- Corresponding eigenvectors:

$$\mathsf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathsf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Step 4: Transform the Data

- Project the centered data onto the principal components:

$$Z = X_{\text{centered}} V$$

- Transformed data:

$$Z = \begin{bmatrix} -2\sqrt{2} & 0 \\ 0 & 0 \\ 2\sqrt{2} & 0 \end{bmatrix}$$

Step 5: Interpretation

- The first principal component (PC_1) captures all variance ($\lambda_1 = 8$).
- The second principal component (PC_2) has zero variance ($\lambda_2 = 0$).
- Data can be reduced to a single dimension (PC_1).

Conclusion: PCA effectively simplifies the dataset while retaining the most important information.

Scenario: Athlete Performance Analysis

Objective: Evaluate athletes' performance across multiple metrics.

- Features:
  - Speed (X_1)
  - Strength (X_2)
  - Stamina (X_3)
- Data for three athletes:

$$\mathrm{X} = \begin{bmatrix} 2 & 4 & 6 \\ 4 & 6 & 8 \\ 6 & 8 & 10 \end{bmatrix}$$

Step 1: Standardize the Data

- Compute the mean of each feature:

$$\text{Mean of } X_1 = 4, \quad \text{Mean of } X_2 = 6, \quad \text{Mean of } X_3 = 8$$

- Center the data:

$$\mathrm{X}_{\text{centered}} = \begin{bmatrix} -2 & -2 & -2 \\ 0 & 0 & 0 \\ 2 & 2 & 2 \end{bmatrix}$$

Step 2: Covariance Matrix

- Covariance matrix:

$$C = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}$$

Step 3: Eigenvalues and Eigenvectors

- Eigenvalues:

$$\lambda_1 = 12, \quad \lambda_2 = 0, \quad \lambda_3 = 0$$

- Eigenvector for λ_1:

$$\mathsf{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Step 4: Transform the Data

- Project the centered data onto PC_1:

$$Z = \begin{bmatrix} -2\sqrt{3} \\ 0 \\ 2\sqrt{3} \end{bmatrix}$$

Step 5: Interpretation

- **Principal Component 1 (PC1):** Captures all variance (12).
- **PC2 and PC3:** No variance, dataset lies entirely along PC_1.
- **Conclusion:** Data can be reduced to one dimension without information loss.

Scenario: Student Performance Analysis

- Three features:
  - Math (X1)
  - Science (X2)
  - English (X3)
- Scores for three students:

$$X = \begin{bmatrix} 80 & 90 & 70 \\ 85 & 85 & 75 \\ 90 & 95 & 80 \end{bmatrix}$$

Step 1: Standardize the Data

- Calculate the mean for each feature:

$$\text{Mean of } X_1 = 85, \quad \text{Mean of } X_2 = 90, \quad \text{Mean of } X_3 = 75$$

- Center the data:

$$\mathrm{X}_{\text{centered}} = \begin{bmatrix} -5 & 0 & -5 \\ 0 & -5 & 0 \\ 5 & 5 & 5 \end{bmatrix}$$

Step 2: Compute the Covariance Matrix

- Formula:

$$C = \frac{1}{n-1} X_{\text{centered}}^{\top} X_{\text{centered}}$$

- Covariance matrix:

$$C = \begin{bmatrix} 25 & 25 & 25 \\ 25 & 25 & 25 \\ 25 & 25 & 25 \end{bmatrix}$$

Step 3: Eigenvalues and Eigenvectors

- Eigenvalues:

$$\lambda_1 = 75, \quad \lambda_2 = 0, \quad \lambda_3 = 0$$

- Eigenvectors:

$$\mathrm{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathrm{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \quad \mathrm{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Step 4: Transform the Data

- Project the centered data onto the principal components:

$$Z = X_{\text{centered}} V$$

- Transformed data:

$$Z = \begin{bmatrix} -8.66 & 0 & 0 \\ 0 & -7.07 & 0 \\ 8.66 & 7.07 & 0 \end{bmatrix}$$

Step 5: Interpretation

- **Principal Component 1 (PC1):** Captures all the variance (75) and represents overall performance across subjects.
- **Principal Component 2 (PC2):** Minor variation among scores.
- **Principal Component 3 (PC3):** Zero variance; does not contribute to the data.

Conclusion: The data can be reduced to one dimension ($PC1$) while retaining all information.

Overview of PCA

- PCA reduces dimensionality while retaining variance in the data.
- It transforms correlated features into a set of uncorrelated components.
- Widely used for visualization, noise reduction, and feature extraction.

Step 1: Standardize the Data

- Compute the mean for each feature:

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Center the data:

$$X_{\text{centered}} = X - \text{Mean}$$

- (Optional) Scale the data to unit variance if features have different scales.

Step 2: Compute the Covariance Matrix

- Understand relationships between features.
- Formula:

$$C = \frac{1}{n-1} X_{\text{centered}}^{\top} X_{\text{centered}}$$

- The covariance matrix is symmetric, with diagonal elements representing variances.

Step 3: Find Eigenvalues and Eigenvectors

- Solve the eigenvalue problem for the covariance matrix:

$$\mathsf{C}\mathsf{v} = \lambda \mathsf{v}$$

- λ: Eigenvalues (variance explained by each component).
- v: Eigenvectors (directions of principal components).

Step 4: Select Principal Components

- Sort eigenvalues in descending order.
- Select the top k components that capture the most variance.
- Principal component matrix:

$$V = \begin{bmatrix} v_1 & v_2 & \cdots & v_k \end{bmatrix}$$

Step 5: Transform the Data

- Project the data onto the principal components:

$$Z = X_{\text{centered}} V$$

- Z is the reduced-dimensional representation of the data.

Step 6: Interpretation

- First principal component explains the highest variance.
- Each subsequent component explains remaining variance orthogonally.
- Use transformed data for analysis, modeling, or visualization.

Key Takeaway: PCA simplifies data while preserving essential patterns.