Principal Component Analysis

*"Center the data, find covariance neat,
Eigen the matrix, sort the elite.
Choose the top ones, projection complete —
PCA makes dimensions discrete."*

📐 PCA Formula Summary

Covariance matrix:

$$\Sigma = \frac{1}{n-1} X^\top X$$

Eigen decomposition:

$$\Sigma v = \lambda v$$

Projection onto principal components:

$$X_{\text{new}} = X \cdot W_k$$

Where W_k = matrix of top k eigenvectors.

🔢 The PCA Pipeline (Step-by-Step)

This is the official method used in ML and Data Science:

1️⃣ **Standardize** the data
2️⃣ **Compute covariance matrix** Σ
3️⃣ **Find eigenvalues & eigenvectors** of Σ
4️⃣ **Sort eigenvectors by eigenvalues** (largest first)
5️⃣ **Select top k components**
6️⃣ **Transform the data** into the new reduced space

$$X = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 2 & 2 \end{bmatrix}$$

1️⃣ Step 1: Compute the Mean of Each Feature

$$\text{mean}_{x_1} = \frac{2+0+2}{3} = \frac{4}{3}$$

$$\text{mean}_{x_2} = \frac{0+2+2}{3} = \frac{4}{3}$$

Mean vector:

$$\mu = \begin{bmatrix} 4/3 \\ 4/3 \end{bmatrix}$$

2️⃣ Step 2: Subtract Mean (Center the Data)

Compute $X_{\text{centered}} = X - \mu$:

1. $(2,0) - (4/3, 4/3) = \left(2 - \frac{4}{3},\ 0 - \frac{4}{3}\right) = \left(\frac{2}{3},\ -\frac{4}{3}\right)$
2. $(0,2) - (4/3, 4/3) = \left(-\frac{4}{3},\ \frac{2}{3}\right)$
3. $(2,2) - (4/3, 4/3) = \left(\frac{2}{3},\ \frac{2}{3}\right)$

So

$$X_c = \begin{bmatrix} \frac{2}{3} & -\frac{4}{3} \\ -\frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

3️⃣ Step 3: Covariance Matrix

Formula (for n = 3 samples):

$$\Sigma = \frac{1}{n-1} X_c^\top X_c = \frac{1}{2} X_c^\top X_c$$

First compute $X_c^\top X_c$.

$X_c^\top$ is:

$$X_c^\top = \begin{bmatrix} \frac{2}{3} & -\frac{4}{3} & \frac{2}{3} \\ -\frac{4}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix}$$

Now multiply:

- $(1,1)$:
  $\left(\frac{2}{3}\right)^2 + \left(-\frac{4}{3}\right)^2 + \left(\frac{2}{3}\right)^2 = \frac{4}{9} + \frac{16}{9} + \frac{4}{9} = \frac{24}{9} = \frac{8}{3}$
- $(2,2)$: same numbers permuted → also $\frac{8}{3}$
- $(1,2)$:
  $\frac{2}{3} \cdot \left(-\frac{4}{3}\right) + \left(-\frac{4}{3}\right) \cdot \frac{2}{3} + \frac{2}{3} \cdot \frac{2}{3} = -\frac{8}{9} - \frac{8}{9} + \frac{4}{9} = -\frac{12}{9} = -\frac{4}{3}$

Symmetry ⇒ $(2,1) = -\frac{4}{3}$

So:

$$X_c^\top X_c = \begin{bmatrix} \frac{8}{3} & -\frac{4}{3} \\ -\frac{4}{3} & \frac{8}{3} \end{bmatrix}$$

Now divide by 2:

$$\Sigma = \frac{1}{2} \begin{bmatrix} \frac{8}{3} & -\frac{4}{3} \\ -\frac{4}{3} & \frac{8}{3} \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix}$$

4️⃣ Step 4: Eigenvalues and Eigenvectors of Σ

We solve:

$$\det(\Sigma - \lambda I) = 0$$

$$\Sigma - \lambda I = \begin{bmatrix} \frac{4}{3} - \lambda & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} - \lambda \end{bmatrix}$$

Determinant:

$$\left(\frac{4}{3} - \lambda\right)^2 - \left(-\frac{2}{3}\right)^2 = 0$$

$$\left(\frac{4}{3} - \lambda\right)^2 - \frac{4}{9} = 0$$

Let $a = \frac{4}{3} - \lambda$:

$$a^2 = \frac{4}{9} \Rightarrow a = \pm\frac{2}{3}$$

So:

1. $\frac{4}{3} - \lambda = \frac{2}{3} \Rightarrow \lambda = \frac{2}{3}$
2. $\frac{4}{3} - \lambda = -\frac{2}{3} \Rightarrow \lambda = 2$

So eigenvalues:

$$\lambda_1 = 2, \quad \lambda_2 = \frac{2}{3}$$

The **largest eigenvalue** is 2 → this is **PC1**.

Eigenvector for λ = 2 (the 1st principal component)

Solve $(\Sigma - 2I)v = 0$:

$$\Sigma - 2I = \begin{bmatrix} \frac{4}{3} - 2 & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} - 2 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & -\frac{2}{3} \end{bmatrix}$$

Equation:

$$-\frac{2}{3}x - \frac{2}{3}y = 0 \Rightarrow x + y = 0 \Rightarrow y = -x$$

So an eigenvector is:

$$v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Normalize (optional for PCA):

Length $= \sqrt{1^2 + (-1)^2} = \sqrt{2}$

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

This is **PC1 direction**.

5️⃣ Step 5: Project Data onto PC1 (Dimensionality Reduction to 1D)

We use centered data X_c and project:

$$z = X_c \cdot u_1$$

Recall

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

Take each row of X_c:

1. $\left(\frac{2}{3}, -\frac{4}{3}\right) \cdot u_1 = \frac{1}{\sqrt{2}}\left(\frac{2}{3} - \left(-\frac{4}{3}\right)\right) = \frac{1}{\sqrt{2}} \cdot \frac{6}{3} = \frac{2}{\sqrt{2}} = \sqrt{2}$
2. $\left(-\frac{4}{3}, \frac{2}{3}\right) \cdot u_1 = \frac{1}{\sqrt{2}}\left(-\frac{4}{3} - \frac{2}{3}\right) = \frac{1}{\sqrt{2}} \cdot \left(-\frac{6}{3}\right) = -\sqrt{2}$
3. $\left(\frac{2}{3}, \frac{2}{3}\right) \cdot u_1 = \frac{1}{\sqrt{2}}\left(\frac{2}{3} - \frac{2}{3}\right) = 0$

So **1D PCA coordinates** are:

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