

MACHINE LEARNING

Machine learning (ML) is a branch of artificial intelligence (AI) and computer science that focuses on the using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy.

Types of Machine Learning

There are several types of machine learning, each with special characteristics and applications. Some of the main types of machine learning algorithms are as follows:

① Supervised Machine Learning - Supervised learning is defined as when a model gets trained on a 'Labelled Dataset'. Labelled datasets have both input and output parameters. In supervised learning algorithms learn to map points between inputs and correct outputs. It has both training and validation datasets labelled.

Example: If you feed the datasets of dogs and cats labelled images to the algorithm, the machine will learn to classify between a dog or a cat from these labeled images. When we input new dog or cat from these labeled images, when we input new dog or cat images that it has never seen before, it will use the learned algorithms and predict whether it is a dog or a cat. This is how supervised learning works, and this is particularly on image classification.

There are two main categories of supervised learning that are mentioned below:

① Classification · ② Regression

Classification: Classification deals with predicting categorical target variables, which represent discrete classes or labels. For instance, classifying emails as spam or not spam, or predicting whether a patient has a high risk of heart disease. Classification algorithms learn to map the input features to one of the predefined classes.

Here are some classification algorithms:

- ↳ Logistic Regression
- ↳ Support Vector Machine
- ↳ Random Forest
- ↳ Decision Tree
- ↳ K-Nearest Neighbors (KNN)
- ↳ Naïve Bayes

Regression: On the other hand, deals with predicting continuous target variables, which represent numerical values. For example, predicting the price of a house based on its size, location, and amenities, or forecasting the sales of a product. Regression algorithms learn to map the input features to a continuous numerical value.

Here are some regression algorithms:

- ↳ Linear Regression
- ↳ Polynomial Regression
- ↳ Ridge Regression
- ↳ Lasso Regression
- ↳ Decision tree
- ↳ Random Forest

Example of (Numerical) Regression

Age	Experience	Skill	Communication	Salary
24	2	6	7	18000

↓
dependent variable / target variable

Example of Classification (Numerical, Categorical)

course	x variable		placement y variable
	python score	project score	
ug	7	8	Yes (1)
pn	2	6	No (0)

Advantages of Supervised Machine Learning

- ↳ Supervised Learning models can have high accuracy as they are trained on labelled data.
- ↳ The process of decision-making in supervised learning models is often interpretable.
- ↳ It can often be used in pre-trained models which saves time and resources when developing new models from scratch.

Disadvantages of Supervised Machine Learning

- ↳ It has limitations in knowing patterns and may struggle with unseen or unexpected patterns that are not present in the training data.
- ↳ It can be time-consuming and costly as it relies on labeled data only.
- ↳ It may lead to poor generalizations based on new data.

Applications of supervised Learning

Supervised learning is used in a wide variety of applications, including:

- Image classification: Identify objects, faces, and other features in images.
- Natural language processing (NLP): Extract information from text, such as sentiment, entities, and relationships.
- Speech recognition: Convert spoken language into text.

Fraud detection; Email-spam detection, Crimning.

② Unsupervised Machine Learning: Unsupervised learning is a type of machine learning technique in which an algorithm discovers patterns and relationships using unlabeled data. Unlike supervised learning, unsupervised learning doesn't involve providing the algorithm with labeled target outputs. The primary goal of Unsupervised learning is often to discover hidden patterns, similarities or clusters within the data, which can then be used for various purposes such as data exploration, visualization, dimensionality reduction, and more.

Example Consider that you have a dataset that contains information about the purchases you made from the shop. Through clustering, the algorithm can group the same purchasing behavior among you and other customers, which reveals potential customers without predefined labels. This type of information can help businesses get target customers as well as identify outliers.

There are two main categories of unsupervised learning that are mentioned below:

Clustering

Clustering is the process of grouping data points into clusters based on their similarity. This technique is useful for identifying patterns and relationships in the data without the need for labeled examples.

Here are some clustering algorithms:

- K-Means Clustering algorithm
- Mean-shift algorithm
- DBSCAN Algorithm

Principal Component Analysis

Independent Component Analysis.

Association Association rule learning is a technique for discovering relationships between items in a dataset. It identifies rules that indicate the presence of one item implies the presence of another item with a specific probability.

Here are some association rule learning algorithms;

- ⊙ Apriori Algorithm
- ⊙ Eclat
- ⊙ FP-growth Algorithm

Advantages of Unsupervised Machine Learning -

- ⊙ It helps to discover hidden patterns and various relationships between the data.
- ⊙ Used for tasks such as customer segmentation, anomaly detection, and data exploration.
- ⊙ It does not require labeled data and reduces the effort of data labelling.

Disadvantages of Unsupervised Machine Learning

- ⊙ Without using labels, it may be difficult to predict the quality of the model's output.
- ⊙ Cluster Interpretability may not be clear and may not have meaningful interpretations.
- ⊙ It has techniques such as autoencoders and dimensionality reduction that can be used to extract meaningful features from raw data.

Applications of Unsupervised Learning

Here are some common applications of unsupervised learning:

Clustering - Group similar data points into clusters.

Anomaly detection: Identify outliers or anomalies in data.

Dimensionality reduction: Reduce the dimensionality of data while preserving its essential information.

Recommendation system: Suggest products, movies, or content to users based on their historical behavior or preferences.

③ Semi-Supervised Learning: It is a machine learning algorithm that works between the supervised and unsupervised learning so it uses both labelled and unlabelled data.

It's particularly useful when obtaining labeled data is costly, time-consuming, or resource intensive. This approach is useful when the dataset is expensive and time-consuming. Semi-supervised learning is chosen when labeled data requires skills and relevant resources in order to train or learn from it.

Example - Consider that we are building a language translation model, having labeled translations for every sentence pair can be resource intensive. It allows the models to learn from labeled and unlabeled sentence pairs, making them more accurate. This technique has led to significant improvements in the quality of machine translation services.