Supervised VS Unsupervised Machine Learning

Supervised learning:

Supervised learning is the learning of the model where with input variable (say, x) and an output variable (say, Y) and an algorithm to map the input to the output.

That is, Y = f(X)

Why supervised learning?

The basic aim is to approximate the mapping function(mentioned above) so well that when there is a new input data (x) then the corresponding output variable can be predicted.

It is called supervised learning because the process of an learning(from the training dataset) can be thought of as a teacher who is supervising the entire learning process. Thus, the "learning algorithm" iteratively makes predictions on the training data and is corrected by the "teacher", and the learning stops when the algorithm achieves an acceptable level of performance(or the desired accuracy).

Example of Supervised Learning

Suppose there is a basket which is filled with some fresh fruits, the task is to arrange the same type of fruits at one place.

Also, suppose that the fruits are apple, banana, cherry, grape.

Suppose one already knows from their previous work (or experience) that, the shape of each and every fruit present in the basket so, it is easy for them to arrange the same type of fruits in one place.

Here, the previous work is called as training data in Data Mining terminology. So, it learns the things from the training data. This is because it has a response variable which says y that if some fruit has so and so features then it is grape, and similarly for each and every fruit.
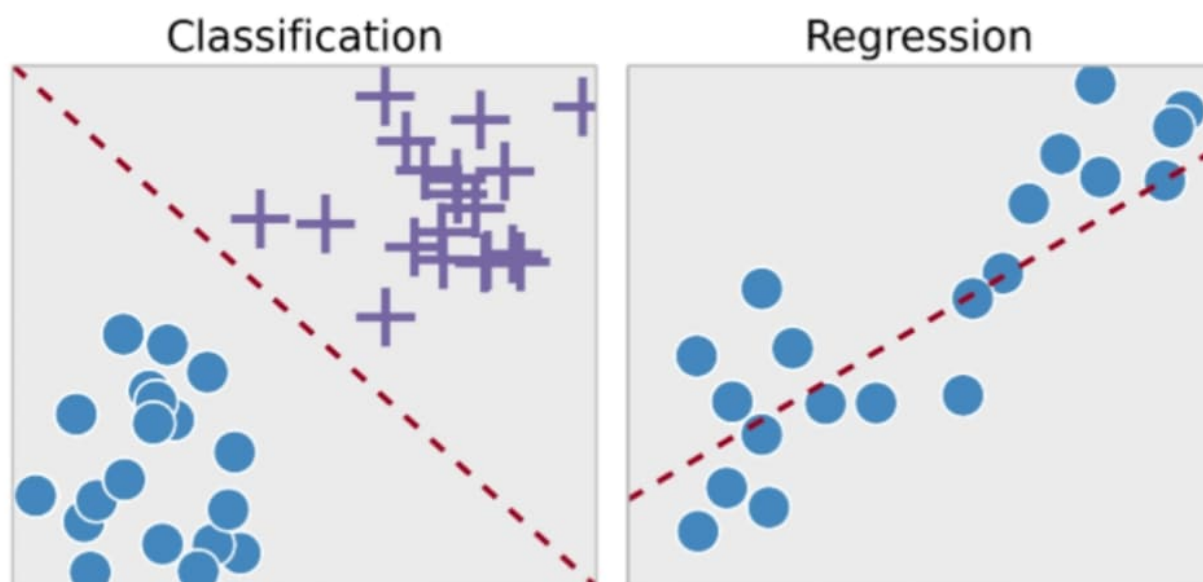
This type of information is deciphered from the data that is used to train the model.

This type of learning is called Supervised Learning.

Such problems are listed under classical Classification Tasks.

Types of Supervised Machine Learning Techniques

Regression:

Regression technique predicts a single output value using training data.
Example: You can use regression to predict the house price from training data. The input variables will be locality, size of a house, etc.

Classification:
Classification means to group the output inside a class. If the algorithm tries to label input into two distinct classes, it is called binary classification. Selecting between more than two classes is referred to as multiclass classification.

Example: Determining whether or not someone will be a defaulter of the loan.

Strengths: Outputs always have a probabilistic interpretation, and the algorithm can be regularized to avoid overfitting.

Weaknesses: Logistic regression may underperform when there are multiple or non-linear decision boundaries. This method is not flexible, so it does not capture more complex relationships.

Unsupervised Learning:
Unsupervised learning is where only the input data (say, X) is present and no corresponding output variable is there.

Why Unsupervised Learning?
The main aim of Unsupervised learning is to model the distribution in the data in order to learn more about the data.
It is called so, because there is no correct answer and there is no such teacher(unlike supervised learning). Algorithms are left to their own devises to discover and present the interesting structure in the data.

Example of Unsupervised Learning

Again, Suppose there is a basket and it is filled with some fresh fruits. The task is to arrange the same type of fruits at one place.
This time there is no information about those fruits beforehand, its the first time that the fruits are being seen or discovered
So how to group similar fruits without any prior knowledge about those.
First, any physical characteristic of a particular fruit is selected. Suppose color.

Then the fruits are arranged on the basis of the color. The groups will be something as shown below:
RED COLOR GROUP: apples & cherry fruits.
GREEN COLOR GROUP: bananas & grapes.
So now, take another physical character say, size, so now the groups will be something like this.

RED COLOR AND BIG SIZE: apple.
RED COLOR AND SMALL SIZE: cherry fruits.
GREEN COLOR AND BIG SIZE: bananas.
GREEN COLOR AND SMALL SIZE: grapes.

Here, there is no need to know or learn anything beforehand. That means, no train data and no response variable. This type of learning is known as Unsupervised Learning.

Types of Unsupervised Machine Learning Techniques

Unsupervised learning problems further grouped into clustering and association problems.

Clustering

Clustering is an important concept when it comes to unsupervised learning. It mainly

| sample | Cluster/group |

deals with finding a structure or pattern in a collection of uncategorized data. Clustering algorithms will process your data and find natural clusters(groups) if they exist in the data. You can also modify how many clusters your algorithms should identify. It allows you to adjust the granularity of these groups.

Association
Association rules allow you to establish associations amongst data objects inside large databases. This unsupervised technique is about discovering exciting relationships between variables in large databases. For example, people that buy a new home most likely to buy new furniture.

Other Examples:
*   A subgroup of cancer patients grouped by their gene expression measurements
*   Groups of shopper based on their browsing and purchasing histories
*   Movie group by the rating given by movies viewers

Difference between Supervised & Unsupervised Machine Learning

Parameters	Supervised machine learning technique	Unsupervised machine learning technique
Process	In a supervised learning model, input and output variables will be given.	In unsupervised learning model, only input data will be given
Input Data	Algorithms are trained using labeled data.	Algorithms are used against data which is not labeled
Algorithms Used	Support vector machine, Neural network, Linear and logistics regression, random forest, and Classification trees.	Unsupervised algorithms can be divided into different categories: like Cluster algorithms, K-means, Hierarchical clustering, etc.
Computational Complexity	Supervised learning is a simpler method.	Unsupervised learning is computationally complex
Use of Data	Supervised learning model uses training data to learn a link between the input and the outputs.	Unsupervised learning does not use output data.
Accuracy of Results	Highly accurate and trustworthy method.	Less accurate and trustworthy method.
Real Time Learning	Learning method takes place offline.	Learning method takes place in real time.
Number of Classes	Number of classes is known.	Number of classes is not known.
Main Drawback	Classifying big data can be a real challenge in Supervised Learning.	You cannot get precise information regarding data sorting, and the output as data used in unsupervised learning is labeled and not known.

Summary

• In Supervised learning, you train the machine using data which is well "labeled."
• Unsupervised learning is a machine learning technique, where you do not need to supervise the model.
• Supervised learning allows you to collect data or produce a data output from the previous experience.
• Unsupervised machine learning helps you to finds all kind of unknown patterns in data.
• For example, you will able to determine the time taken to reach back come base on weather condition, Times of the day and holiday.
• For example, Baby can identify other dogs based on past supervised learning.
• Regression and Classification are two types of supervised machine learning techniques.
• Clustering and Association are two types of Unsupervised learning.
• In a supervised learning model, input and output variables leill be given while with unsupervised learning model, only input data will be given.