

# Dimensionality Reduction

## Introduction

Dimensionality Reduction is a technique used to reduce the number of input features in a dataset while preserving as much important information as possible. It helps improve model performance, reduce computation, and avoid overfitting.

## Curse of Dimensionality

As the number of features increases, the amount of data required to train a model effectively grows exponentially. This problem is known as the curse of dimensionality.

More Features → Sparse Data → Poor Model Performance

## Why Dimensionality Reduction is Needed

- 1 Reduces overfitting
- 2 Improves training speed
- 3 Removes redundant features
- 4 Simplifies data visualization

## Types of Dimensionality Reduction

- 1 Feature Selection
- 2 Feature Extraction

### Feature Selection

Feature selection chooses the most relevant features and removes irrelevant ones without changing the original feature space.

### Feature Extraction

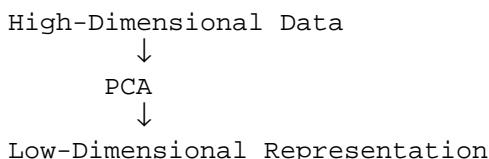
Feature extraction transforms the original features into a new lower-dimensional space.

## Principal Component Analysis (PCA)

PCA is the most commonly used dimensionality reduction technique. It transforms features into principal components that capture maximum variance.

- 1 Reduces correlated features
- 2 Unsupervised technique
- 3 Orthogonal components

Diagram:



## **Advantages**

- 1 Improves model efficiency
- 2 Reduces noise
- 3 Better visualization

## **Disadvantages**

- 1 Loss of interpretability
- 2 Possible information loss
- 3 Extra preprocessing step

## **Real-Life Example**

In image processing, dimensionality reduction is used to compress images by reducing pixel dimensions while preserving important visual information.

## **Summary**

Dimensionality reduction helps simplify data without losing essential information. It is widely used in preprocessing high-dimensional datasets.