Worked-Out Example: Applying PCA with Covariance Matrix and Eigenvalues

Let’s dive deeper into the example provided and compute everything step-by-step.

Step 1: Input Matrix

We start with the matrix of study attributes (Importance and Time Required):

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

Step 2: Mean-Center the Data

To standardize the data:

- **Mean of Importance (x)**:

$$\bar{x} = \frac{3+3+2+3+2}{5} = 2.6$$

- **Mean of Time Required (y)**:

$$\bar{y} = \frac{6+5+4+5+6}{5} = 5.2$$

Subtract the mean from each column:

$$A' = \begin{bmatrix} 3-2.6 & 6-5.2 \\ 3-2.6 & 5-5.2 \\ 2-2.6 & 4-5.2 \\ 3-2.6 & 5-5.2 \\ 2-2.6 & 6-5.2 \end{bmatrix} = \begin{bmatrix} 0.4 & 0.8 \\ 0.4 & -0.2 \\ -0.6 & -1.2 \\ 0.4 & -0.2 \\ -0.6 & 0.8 \end{bmatrix}$$

Step 3: Compute Covariance Matrix

The covariance matrix measures relationships between attributes (columns):

$$\text{Covariance Matrix: } \text{Cov}(A') = [\begin{matrix} \text{Var}(x) & \text{Cov}(x,y) \\ \text{Cov}(y,x) & \text{Var}(y) \end{matrix}]$$

- **Variance of Importance (x):**

$$\text{Var}(x) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

  Substituting values:

$$\text{Var}(x) = \frac{(0.4)^2 + (0.4)^2 + (-0.6)^2 + (0.4)^2 + (-0.6)^2}{4} = \frac{0.16 + 0.16 + 0.36 + 0.16 + 0.36}{4} = 0.3$$

- **Variance of Time Required (y):**

$$\text{Var}(y) = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

  Substituting values:

$$\text{Var}(y) = \frac{(0.8)^2 + (-0.2)^2 + (-1.2)^2 + (-0.2)^2 + (0.8)^2}{4} = \frac{0.64 + 0.04 + 1.44 + 0.04 + 0.64}{4} = 0.7$$

- **Covariance (x, y):**

$$\text{Cov}(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

Substituting values:

$$\text{Cov}(x, y) = \frac{(0.4)(0.8) + (0.4)(-0.2) + (-0.6)(-1.2) + (0.4)(-0.2) + (-0.6)(0.8)}{4}$$

$$= \frac{0.32 - 0.08 + 0.72 - 0.08 - 0.48}{4} = \frac{0.4}{4} = 0.1$$

The covariance matrix becomes:

$$\text{Cov}(A') = \begin{bmatrix} 0.3 & 0.1 \\ 0.1 & 0.7 \end{bmatrix}$$

Step 4: Eigenvalue Decomposition

To identify the principal components, solve for the eigenvalues and eigenvectors.

- **Characteristic Equation:** Solve:

$$\det(\text{Cov} - \lambda I) = 0$$

Substituting:

$$\det \begin{bmatrix} 0.3 - \lambda & 0.1 \\ 0.1 & 0.7 - \lambda \end{bmatrix} = 0$$

Expanding the determinant:

$$(0.3 - \lambda)(0.7 - \lambda) - (0.1)(0.1) = 0$$

$$\lambda^2 - (0.3 + 0.7)\lambda + (0.3)(0.7) - 0.01 = 0$$

$$\lambda^2 - 1.0\lambda + 0.2 = 0$$

- **Solve for Eigenvalues (λ):** Using the quadratic formula:

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

  Substituting:

$$\lambda = \frac{-(-1.0) \pm \sqrt{(1.0)^2 - 4(1)(0.2)}}{2(1)}$$

$$\lambda = \frac{1.0 \pm \sqrt{1.0 - 0.8}}{2} = \frac{1.0 \pm \sqrt{0.2}}{2}$$

$$\lambda_1 = 0.85, \quad \lambda_2 = 0.15$$

- **Eigenvectors:** Substitute each λ back into:

$$(\text{Cov} - \lambda I)v = 0$$

  Solve for the eigenvectors corresponding to $\lambda_1 = 0.85$ and $\lambda_2 = 0.15$.

Step 5: Result - Reduced Matrix

Using the eigenvector corresponding to the largest eigenvalue (0.85), project the data onto the principal component:

$$A_{\text{PCA}} = \begin{bmatrix} 3 & 6 \\ 3 & 5 \\ 2 & 4 \\ 3 & 5 \\ 2 & 6 \end{bmatrix} \rightarrow \text{Reduced to 1 Component}$$

This highlights that the most critical dimension for study topics is the combined effect of *Importance* and *Time Required*.

Why This is Helpful for Study Preparation?

1. **Reduces Complexity:** Focus on the most impactful attributes, saving time and effort.
2. **Prioritizes Key Areas:** Ensures study time is allocated to the most important topics.
3. **Efficient Decision-Making:** Simplifies the process of deciding which topics to focus on.