Principal Component Analysis

Dr. Chandranath Adak

Dept. of CSE, Indian Institute of Technology Patna

August 20, 2025

Necessity of Feature Reduction

- Avoid Curse of Dimensionality:
    - High-dimensional data often leads to sparse distributions, reducing model performance
- Improve Computational Efficiency:
    - Lower-dimensional representations reduce the time and resources required for training and inference
- Enhance Generalization:
    - Reducing noise and irrelevant features prevents over-fitting and improves model accuracy

Focus

Understand a feature reduction technique, called Principal Component Analysis (PCA)

Intuition for Selecting a New Set of Features

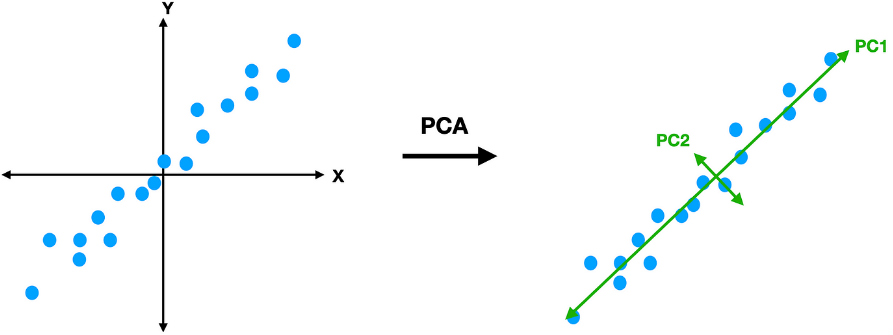
- Features should provide clear separability for data points:
  - Ensure high variance along each feature direction
- Each feature should contribute unique information:
  - Features must be uncorrelated (covariance between features $= 0$)
- Dimensionality reduction should preserve essential information:
  - Minimize information loss during the reduction process

PCA is an unsupervised linear feature reduction technique that aligns with the intuition for selecting a new set of features

Introducing Principal Component Analysis (PCA)

- **Clear Separability Through Variance:**
  - PCA identifies principal components by maximizing variance along new orthogonal axes, ensuring that the data points are distributed distinctly in the transformed feature space
- **Capturing Unique Information:**
  - By constructing uncorrelated principal components (zero covariance), PCA ensures that each component contributes non-redundant information about the data
- **Minimizing Information Loss:**
  - PCA ranks components based on their contribution to total variance, allowing for dimensionality reduction while preserving most of the essential information
- **Efficient Dimensionality Reduction:**
  - Reduces the number of features, addressing issues of high dimensionality such as sparsity, computational complexity, and overfitting

Visualization



Original Variables

Principle Components

- Consider $x_1, x_2, \ldots, x_m$ as m data points
- Each data point $x_i \in \mathbb{R}^n$ is represented as an n-dimensional vector
- Let $X_{m \times n}$ represent the data matrix, where $x_1, x_2, \ldots, x_m$ are the rows of X
- Assumption: The data is preprocessed to have 0-mean and unit variance:
  - Feature scaling is performed using standardization

- Consider $x_1, x_2, \dots, x_m$ as m data points
- Each data point $x_i \in \mathbb{R}^n$ is represented as an n-dimensional vector
- Let $X_{m \times n}$ represent the data matrix, where $x_1, x_2, \dots, x_m$ are the rows of X
- Assumption: The data is preprocessed to have 0-mean and unit variance:
  - Feature scaling is performed using standardization

PCA applies a linear transformation to project the data onto a new coordinate system

- Let $z_1, z_2, \dots, z_n$ denote a set of n orthonormal vectors forming the new basis
- Define Z as an $n \times n$ matrix where $z_1, z_2, \dots, z_n$ are the column vectors
- Using this new basis, each x_i can be expressed as: $x_i = \sum_{j=1}^{n} \alpha_{ij} z_j$

We aim to derive Z that satisfies feature reduction criteria while balancing dimensionality reduction and information retention

$$x_i = \alpha_{i1} z_1 + \alpha_{i2} z_2 + \ldots + \alpha_{in} z_n$$

The z_i's being orthonormal exhibit the following characteristics:

$$\langle z_i, z_j \rangle = 0, \quad \text{for} \quad i \neq j$$

$$||z_i|| = \langle z_i, z_i \rangle = 1, \quad \text{for} \quad i = 1, 2, \ldots, n$$

Thus, for an orthonormal basis, the coefficients α_{ij} can be expressed as:

$$\alpha_{ij} \quad = \quad \langle x_i, z_j \rangle \quad = \quad \begin{bmatrix} \longleftarrow & x_i^\top & \longrightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ z_j \\ \downarrow \end{bmatrix}$$

The transformed data $\hat{x}_i$ is given by:

$$\hat{x}_i^\top \quad = \quad x_i^\top Z \quad = \quad \begin{bmatrix} \longleftarrow & x_i^\top & \longrightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & \ldots & \uparrow \\ z_1 & z_2 & \ldots & z_n \\ & & \ldots & \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

The transformed data $\hat{x}_i$ is given by:

$$\hat{x}_i^\top = x_i^\top Z = \begin{bmatrix} \longleftarrow & x_i^\top & \longrightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ z_1 & z_2 & \cdots & z_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

Thus, the transformed data $\hat{X}$ is given by:

$$\hat{X} = XZ = \begin{bmatrix} \longleftarrow & x_1^\top & \longrightarrow \\ \longleftarrow & x_2^\top & \longrightarrow \\ & \vdots & \\ \longleftarrow & x_m^\top & \longrightarrow \end{bmatrix} \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ z_1 & z_2 & \cdots & z_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

We aim for the covariance between the features of the transformed data to be zero.

Covariance Matrix of X

The covariance matrix of X, denoted Σ, is given by:

$$\Sigma = \frac{1}{m} X^\top X$$

Each entry Σ_{ij} represents the covariance between the i^{th} and j^{th} columns of X. ▸▸ Proof

We aim for the covariance between the features of the transformed data to be zero.

Covariance Matrix of X

The covariance matrix of X, denoted Σ, is given by:

$$\Sigma = \frac{1}{m} X^\top X$$

Each entry Σ_{ij} represents the covariance between the i^{th} and j^{th} columns of X. ⏩ Proof

Covariance Matrix of Transformed Data $\hat{X}$

The covariance matrix of the transformed data $\hat{X}$, denoted $\hat{\Sigma}$, is given by:

$$\hat{\Sigma} = \frac{1}{m} \hat{X}^\top \hat{X}$$

Since X has zero-mean columns and $\hat{X} = XZ$, the columns of $\hat{X}$ will also have zero mean. ⏩ Proof

We aim for covariance between features of transformed data to be zero.

We can derive the following expression for the covariance matrix of the transformed data:

$$\hat{\Sigma} = \frac{1}{m}\hat{X}^\top \hat{X} = \frac{1}{m}(XZ)^\top (XZ) = \frac{1}{m} Z^\top X^\top X Z = Z^\top \left(\frac{1}{m} X^\top X \right) Z = Z^\top \Sigma Z$$

Each element $\hat{\Sigma}_{ij}$ of the covariance matrix $\hat{\Sigma}$ represents the covariance between the i^{th} and j^{th} columns of $\hat{X}$.

We aim for covariance between features of transformed data to be zero.

We can derive the following expression for the covariance matrix of the transformed data:

$$\hat{\Sigma} = \frac{1}{m}\hat{X}^\top \hat{X} = \frac{1}{m}(XZ)^\top (XZ) = \frac{1}{m}Z^\top X^\top XZ = Z^\top \left(\frac{1}{m}X^\top X\right) Z = Z^\top \Sigma Z$$

Each element $\hat{\Sigma}_{ij}$ of the covariance matrix $\hat{\Sigma}$ represents the covariance between the i^{th} and j^{th} columns of $\hat{X}$.
We want the following conditions to hold:

$$\hat{\Sigma}_{ij} = \left(\frac{1}{m}\hat{X}^\top \hat{X}\right)_{ij} = 0 \quad \text{for} \quad i \neq j$$

$$\hat{\Sigma}_{ij} = \left(\frac{1}{m}\hat{X}^\top \hat{X}\right)_{ij} \neq 0 \quad \text{for} \quad i = j$$

This implies that we want: $\hat{\Sigma} = \frac{1}{m}\hat{X}^\top \hat{X} = Z^\top \Sigma Z = \mathcal{D}$, where $\mathcal{D}$ is a diagonal matrix.

We aim to ensure that the covariance between the features of the transformed data is zero.
To achieve this, we want $Z^\top \Sigma Z$ to be a diagonal matrix $\mathcal{D}$.

- $\Sigma = \frac{1}{m} X^\top X$ is a square symmetric matrix of dimension $n \times n$.
- Z is an orthogonal matrix.

Which orthogonal matrix diagonalizes Σ?

We aim to ensure that the covariance between the features of the transformed data is zero.
To achieve this, we want $Z^\top \Sigma Z$ to be a diagonal matrix $\mathcal{D}$.

- $\Sigma = \frac{1}{m} X^\top X$ is a square symmetric matrix of dimension $n \times n$.
- Z is an orthogonal matrix.

Which orthogonal matrix diagonalizes Σ?

The matrix Z, whose columns are the eigenvectors of Σ. ⏩ Explanation

Why do the vectors in Z form a good basis?

We aim to ensure that the covariance between the features of the transformed data is zero.
To achieve this, we want $Z^\top \Sigma Z$ to be a diagonal matrix $\mathcal{D}$.

- $\Sigma = \frac{1}{m} X^\top X$ is a square symmetric matrix of dimension $n \times n$.
- Z is an orthogonal matrix.

Which orthogonal matrix diagonalizes Σ?

The matrix Z, whose columns are the eigenvectors of Σ. ⏩ Explanation

Why do the vectors in Z form a good basis?

- Because the eigenvectors of Σ are linearly independent:
  - A matrix $A \in \mathbb{R}^{n \times n}$ with distinct eigenvalues has linearly independent eigenvectors. ⏩ Explanation
- Because the eigenvectors of Σ are orthogonal:
  - The eigenvectors of a square symmetric matrix are orthogonal. ⏩ Explanation

PCA

This method is called Principal Component Analysis (PCA) for transforming the data to a new basis where the dimensions are non-redundant (low covariance) and not noisy (high variance).

In practice, we select only top-k dimensions along which the variance is high.

Self-study: Scree Plot

Our next objective is to minimize the information loss.

- Given n orthogonal linearly independent vectors $Z = \{z_1, z_2, \cdots, z_n\}$, we can represent x_i exactly as a linear combination of these vectors:

$$x_i = \sum_{j=1}^{n} \alpha_{ij} z_j$$

  where α_{ij} are the coefficients

- Top-k Dimensional Approximation: We are interested in reducing noisy and redundant dimensions by keeping only the top-k dimensions. The approximated vector $\hat{x}_i$ is given by:

$$\hat{x}_i = \sum_{j=1}^{k} \alpha_{ij} z_j$$

- We aim to select z_j such that the reconstruction error e is minimized:

$$e = \sum_{i=1}^{m} \left((x_i - \hat{x}_i)^\top (x_i - \hat{x}_i) \right)$$

  where m is the total number of data points

- The reconstruction error e is given by:

$$e = \sum_{i=1}^{m} (x_i - \hat{x}_i)^\top (x_i - \hat{x}_i)$$

Substituting $\hat{x}_i = \sum_{j=1}^{k} \alpha_{ij} z_j$, we get:

$$e = \sum_{i=1}^{m} \left(x_i - \sum_{j=1}^{k} \alpha_{ij} z_j \right)^\top \left(x_i - \sum_{j=1}^{k} \alpha_{ij} z_j \right)$$

Expanding, we obtain:

$$e = \sum_{i=1}^{m} \sum_{j=k+1}^{n} (\alpha_{ij} z_j)^\top (\alpha_{ij} z_j) = \sum_{i=1}^{m} \sum_{j=k+1}^{n} \left(\alpha_{ij}^2 \| z_j \|^2 \right) = \sum_{i=1}^{m} \sum_{j=k+1}^{n} \left(x_i^\top z_j \right)^2$$

- The reconstruction error e is given by:

$$e = \sum_{i=1}^{m} (x_i - \hat{x}_i)^\top (x_i - \hat{x}_i)$$

Substituting $\hat{x}_i = \sum_{j=1}^{k} \alpha_{ij} z_j$, we get:

$$e = \sum_{i=1}^{m} \left(x_i - \sum_{j=1}^{k} \alpha_{ij} z_j \right)^\top \left(x_i - \sum_{j=1}^{k} \alpha_{ij} z_j \right)$$

Expanding, we obtain:

$$e = \sum_{i=1}^{m} \sum_{j=k+1}^{n} (\alpha_{ij} z_j)^\top (\alpha_{ij} z_j) = \sum_{i=1}^{m} \sum_{j=k+1}^{n} \left(\alpha_{ij}^2 \| z_j \|^2 \right) = \sum_{i=1}^{m} \sum_{j=k+1}^{n} \left(x_i^\top z_j \right)^2$$

$$= \sum_{i=1}^{m} \sum_{j=k+1}^{n} \left(z_j^\top x_i \right) \left(x_i^\top z_j \right) = \sum_{j=k+1}^{n} z_j^\top \left(\sum_{i=1}^{m} x_i x_i^\top \right) z_j = \sum_{j=k+1}^{n} z_j^\top \left(m\Sigma \right) z_j$$

We aim to minimize the error:

$$\min_{z_{k+1}, z_{k+2}, \ldots, z_n} \quad \sum_{j=k+1}^{n} z_j^\top (m\Sigma)\, z_j$$

subject to the constraint:

$$z_j^\top z_j = 1 \quad \forall j = k+1, k+2, \ldots, n$$

We aim to minimize the error:

$$\min_{z_{k+1}, z_{k+2}, \ldots, z_n} \quad \sum_{j=k+1}^{n} z_j^\top (m\Sigma) z_j$$

subject to the constraint:

$$z_j^\top z_j = 1 \quad \forall j = k+1, k+2, \ldots, n$$

Solution

- The solution to the above problem is given by the eigenvectors corresponding to the smallest eigenvalues of Σ
- Thus, we select $Z = \{z_1, z_2, \ldots, z_n\}$ as the eigenvectors of Σ and retain only the top-k eigenvectors to express the data (or discard the eigenvectors $z_{k+1}, \ldots, z_n$).

⏩ Explanation

Our final goal is to maximize the variance along the chosen feature direction.

The i^{th} dimension of the transformed data $\hat{X}$ can be expressed as:

$$\hat{X}_i = Xz_i$$

The variance along this dimension is calculated as:

$$\hat{\Sigma}_{ii} = \frac{1}{m}\hat{X}_i^\top \hat{X}_i = \frac{1}{m}(Xz_i)^\top (Xz_i) = \frac{1}{m} z_i^\top X^\top X z_i = z_i^\top \left(\frac{1}{m}X^\top X\right) z_i = z_i^\top \lambda_i z_i$$

where z_i is the eigenvector of $\frac{1}{m}X^\top X$. Therefore, $\left(\frac{1}{m}X^\top X\right) z_i = \lambda_i z_i$

$$\hat{\Sigma}_{ii} = \lambda_i z_i^\top z_i = \lambda_i \qquad (\text{Since } z_i^\top z_i = 1)$$

Conclusion

Thus, the variance along the i^{th} dimension, which corresponds to the eigenvector of $\frac{1}{m}X^\top X$, is determined by the associated eigenvalue. Therefore, we make the correct choice by discarding the dimensions (eigenvectors) linked to smaller eigenvalues!

Summary

- Principal Component Analysis (PCA) is
  - A dimensionality reduction technique
  - Used to transform data into a new coordinate system
- The key objectives of PCA are:
  - **Maximize Variance**: It identifies new axes (principal components) that capture the maximum variance in the data, helping to retain important information while reducing dimensionality.
  - **Minimize Covariance**: PCA ensures that the transformed dimensions are uncorrelated, minimizing the covariance between them.
  - **Feature Reduction**: By selecting the top principal components, PCA reduces the number of dimensions required to represent the data, focusing on the most significant features.
- In essence, PCA helps in simplifying complex data, making it easier to analyze and visualize, while preserving its core structure.

Thank You!

The covariance matrix Σ of X is given by: $\Sigma = \frac{1}{m}X^\top X$
Each entry Σ_{ij} represents the covariance between i^{th} and j^{th} columns of X.

Let μ_i and μ_j represent the means of the i^{th} and j^{th} columns of X, respectively, which are both zero since the data has been preprocessed to have zero mean.

By the definition of covariance, we have:

$$\begin{aligned}
\Sigma_{ij} &= \frac{1}{m}\sum_{k=1}^{m}(X_{ki}-\mu_i)(X_{kj}-\mu_j) \\
&= \frac{1}{m}\sum_{k=1}^{m}X_{ki}X_{kj} \qquad\qquad (\because \mu_i = \mu_j = 0) \\
&= \frac{1}{m}X_i^\top X_j \\
&= \frac{1}{m}(X^\top X)_{ij}
\end{aligned}$$

▸▸ Go back

Zero-Mean Columns in $\hat{X}$

If X is a matrix such that its columns are zero-mean and $\hat{X} = XZ$, then the columns of $\hat{X}$ will also be zero-mean.

- Let $\mathbb{1}$ represent an m-dimensional column vector with entries equal to 1.
- The product $\mathbb{1}^\top X$ results in a row vector, where the i^{th} entry is the sum of the i^{th} column of X.
- Since the columns of X are zero-centered, we have $\mathbb{1}^\top X = 0$.

Consider the following:

$$\mathbb{1}^\top \hat{X} = \mathbb{1}^\top XZ = (\mathbb{1}^\top X)Z = 0$$

Thus, the transformed matrix $\hat{X}$ also has columns with zero-sum.

▸▸ Go back

- Let $u_1, u_2, \ldots, u_n$ be the eigenvectors of a matrix A and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the corresponding eigenvalues
- Consider a matrix U whose columns are $u_1, u_2, \ldots, u_n$

$$\begin{aligned}
AU &= A\begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \ldots & u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ Au_1 & Au_2 & \ldots & Au_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \\
&= \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \lambda_1 u_1 & \lambda_2 u_2 & \ldots & \lambda_n u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \\
&= \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ u_1 & u_2 & \ldots & u_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{bmatrix} = U\Lambda
\end{aligned}$$

$\Lambda = U^{-1}AU$

U^{-1} exists if columns of U are linearly independent

⏩ Go back

Λ is a diagonal matrix

Theorem

The eigenvectors of a matrix $A \in \mathbb{R}^{n\times n}$ with distinct eigenvalues are linearly independent.

- Let $u_1, u_2, \ldots, u_r$ be the eigenvectors corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$ of an $n \times n$ matrix A. The set $\{u_1, u_2, \ldots, u_r\}$ is linearly independent.
- **Assumption:** Suppose $U = \{u_1, u_2, \ldots, u_r\}$ is linearly dependent. Let us order the set U such that u_{p+1} is the first vector in the set that can be expressed as:

$$u_{p+1} = c_1 u_1 + c_2 u_2 + \ldots + c_p u_p, \quad c_1, c_2, \ldots, c_p \in \mathbb{R}, \quad (1)$$

- **Being an eigenvector:**

$$A u_{p+1} = \lambda_{p+1} u_{p+1}, \quad (2)$$

- Multiplying A on both sides of (1):

$$A u_{p+1} = c_1 (A u_1) + c_2 (A u_2) + \ldots + c_p (A u_p)$$

- Using the eigenvalue equation $Au_i = \lambda_i u_i$, we have:

$$Au_{p+1} = c_1(\lambda_1 u_1) + c_2(\lambda_2 u_2) + \ldots + c_p(\lambda_p u_p), \quad (3)$$

- Multiplying λ_{p+1} on both sides of (1) and subtracting it from (3):

$$0 = c_1(\lambda_1 - \lambda_{p+1})u_1 + c_2(\lambda_2 - \lambda_{p+1})u_2 + \ldots + c_p(\lambda_p - \lambda_{p+1})u_p, \quad (4)$$

- Since the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p, \lambda_{p+1}$ are distinct, $(\lambda_i - \lambda_{p+1}) \neq 0$ for all $i = 1, 2, \ldots, p$. This implies that the coefficients $c_1, c_2, \ldots, c_p$ must all be zero for the equality to hold. This contradicts the assumption that u_{p+1} is linearly dependent, proving that $\{u_1, u_2, \ldots, u_r\}$ is linearly independent.

⏩ Go back

Theorem

The eigenvectors of a square symmetric matrix corresponding to distinct eigenvalues are orthogonal.

Let x and y be eigenvectors of a symmetric matrix $A \in \mathbb{R}^{n\times n}$, corresponding to eigenvalues λ_1 and λ_2, respectively, with $\lambda_1 \neq \lambda_2$. Then:

$$Ax = \lambda_1 x \quad \text{and} \quad Ay = \lambda_2 y.$$

Since A is symmetric $(A = A^\top)$:

$$y^\top Ax = \lambda_1 y^\top x \quad \text{and} \quad x^\top A^\top y = \lambda_2 x^\top y.$$

Subtracting these two equations:

$$y^\top Ax - x^\top A^\top y = \lambda_1 y^\top x - \lambda_2 x^\top y.$$

Using the symmetry of the scalar product $(y^\top x = x^\top y)$:

$$0 = (\lambda_1 - \lambda_2) y^\top x.$$

Since $\lambda_1 \neq \lambda_2$, it must be that:

$$y^\top x = 0.$$

Thus, x and y are orthogonal.

⏩ Go back

Theorem

Each covariance matrix is positive semi-definite:

$$a^\top \Sigma_{XX} a \geq 0 \quad \text{for all } a \in \mathbb{R}^n.$$

Proof:

The covariance matrix of X can be expressed in terms of expected values as:

$$\Sigma_{XX} = \Sigma(X) = \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top]. \tag{1}$$

A matrix M is positive semi-definite if and only if:

$$M \text{ is pos. semi-def.} \iff x^\top M x \geq 0 \quad \text{for all } x \in \mathbb{R}^n. \tag{2}$$

For an arbitrary real column vector $a \in \mathbb{R}^n$, we have:

$$a^\top \Sigma_{XX} a = a^\top \mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top] a. \tag{3}$$

Using the linearity of the expected value, we can rewrite this as:

$$a^\top \Sigma_{XX} a = \mathbb{E}[a^\top (X - \mathbb{E}[X])(X - \mathbb{E}[X])^\top a]. \tag{4}$$

Define the scalar random variable:

$$Y = a^\top (X - \mu_X), \tag{5}$$

where $\mu_X = \mathbb{E}[X]$, and note that:

$$a^\top (X - \mu_X) = (X - \mu_X)^\top a. \tag{6}$$

Substituting this into equation (4), we get:

$$a^\top \Sigma_{XX} a = \mathbb{E}[Y^2]. \tag{7}$$

Since Y^2 is a non-negative random variable, and the expected value of a non-negative random variable is also non-negative, we conclude:

$$a^\top \Sigma_{XX} a \geq 0. \tag{8}$$

Thus, Σ_{XX} is positive semi-definite.

⏩ Go back

Theorem

If A is a square symmetric $N \times N$ matrix, then:

- The solution to the following optimization problem:

$$\max_{x} x^{\top} A x \quad \text{s.t.} \quad \|x\| = 1$$

  is given by the eigenvector corresponding to the largest eigenvalue of A.

- The solution to the following optimization problem:

$$\min_{x} x^{\top} A x \quad \text{s.t.} \quad \|x\| = 1$$

  is given by the eigenvector corresponding to the smallest eigenvalue of A.

Solving the Optimization Problem Using Lagrange Multipliers

This is a constrained optimization problem that can be solved using Lagrange Multipliers:

$$\mathcal{L} = x^\top Ax - \lambda(x^\top x - 1)$$

Taking the derivative with respect to x:

$$\frac{\partial \mathcal{L}}{\partial x} = 2Ax - \lambda(2x) = 0 \implies Ax = \lambda x$$

Hence, x must be an eigenvector of A with eigenvalue λ.

Critical Points

Multiplying both sides of $Ax = \lambda x$ by $x^\top$:

$$x^\top Ax = \lambda x^\top x = \lambda \quad (\text{since } x^\top x = 1).$$

Therefore, the critical points of this constrained problem are the eigenvalues of A.

Conclusion

- The **maximum value** of $x^\top Ax$ is the largest eigenvalue of A.
- The **minimum value** of $x^\top Ax$ is the smallest eigenvalue of A.

▸▸ Go back