

Supervised and Unsupervised Learning in Machine Learning

Machine Learning is the science of making computers learn and act like humans by feeding data and information without being explicitly programmed.

Machine learning algorithms are trained with training data. When new **data** comes in, they can make predictions and decisions accurately based on past data.

For example, whenever you ask Siri to do something, a powerful speech recognition converts the audio into its corresponding textual form. This is sent to the Apple servers for further processing where language processing algorithms are run to understand the user's intent. Then finally, Siri tells you the answer.

There are two types of **machine learning**:

1. Supervised Learning
2. Unsupervised Learning

Supervised Learning

In Supervised Learning, the machine learns under supervision. It contains a model that is able to predict with the help of a labeled dataset. A labeled dataset is one where you already know the target answer.

In this case, we have images that are labeled a spoon or a knife. This known data is fed to the machine, which analyzes and learns the association of these images based on its features such as shape, size, sharpness, etc. Now when a new image is fed to the machine without any label, the machine is able to predict accurately that it is a spoon with the help of the past data.

Supervised learning can be further divided into two types:

1. Classification
2. Regres

1. Classification - Supervised Learning

Classification is used when the output variable is categorical i.e. with 2 or more classes. For example, yes or no, male or female, true or false, etc.

In order to predict whether a mail is spam or not, we need to first teach the machine what a spam mail is. This is done based on a lot of spam filters - reviewing the content of the mail, reviewing the mail header, and then searching if it contains any false information. Certain keywords and blacklist filters that blackmails are used from already blacklisted spammers.

All of these features are used to score the mail and give it a spam score. The lower the total spam score of the email, the more likely that it is not a scam.

Based on the content, label, and the spam score of the new incoming mail, the algorithm decides whether it should land in the inbox or spam folder.

2. Regression - Supervised Learning

Regression is used when the output variable is a real or continuous value. In this case, there is a relationship between two or more variables i.e., a change in one variable is associated with a change in the other variable. For example, salary based on work experience or weight based on height, etc.

Let's consider two variables - humidity and temperature. Here, 'temperature' is the independent variable and 'humidity' is the dependent variable. If the temperature increases, then the humidity decreases.

These two variables are fed to the model and the machine learns the relationship between them. After the machine is trained, it can easily predict the humidity based on the given temperature.

Real-Life Applications of Supervised Learning

- Risk Assessment

Supervised learning is used to assess the risk in financial services or insurance domains in order to minimize the risk portfolio of the companies.

- Image Classification

Image classification is one of the key use cases of demonstrating supervised machine learning. For example, Facebook can recognize your friend in a picture from an album of tagged photos.

- Fraud Detection

To identify whether the transactions made by the user are authentic or not.

- Visual Recognition

The ability of a machine learning model to identify objects, places, people, actions, and images.

Unsupervised Learning

In Unsupervised Learning, the machine uses unlabeled data and learns on itself without any supervision. The machine tries to find a pattern in the unlabeled data and gives a response. Let's take a similar example as before, but this time we do not tell the machine whether it's a spoon or a knife. The machine identifies patterns from the given set and groups them based on their patterns, similarities, etc.

Unsupervised learning can be further grouped into types:

1. Clustering
2. Association

1. Clustering - Unsupervised Learning

Clustering is the method of dividing the objects into clusters that are similar between them and are dissimilar to the objects belonging to another cluster. For example, finding out which customers made similar product purchases.

Suppose a telecom company wants to reduce its customer churn rate by providing personalized call and data plans. The behavior of the customers is studied and the model segments the customers with similar traits. Several strategies are adopted to minimize churn rate and maximize profit through suitable promotions and campaigns.

On the right side of the image, you can see a graph where customers are grouped. Group A customers use more data and also have high call durations. Group B customers are heavy Internet users, while Group C customers have high call duration. So, Group B will be given more data benefit plans, while Group C will be given cheaper called call rate plans and group A will be given the benefit of both.

2. Association - Unsupervised Learning

Association is a rule-based machine learning to discover the probability of the co-occurrence of items in a collection. For example, finding out which products were purchased together.

Let's say that a customer goes to a supermarket and buys bread, milk, fruits, and wheat. Another customer comes and buys bread, milk, rice, and butter. Now, when another customer comes, it is highly likely that if he buys bread, he will buy milk too. Hence, a relationship is established based on customer behavior and recommendations are made **Real-Life Applications of Unsupervised Learning**

- Market Basket Analysis

It is a machine learning model based on the algorithm that if you buy a certain group of items, you are less or more likely to buy another group of items.

- Semantic Clustering

Semantically similar words share a similar context. People post their queries on websites in their own ways. Semantic clustering groups all these responses with the same meaning in a cluster to ensure that the customer finds the information they want quickly and easily. It plays an important role in information retrieval, good browsing experience, and comprehension.

- Delivery Store Optimization

Machine learning models are used to predict the demand and keep up with supply. They are also used to open stores where the demand is higher and optimizing routes for more efficient deliveries according to past data and behavior.

- Identifying Accident Prone Areas

Unsupervised machine learning models can be used to identify accident-prone areas and introduce safety measures based on the intensity of those accidents.

Difference Between Supervised and Unsupervised Learning

Supervised Learning	Unsupervised Learning
It uses known and labeled data as input	It uses unlabeled data as input
It has a feedback mechanism	It has no feedback mechanism

The most commonly used supervised learning algorithms are:

- Decision tree
- Logistic regression
- Support vector machine

The most commonly used unsupervised learning algorithms are:

- K-means clustering
- Hierarchical clustering
- Apriori algorithm