Dimension Reduction and PCA

### As the number of features or dimensions in a dataset increases, the amount of data required to obtain a statistically significant result increases exponentially. This can lead to issues such as overfitting, increased computation time, and reduced accuracy of machine learning models.

#### Dimensionality reduction is a type of feature extraction technique that aims to reduce the number of input features while retaining as much of the original information as possible.

# Principal Component Analysis (PCA)

* A statistical procedure that uses an orthogonal transformation that converts a set of correlated variables to a set of uncorrelated variables.
* PCA is an unsupervised learning algorithm technique used to examine the interrelations among a set of variables.

# Implementation Steps of PCA

**Step 1:** Standardization

First, we need to standardize our dataset to ensure that each variable has a mean of 0 and a standard deviation of 1.

**Step 2:** Covariance Matrix Computation

Covariance measures the strength of joint variability between two or more variables, indicating how much they change in relation to each other.

The value of covariance can be positive, negative, or zeros.

**Step 3:** Compute Eigenvalues and Eigenvectors of Covariance Matrix to Identify Principal Components