A Complete PCA Example in Linear Algebra

Fei Ge

September 16, 2025

Conceptual Overview: Eigenvalues and Eigenvectors

At a high level, eigenvalues and eigenvectors reveal the fundamental structure of a square matrix. They describe how a linear transformation (represented by the matrix) acts on space.

- **Eigenvectors** are special, non-zero vectors that do not change their direction when the matrix transformation is applied to them. They are the "axes of transformation," pointing in directions that are only stretched or shrunk.

- **Eigenvalues** are the scalar values corresponding to each eigenvector. An eigenvalue tells you by exactly how much its eigenvector is stretched or shrunk. A positive eigenvalue means stretching, while a negative one means stretching and flipping direction.

Mathematically, they are defined by the equation:

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

Where $\mathbf{A}$ is the matrix, $\mathbf{v}$ is the eigenvector, and λ is the corresponding eigenvalue.

In the context of Principal Component Analysis (PCA), we perform an eigen-decomposition on a covariance matrix. The eigenvectors represent the principal components (the directions of maximum variance in the data), and the eigenvalues represent the magnitude of this variance.

---

1 Step 1: The Raw Data

We start with a dataset for 5 students and 2 variables: **Study Hours** and **Exam Score**.

Student	Study Hours (x)	Exam Score (y)
A	2	3
B	3	5
C	5	6
D	7	8
E	8	8

2 Step 2: Center the Data

The first step in PCA is to center the data by subtracting the mean of each variable from all its observations. This ensures that the analysis focuses on the variance within the data.

- **Mean of Study Hours** $= (2 + 3 + 5 + 7 + 8)/5 = 5$

- **Mean of Exam Score** $= (3 + 5 + 6 + 8 + 8)/5 = 6$

This gives us the **centered data matrix (C)**:

$$
\mathbf{C} = \begin{bmatrix} 2-5 & 3-6 \\ 3-5 & 5-6 \\ 5-5 & 6-6 \\ 7-5 & 8-6 \\ 8-5 & 8-6 \end{bmatrix} = \begin{bmatrix} -3 & -3 \\ -2 & -1 \\ 0 & 0 \\ 2 & 2 \\ 3 & 2 \end{bmatrix}
$$

3 Step 3: Calculate the Covariance Matrix

With the data centered, we now compute the covariance matrix. The formula is $\text{Cov} = \frac{1}{n-1}\mathbf{C}^T\mathbf{C}$, where $n = 5$.

3.1 Calculate $\mathbf{C}^T\mathbf{C}$

The sum of squares and cross-products matrix is:

$$
\mathbf{C^T C} = \begin{bmatrix} -3 & -2 & 0 & 2 & 3 \\ -3 & -1 & 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} -3 & -3 \\ -2 & -1 \\ 0 & 0 \\ 2 & 2 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 26 & 21 \\ 21 & 18 \end{bmatrix}
$$

3.2 Divide by (n-1)

We divide by $5 - 1 = 4$ to get the sample covariance matrix, which we will call $\mathbf{A}$.

$$
\text{Covariance Matrix (A)} = \frac{1}{4}\begin{bmatrix} 26 & 21 \\ 21 & 18 \end{bmatrix} = \begin{bmatrix} 6.5 & 5.25 \\ 5.25 & 4.5 \end{bmatrix}
$$

4 Step 4: Calculate the Eigenvalues (λ)

The eigenvalues of the covariance matrix represent the variance along each principal component. We find them by solving the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$.

$$
(6.5 - \lambda)(4.5 - \lambda) - (5.25)^2 = 0
$$

This expands to the quadratic equation:

$$
\lambda^2 - 11\lambda + 1.6875 = 0
$$

Using the quadratic formula, we find the two eigenvalues:

- $\lambda_1 \approx 10.84$

- $\lambda_2 \approx 0.16$

5 Step 5: Calculate the Eigenvectors (Principal Components)

The eigenvectors of the covariance matrix give the direction of the principal components. We solve $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$ for each eigenvalue.

5.1   For $\lambda_1 \approx 10.84$ (PC1)

1. **Set up the equation**: $(\mathbf{A} - 10.84\mathbf{I})\mathbf{v} = 0$ gives the system $-4.34v_1 + 5.25v_2 = 0$.

2. **Find the relationship**: This simplifies to $v_2 \approx 0.827v_1$.

3. **Normalize**: We choose a vector in this direction like $\begin{bmatrix} 1 \\ 0.827 \end{bmatrix}$ and divide by its length, which is $\sqrt{1^2 + 0.827^2} \approx 1.298$. The normalized eigenvector is:

$$\mathbf{PC1} \approx \begin{bmatrix} 0.77 \\ 0.64 \end{bmatrix}$$

5.2   For $\lambda_2 \approx 0.16$ (PC2)

1. **Set up the equation**: $(\mathbf{A} - 0.16\mathbf{I})\mathbf{v} = 0$ gives the system $6.34v_1 + 5.25v_2 = 0$.

2. **Find the relationship**: This simplifies to $v_2 \approx -1.208v_1$.

3. **Normalize**: We choose a vector in this direction like $\begin{bmatrix} 1 \\ -1.208 \end{bmatrix}$ and divide by its length, which is $\sqrt{1^2 + (-1.208)^2} \approx 1.568$. The normalized eigenvector is:

$$\mathbf{PC2} \approx \begin{bmatrix} 0.64 \\ -0.77 \end{bmatrix}$$

6   Step 6: Final Result Summary

The eigenvalues tell us how much variance is captured by each principal component.

Component	Eigenvalue (Variance)	% of Variance	Eigenvector (Direction)
PC1	10.84	98.5%	$\begin{bmatrix} 0.77 \\ 0.64 \end{bmatrix}$
PC2	0.16	1.5%	$\begin{bmatrix} 0.64 \\ -0.77 \end{bmatrix}$

This result shows that the first principal component (PC1) captures 98.5% of the total variance in the data. It represents the strong positive relationship between study hours and exam scores.