Positive Semi-Definiteness of the Covariance Matrix

Theorem

Let X be an $m \times n$ real data matrix, and let $\tilde{X}$ be the centered matrix obtained by subtracting the column means from X. Then the sample covariance matrix

$$C = \frac{1}{m} \tilde{X}^\top \tilde{X}$$

is symmetric and positive semi-definite. Consequently, the correlation matrix (which is the covariance matrix of standardized data) is also positive semi-definite.

Proof

1. Symmetry

Since

$$C^\top = \left(\frac{1}{m} \tilde{X}^\top \tilde{X} \right)^\top = \frac{1}{m} \tilde{X}^\top \tilde{X} = C,$$

the matrix C is symmetric.

2. Positive Semi-Definiteness

Let $\mathbf{v} \in \mathbb{R}^n$ be an arbitrary vector. Consider the quadratic form:

$$\mathbf{v}^\top C \mathbf{v} = \mathbf{v}^\top \left(\frac{1}{m} \tilde{X}^\top \tilde{X} \right) \mathbf{v} = \frac{1}{m} (\tilde{X}\mathbf{v})^\top (\tilde{X}\mathbf{v}) = \frac{1}{m} \|\tilde{X}\mathbf{v}\|^2.$$

The squared Euclidean norm $\|\tilde{X}\mathbf{v}\|^2$ is always nonnegative, so

$$\mathbf{v}^\top C \mathbf{v} \geq 0 \quad \text{for all } \mathbf{v} \in \mathbb{R}^n.$$

Therefore, C is positive semi-definite.

3. Correlation Matrix

The correlation matrix is obtained by first standardizing each column of X to have mean 0 and variance 1, resulting in a new centered matrix $\tilde{X}_{\text{std}}$. Its covariance matrix is

$$R = \frac{1}{m}\tilde{X}_{\text{std}}^{\top}\tilde{X}_{\text{std}},$$

which is of the same form as C. Hence, R is also symmetric and positive semi-definite.

When is C Positive Definite?

The matrix C is positive definite if and only if $\|\tilde{X}\mathbf{v}\|^2 > 0$ for all $\mathbf{v} \neq \mathbf{0}$, which occurs precisely when the columns of $\tilde{X}$ are linearly independent (i.e., $\tilde{X}$ has full column rank). This requires $m \geq n$ and no perfect multicollinearity among the features.

Standard Equation of an Ellipse

In two dimensions, the standard form of an ellipse centered at the origin, with axes aligned to the coordinate axes, is

$$\frac{y_1^2}{a^2} + \frac{y_2^2}{b^2} = 1,$$

where:

- $a > 0$ is the **semi-axis length** along the y_1-axis,
- $b > 0$ is the **semi-axis length** along the y_2-axis. Indeed, setting $y_2 = 0$ gives $y_1 = \pm a$, and setting $y_1 = 0$ gives $y_2 = \pm b$.

Quadratic Form Equation

Suppose we have the equation

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 = 1, \quad \lambda_1 > 0,\ \lambda_2 > 0.$$

We wish to identify the semi-axis lengths.

Rewrite the equation as:

$$\frac{y_1^2}{1/\lambda_1} + \frac{y_2^2}{1/\lambda_2} = 1.$$

To match the standard ellipse form, express the denominators as squares:

$$\frac{y_1^2}{\left(\frac{1}{\sqrt{\lambda_1}}\right)^2} + \frac{y_2^2}{\left(\frac{1}{\sqrt{\lambda_2}}\right)^2} = 1.$$

Conclusion

This is the standard equation of an ellipse with:

$$\text{semi-axis along } y_1 = \frac{1}{\sqrt{\lambda_1}}, \qquad \text{semi-axis along } y_2 = \frac{1}{\sqrt{\lambda_2}}.$$

Intuition

A larger eigenvalue λ_i means the quadratic form grows more rapidly in the y_i-direction. To maintain the value 1, the coordinate y_i must be smaller. Hence:

$$\lambda_i \uparrow \quad \Rightarrow \quad \text{semi-axis length } \frac{1}{\sqrt{\lambda_i}} \downarrow .$$

Example

Let $\lambda_1 = 4$, $\lambda_2 = 1$. Then:

$$4y_1^2 + y_2^2 = 1.$$

- When $y_2 = 0$: $4y_1^2 = 1 \Rightarrow y_1 = \pm\frac{1}{2} = \pm\frac{1}{\sqrt{4}}$, - When $y_1 = 0$: $y_2^2 = 1 \Rightarrow y_2 = \pm 1 = \pm\frac{1}{\sqrt{1}}$.

Thus, the semi-axes are $\frac{1}{2}$ and 1, as predicted.

Introduction

Let A be an $n \times n$ real symmetric positive semi-definite matrix. The quadratic form

$$Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}$$

defines a geometric shape via its level set $\mathbf{x}^\top A\mathbf{x} = 1$. We explain why this shape is an ellipsoid (when A is positive definite) and why its principal axes align with the eigenvectors of A.

1. Quadratic Forms and Level Sets

- If A is **positive definite** (all eigenvalues > 0), the set $\{\mathbf{x} : \mathbf{x}^\top A\mathbf{x} = 1\}$ is an **ellipsoid**. - If A is **positive semi-definite but singular** (some eigenvalues $= 0$), the set is a **degenerate ellipsoid** (e.g., a line, plane, or empty set).

2. Diagonalization via the Spectral Theorem

Since A is symmetric, the Spectral Theorem gives

$$A = Q\Lambda Q^\top,$$

where:

- $Q \in \mathbb{R}^{n\times n}$ is orthogonal ($Q^\top Q = I$); its columns $\mathbf{q}_1, \dots, \mathbf{q}_n$ are orthonormal eigenvectors of A,
- $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, with $\lambda_i \geq 0$ (since $A \succeq 0$).

3. Change of Coordinates

Substitute into the quadratic form:

$$\mathbf{x}^\top A\mathbf{x} = \mathbf{x}^\top Q\Lambda Q^\top \mathbf{x} = (Q^\top \mathbf{x})^\top \Lambda (Q^\top \mathbf{x}).$$

Define $\mathbf{y} = Q^\top \mathbf{x}$. Since Q is orthogonal, this is a rotation (or reflection) of coordinates. The equation becomes:

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 = 1.$$

4. Principal Axes Are Eigenvectors

In $\mathbf{y}$-coordinates, the ellipsoid is axis-aligned—the coordinate axes $y_1, \dots, y_n$ are its principal axes. But $\mathbf{x} = Q\mathbf{y}$, so the i-th $\mathbf{y}$-axis corresponds to the direction of the i-th column of Q in $\mathbf{x}$-space. Since the columns of Q are the eigenvectors of A, we conclude:

The principal axes of the ellipsoid are in the directions of the eigenvectors of A.

5. Axis Lengths

Set all $y_j = 0$ for $j \neq i$. Then:

$$\lambda_i y_i^2 = 1 \quad \Rightarrow \quad y_i = \pm \frac{1}{\sqrt{\lambda_i}}.$$

Because $\mathbf{x} = Q\mathbf{y}$ is a rotation (distance-preserving), the semi-axis length in the direction of eigenvector $\mathbf{q}_i$ is also $1/\sqrt{\lambda_i}$.

- Larger λ_i shorter axis (squished),
- Smaller λ_i longer axis (stretched).

6. Example in 2D

Let

$$A = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}.$$

- Eigenvalues: $\lambda_1 = 8$, $\lambda_2 = 2$, - Eigenvectors: $\mathbf{q}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{q}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

The equation $\mathbf{x}^\top A\mathbf{x} = 1$ describes an ellipse with:

- Minor axis (length $1/\sqrt{8} \approx 0.35$) in direction $\mathbf{q}_1$,
- Major axis (length $1/\sqrt{2} \approx 0.71$) in direction $\mathbf{q}_2$. The ellipse is rotated 45°—exactly the eigenvector directions.

Conclusion

The eigenvectors of a symmetric positive definite matrix A reveal the intrinsic geometry of the ellipsoid $\mathbf{x}^\top A\mathbf{x} = 1$: they are the directions of its principal axes, and the eigenvalues determine the axis lengths. This is why eigenvectors are called **principal axes**.

Introduction

In Principal Component Analysis (PCA), the **direction of maximum variance** in the data coincides with the **longest principal axis** of the associated data ellipsoid. We explain this deep connection between statistics and geometry.

1. The Data Ellipsoid

Let the data be centered (mean zero) with sample covariance matrix C. The **data ellipsoid** is defined by

$$\mathbf{x}^\top C^{-1}\mathbf{x} = 1.$$

This ellipsoid captures the shape and orientation of the data cloud.

2. Variance Along a Direction

For any unit vector $\mathbf{u} \in \mathbb{R}^n$, the variance of the data projected onto $\mathbf{u}$ is

$$\mathrm{Var}(\mathbf{u}^\top \mathbf{x}) = \mathbf{u}^\top C\mathbf{u}.$$

3. Maximizing Variance

The direction of maximum variance solves

$$\max_{\|\mathbf{u}\|=1} \mathbf{u}^\top C\mathbf{u}.$$

Using Lagrange multipliers, this leads to the eigenvalue problem

$$C\mathbf{u} = \lambda\mathbf{u}.$$

The maximum is achieved at the eigenvector $\mathbf{u}_1$ corresponding to the largest eigenvalue λ_1 of C.

4. Geometry of the Data Ellipsoid

The ellipsoid $\mathbf{x}^\top C^{-1}\mathbf{x} = 1$ has:

- **Principal axes** in the directions of the eigenvectors of C (since C and C^{-1} share eigenvectors),
- **Semi-axis lengths** equal to $\sqrt{\lambda_i}$, where λ_i are the eigenvalues of C. Indeed, if $C\mathbf{u}_i = \lambda_i\mathbf{u}_i$, then $C^{-1}\mathbf{u}_i = \frac{1}{\lambda_i}\mathbf{u}_i$, so the axis length in direction $\mathbf{u}_i$ is $1/\sqrt{1/\lambda_i} = \sqrt{\lambda_i}$.

5. The Connection

- The **longest axis** of the ellipsoid corresponds to the **largest eigenvalue** λ_1,
- This axis lies in the direction of the eigenvector $\mathbf{u}_1$,
- But $\mathbf{u}_1$ is precisely the **direction of maximum variance**.

6. Geometric Intuition

- High variance $\Rightarrow$ data points are spread out $\Rightarrow$ ellipsoid is long in that direction,
- Low variance $\Rightarrow$ data points are clustered $\Rightarrow$ ellipsoid is short in that direction. Thus, the ellipsoid's shape directly reflects the variance structure of the data.

Conclusion

The direction of maximum variance and the longest principal axis of the data ellipsoid are one and the same: the eigenvector of the covariance matrix C corresponding to its largest eigenvalue. This elegant correspondence is the geometric foundation of Principal Component Analysis.