

# Principal Component Analysis

PHAUK Sökkhey, PhD

## Eigenvector and Eigenvalues

### Matrix Transformations

$$\vec{y} = A\vec{x}$$

- In general,  $\mathbf{A}$  maps the set of vectors in  $\mathbb{R}^2$  onto another set of vectors in  $\mathbb{R}^2$ .
- The transformation is invertible if  $\det(A) \neq 0$ .

### Matrix

### Matrix Operations

### Matrix Multiplication

### Linear Transformation

### Linear Transformation

### Linear Transformation

### Eigenvectors

### Review: Find Eigenvalues

Given the matrix:  $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$

### Steps:

1. Solve the characteristic equation:

$$|A - \lambda I| = 0$$

2. Substituting  $A - \lambda I$ :

$$\begin{vmatrix} 2 - \lambda & 0 \\ 0 & 1 - \lambda \end{vmatrix} = 0$$

3. Compute the determinant:

$$(2 - \lambda)(1 - \lambda) = 0$$

4. Solve for  $\lambda$ :

$$\lambda = 2, 1$$

Thus, the eigenvalues of  $A$  are  $\lambda = 2$  and  $\lambda = 1$ .

## Eigenvectors

### Eigenvalues and Eigenvectors

#### What do the eigenvalue and eigenvector indicate?

- The linear transformation results in a scaling of  $\lambda$  along the eigenvector corresponding to  $\lambda$ .
  - In the example presented:
    - \*  $\lambda = 2$  along the x-axis
    - \*  $\lambda = 1$  along the y-axis (intuitive!)
  - Stated differently, the eigenvalue indicates the percentage of transformation present along a particular direction:
    - \* **66.66% along x-axis**
    - \* **33.33% along y-axis**

## Mathematical Reviews

### Mean

- Let  $X_1, X_2, \dots, X_n$  be  $n$  observations of a random variable  $X$

- The mean is given by:

$$\mu = \bar{X} = \mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^n X_i$$

- Mean is a measure of central tendency (other measures include **mode** and **median**).
- 

## Standard Deviation

- A measure of variability (the square root of variance).
- It is defined as:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2}$$

## Covariance

- **Definition:** A measure of how two variables change together.

The covariance between  $X$  and  $Y$  is given by:

$$\Sigma_{XY} = \frac{1}{n} \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y)^T$$

Alternatively:

$$\text{Cov}(X, Y) = \Sigma_{XY} = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

## Covariance Matrix

- **Definition:** A matrix capturing the covariance between multiple variables.

The covariance matrix  $\Sigma$  is defined as:

$$\Sigma = \begin{pmatrix} \sigma_{X_1}^2 & \Sigma_{X_1 X_2} & \dots & \Sigma_{X_1 X_d} \\ \Sigma_{X_1 X_2} & \sigma_{X_2}^2 & \dots & \Sigma_{X_2 X_d} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{X_1 X_d} & \Sigma_{X_2 X_d} & \dots & \sigma_{X_d}^2 \end{pmatrix}$$


---

- Each diagonal element  $\sigma_{X_i}^2$  represents the variance of the variable  $X_i$ , while off-diagonal elements  $\Sigma_{X_i X_j}$  represent covariances between variables  $X_i$  and  $X_j$ .

## Correlation

- **Definition:** A normalized measure of how two variables change together.

The correlation between  $X$  and  $Y$  is given by:

$$\rho_{XY} = \frac{\Sigma_{XY}}{\sigma_X \sigma_Y}$$

Where:

- $\Sigma_{XY}$  is the covariance between  $X$  and  $Y$ ,
- $\sigma_X$  and  $\sigma_Y$  are the standard deviations of  $X$  and  $Y$ .

## Variance and multivariate Gaussian distributions

### Variance

- Let's say we have  $m$  observations of a variable  $x$ :

$$x^{(j)}, \quad j = 1, 2, 3, \dots, m$$

- $x^{(j)}$ :  $j$ -th observation of  $x$ .

- The **mean** of these observations is:

$$\mu = \langle x \rangle = \frac{1}{m} \sum_{i=1}^m x^{(i)}$$

- The **variance** is:

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (x^{(i)} - \mu)^2$$

**Variance**

**Covariance**

**Gaussian Distribution**

**Multivariate Gaussian Distribution**

**Multivariate Gaussian Distribution**

## **Principal Component Analysis**

### **What is Demensional Reduction?**

Principal component analysis (PCA) is a dimensionality reduction and machine learning method used to simplify a large data set into a smaller set while still maintaining significant patterns and trends.

### **The big Picture**

#### Tip

- The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set which consists of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set.
- This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

### **Steps in PCA**

#### How Do You Do a Principal Component Analysis?

1. Standardize the range of continuous initial variables
2. Compute the covariance matrix to identify correlations
3. Compute the eigenvectors and eigenvalues of the covariance matrix to identify the

principal components

4. Create a feature vector to decide which principal components to keep
5. Recast the data along the principal components axes

## PCA

### Applications of PCA in ML

#### 💡 PCA in ML

- PCA is used to visualize multidimensional data.
- It is used to reduce the number of dimensions in healthcare data.
- PCA can help resize an image.
- It can be used in finance to analyze stock data and forecast returns.
- PCA helps to find patterns in the high-dimensional datasets

### Advantages of PCA

#### 💡 Advantages of PCA

In terms of data analysis, PCA has a number of benefits, including:

- **Dimensionality Reduction:** Simplifies data by identifying key components, making it easier to analyze and visualize high-dimensional datasets.
- **Feature Extraction:** Derives meaningful new features from correlated or noisy variables.
- **Data Visualization:** Projects data onto 2D or 3D space for better pattern and cluster detection.
- **Noise Reduction:** Reduces the impact of noise or measurement errors by focusing on underlying patterns.
- **Handling Multicollinearity:** Addresses issues caused by highly correlated variables by emphasizing the most important components.

## Limitations of PCA

### Limitations of PCA

Drawbacks of PCA, including:

- **Interpretability:** Principal components may be hard to interpret in terms of the original features.
- **Information Loss:** Crucial features might be excluded, leading to potential loss of important information.
- **Outliers:** Sensitive to anomalies, which can distort the covariance matrix and affect results.
- **Scaling:** Requires properly scaled and centered data; otherwise, patterns may be misrepresented.
- **Computational Complexity:** Can be resource-intensive for large datasets, limiting scalability.

## How Does PCA Works?

### Step 1: Normalize Data

- **Standardize the data** before performing PCA to ensure:

- Each feature has a mean of 0.
- Each feature has a variance of 1.

- Formula for standardization:

$$Z = \frac{x - \mu}{\sigma}$$

- Where:

- $x$ : Original feature value.
- $\mu$ : Mean of the feature.
- $\sigma$ : Standard deviation of the feature.

## Step 2: Build the Covariance Matrix

Construct a square matrix to express the correlation between two or more features in a multi-dimensional dataset.

$$\Sigma = \begin{pmatrix} \sigma_{X_1}^2 & \Sigma_{X_1 X_2} & \dots & \Sigma_{X_1 X_d} \\ \Sigma_{X_1 X_2} & \sigma_{X_2}^2 & \dots & \Sigma_{X_2 X_d} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{X_1 X_d} & \Sigma_{X_2 X_d} & \dots & \sigma_{X_d}^2 \end{pmatrix}$$

## Step 3: Find the Eigenvectors and Eigenvalues

Calculate the eigenvectors/unit vectors and eigenvalues. Eigenvalues are scalars by which we multiply the eigenvector of the covariance matrix.

## Step 4: Sort the Eigenvectors in Highest to Lowest Order and Select the Number of Principal Components.

## Step 5: Recast the Data Along the Principal Components Axes

### PCA demo

[Demo 1\\_PCA](#)

[Demo 2\\_PCA](#)