

# Principal Component Analysis

By: Aprameyan

---

## Curse of Dimensionality

- Issues that arise when number of datapoints is small relative to the intrinsic dimension of the data.
- When dimensionality increases, the volume of the space increases so fast that the data becomes sparse.
- To obtain a good result, the amount of data increases exponentially.
- As number of features increases, data required to train models accurately grows rapidly.

## Principal Component Analysis

### Introduction

- Dimensionality Reduction Technique
- Applications:
  - Exploratory Data Analysis
  - Visualization
  - Data Preprocessing
- Dimensionality reduction:
  - Process of transforming data from a high-dimensional space to a low-dimensional space
  - Low-dimensional representation. Retains some meaningful properties of the original data, close to its intrinsic dimension
- Principal Component Analysis:
  - Technique for dimensionality reduction
  - Identifies a set of orthogonal axes, called principal components

- Captures the maximum variance in the data

## Principal components:

- In a collection of points in a real coordinate space
  - Sequence of  $p$  unit vectors
  - $i^{\text{th}}$  vector is the direction of a line that best fits the data while being perpendicular to the vectors 0 to  $i - 1$
- Best fitting line is defined as the line that minimizes the "Average Squared Perpendicular Distance" from the points to the line

## Recursive Feature Elimination