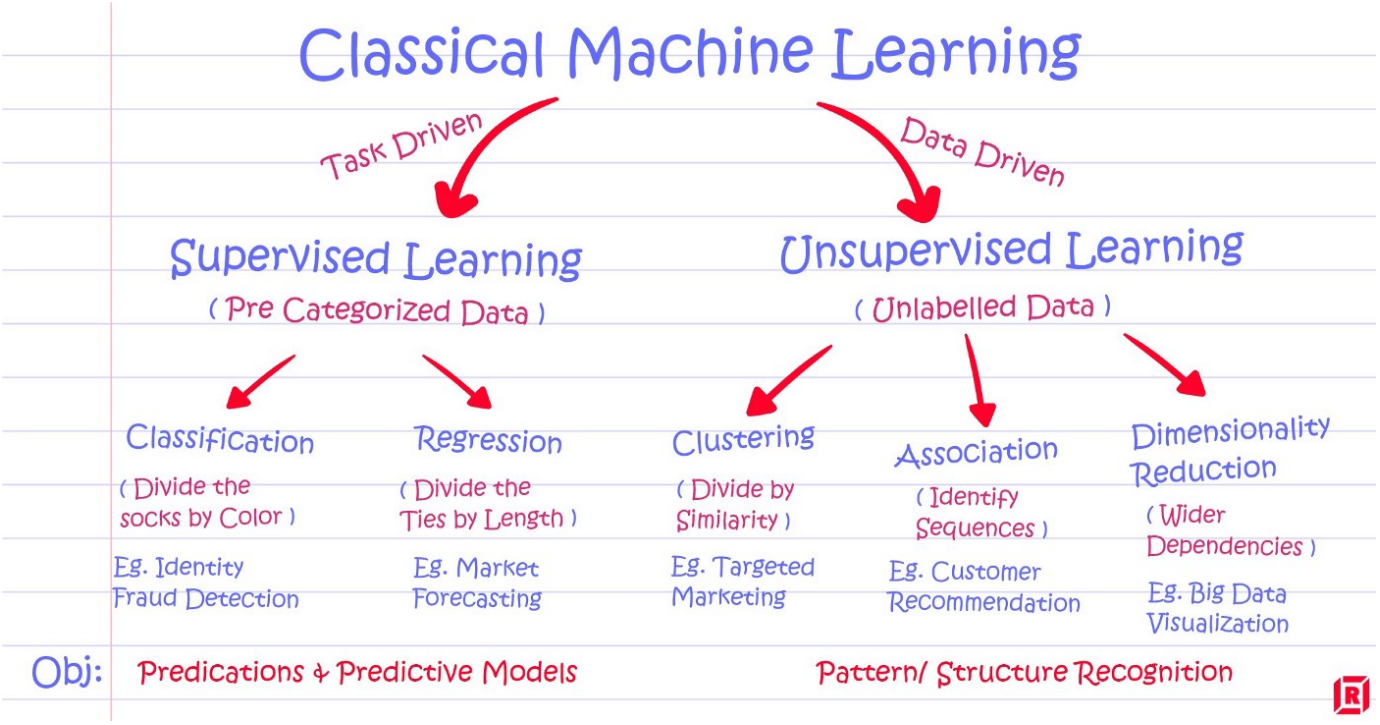
Introduction to Unsupervised Learning



Unsupervised learning is a machine learning technique in which **models are not supervised using training dataset**.

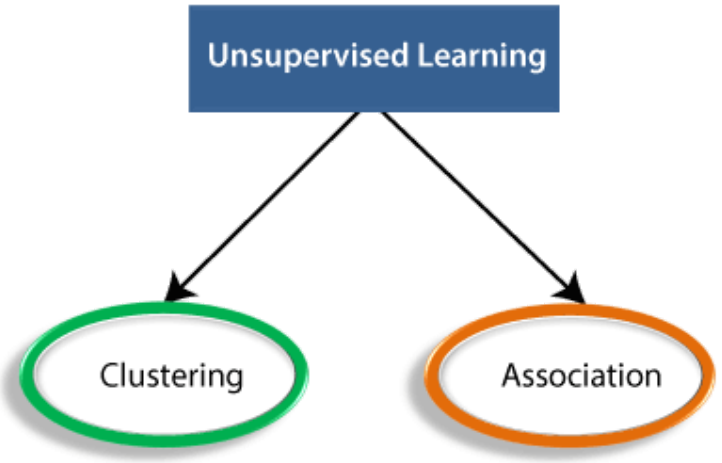
Instead, **models itself find the hidden patterns and insights from the given data**.

It can be compared to learning which takes place in the human brain while learning new things. It can be defined as:

Unsupervised learning is a **type of machine learning in which models are trained using unlabeled dataset** and are **allowed to act on that data without any supervision.**

Unsupervised learning cannot be **directly applied to a regression or classification problem because unlike supervised learning**, we have the input data but no corresponding output data.

The **goal of unsupervised learning is to find the underlying structure of dataset**, group that data according to similarities, and represent that dataset in a compressed format.



Clustering

— K-Means

— DBSCAN

— Hierarchical Cluster Analysis (HCA)

Clustering:

The **goal is to group similar instances together** into clusters.

This is a great tool for data analysis, customer segmentation, recommender systems, search engines, image segmentation, semi-supervised learning, dimensionality reduction, and more.

Anomaly detection and novelty detection

— One-class SVM

— Isolation Forest

Anomaly detection:

The objective is to learn what “normal” data looks like, and use this to detect abnormal instances, such as defective items on a production line or a new trend in a time series

Visualization and dimensionality reduction

— Principal Component Analysis (PCA)

— Kernel PCA

— Locally-Linear Embedding (LLE)

— t-distributed Stochastic Neighbor Embedding (t-SNE)

Association rule learning

— **Apriori**

— **Eclat**