



UNIT I



INTRODUCTION

Machine Learning:

- Machine Learning tutorial covers basic and advanced concepts, specially designed to cater to both students and experienced working professionals.
- This machine learning tutorial helps you gain a solid introduction to the fundamentals of machine learning and explore a wide range of techniques, including supervised, unsupervised, and reinforcement learning.

What is Machine Learning?

- Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed.
- ML is one of the most exciting technologies that one would have ever come across.
- As it is evident from the name, it gives the computer that makes it more similar to humans:
- The ability to learn. Machine learning is actively being used today, perhaps in many more places than one would expect.

Features of Machine learning:

- Machine learning is data driven technology. Large amount of data generated by organizations on daily bases.
- So, by notable relationships in data, organizations makes better decisions.
- Machine can learn itself from past data and automatically improve.
- From the given dataset it detects various patterns on data.
- For the big organizations branding is important and it will become more easy to target relatable customer base.
- It is similar to data mining because it is also deals with the huge amount of data.

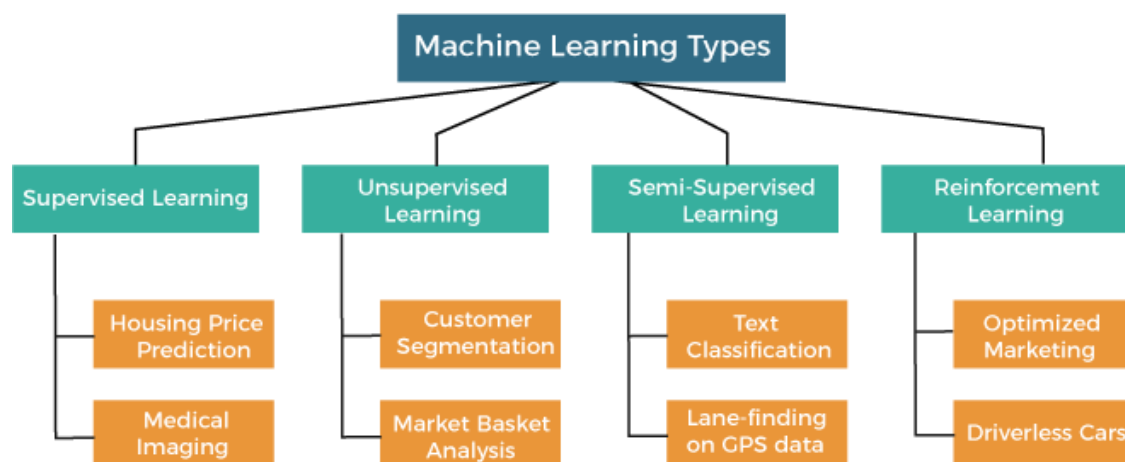


Types of Machine Learning

- Machine learning is a subset of AI, which enables the machine to automatically learn from data, improve performance from past experiences, and make predictions.
- Machine learning contains a set of algorithms that work on a huge amount of data.
- Data is fed to these algorithms to train them, and on the basis of training, they build the model & perform a specific task.
- These ML algorithms help to solve different business problems like Regression, Classification, Forecasting, Clustering, and Associations, etc.

Based on the methods and way of learning, machine learning is divided into mainly four types, which are:

1. Supervised Machine Learning
2. Unsupervised Machine Learning
3. Semi-Supervised Machine Learning
4. Reinforcement Learning





1. Supervised Machine Learning:

- As its name suggests, Supervised machine learning is based on supervision.
- It means in the supervised learning technique, we train the machines using the "labelled" dataset, and based on the training, the machine predicts the output.
- Here, the labelled data specifies that some of the inputs are already mapped to the output.
- More precisely, we can say; first, we train the machine with the input and corresponding output, and then we ask the machine to predict the output using the test dataset.
- Let's understand supervised learning with an example.
- Suppose we have an input dataset of cats and dog images. So, first, we will provide the training to the machine to understand the images, such as the shape & size of the tail of cat and dog, Shape of eyes, colour, height (dogs are taller, cats are smaller), etc.
- After completion of training, we input the picture of a cat and ask the machine to identify the object and predict the output.
- Now, the machine is well trained, so it will check all the features of the object, such as height, shape, colour, eyes, ears, tail, etc., and find that it's a cat. So, it will put it in the Cat category.
- This is the process of how the machine identifies the objects in Supervised Learning.
- The main goal of the supervised learning technique is to map the input variable(x) with the output variable(y).
- Some real-world applications of supervised learning are Risk Assessment, Fraud Detection, Spam filtering, etc.

Categories of Supervised Machine Learning:

Supervised machine learning can be classified into two types of problems, which are given below:

- Classification
- Regression



a) Classification

- Classification algorithms are used to solve the classification problems in which the output variable is categorical, such as "Yes" or No, Male or Female, Red or Blue, etc.
- The classification algorithms predict the categories present in the dataset.
- Some real-world examples of classification algorithms are Spam Detection, Email filtering, etc.

Some popular classification algorithms are given below:

- Random Forest Algorithm
- Decision Tree Algorithm
- Logistic Regression Algorithm
- Support Vector Machine Algorithm

b) Regression:

- Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables.
- These are used to predict continuous output variables, such as market trends, weather prediction, etc.

Some popular Regression algorithms are given below:

- Simple Linear Regression Algorithm
- Multivariate Regression Algorithm
- Decision Tree Algorithm
- Lasso Regression

Advantages of Supervised Learning:

- Since supervised learning work with the labelled dataset so we can have an exact idea about the classes of objects.
- These algorithms are helpful in predicting the output on the basis of prior experience.



Disadvantages of Supervised Learning:

- These algorithms are not able to solve complex tasks.
- It may predict the wrong output if the test data is different from the training data.
- It requires lots of computational time to train the algorithm.

Applications of Supervised Learning:

Some common applications of Supervised Learning are given below:

Image Segmentation:

- Supervised Learning algorithms are used in image segmentation.
- In this process, image classification is performed on different image data with pre-defined labels.

Medical Diagnosis:

- Supervised algorithms are also used in the medical field for diagnosis purposes.
- It is done by using medical images and past labelled data with labels for disease conditions.
- With such a process, the machine can identify a disease for the new patients.

Fraud Detection:

- Supervised Learning classification algorithms are used for identifying fraud transactions, fraud customers, etc.
- It is done by using historic data to identify the patterns that can lead to possible fraud.

Spam detection:

- In spam detection & filtering, classification algorithms are used.
- These algorithms classify an email as spam or not spam.
- The spam emails are sent to the spam folder.



Speech Recognition:

- Supervised learning algorithms are also used in speech recognition.
- The algorithm is trained with voice data, and various identifications can be done using the same, such as voice-activated passwords, voice commands, etc.

2. Unsupervised Machine Learning:

- Unsupervised learning is different from the Supervised learning technique; as its name suggests, there is no need for supervision.
- It means, in unsupervised machine learning, the machine is trained using the unlabelled dataset, and the machine predicts the output without any supervision.
- In unsupervised learning, the models are trained with the data that is neither classified nor labelled, and the model acts on that data without any supervision.
- The main aim of the unsupervised learning algorithm is to group or categories the unsorted dataset according to the similarities, patterns, and differences.
- Machines are instructed to find the hidden patterns from the input dataset.
- Let's take an example to understand it more precisely; suppose there is a basket of fruit images, and we input it into the machine learning model.
- The images are totally unknown to the model, and the task of the machine is to find the patterns and categories of the objects.
- So, now the machine will discover its patterns and differences, such as colour difference, shape difference, and predict the output when it is tested with the test dataset.

Categories of Unsupervised Machine Learning:

Unsupervised Learning can be further classified into two types, which are given below:

- Clustering
- Association



1) Clustering:

- The clustering technique is used when we want to find the inherent groups from the data.
- It is a way to group the objects into a cluster such that the objects with the most similarities remain in one group and have fewer or no similarities with the objects of other groups.
- An example of the clustering algorithm is grouping the customers by their purchasing behaviour.

Some of the popular clustering algorithms are given below:

- K-Means Clustering algorithm
- Mean-shift algorithm
- DBSCAN Algorithm
- Principal Component Analysis
- Independent Component Analysis

2) Association:

- Association rule learning is an unsupervised learning technique, which finds interesting relations among variables within a large dataset.
- The main aim of this learning algorithm is to find the dependency of one data item on another data item and map those variables accordingly so that it can generate maximum profit.
- This algorithm is mainly applied in Market Basket analysis, Web usage mining, continuous production, etc.
- Some popular algorithms of Association rule learning are Apriori Algorithm, Eclat, FP-growth algorithm.



Advantages and Disadvantages of Unsupervised Learning Algorithm

Advantages:

- These algorithms can be used for complicated tasks compared to the supervised ones because these algorithms work on the unlabelled dataset.
- Unsupervised algorithms are preferable for various tasks as getting the unlabelled dataset is easier as compared to the labelled dataset.

Disadvantages:

- The output of an unsupervised algorithm can be less accurate as the dataset is not labelled, and algorithms are not trained with the exact output in prior.
- Working with Unsupervised learning is more difficult as it works with the unlabelled dataset that does not map with the output.

Applications of Unsupervised Learning:

- Network Analysis: Unsupervised learning is used for identifying plagiarism and copyright in document network analysis of text data for scholarly articles.
- Recommendation Systems: Recommendation systems widely use unsupervised learning techniques for building recommendation applications for different web applications and e-commerce websites.
- Anomaly Detection: Anomaly detection is a popular application of unsupervised learning, which can identify unusual data points within the dataset. It is used to discover fraudulent transactions.
- Singular Value Decomposition: Singular Value Decomposition or SVD is used to extract particular information from the database. For example, extracting information of each user located at a particular location.



3. Semi-Supervised Learning:

- Semi-Supervised learning is a type of Machine Learning algorithm that lies between Supervised and Unsupervised machine learning.
- It represents the intermediate ground between Supervised (With Labelled training data) and Unsupervised learning (with no labelled training data) algorithms and uses the combination of labelled and unlabelled datasets during the training period.
- Although Semi-supervised learning is the middle ground between supervised and unsupervised learning and operates on the data that consists of a few labels, it mostly consists of unlabelled data.
- As labels are costly, but for corporate purposes, they may have few labels.
- It is completely different from supervised and unsupervised learning as they are based on the presence & absence of labels.
- To overcome the drawbacks of supervised learning and unsupervised learning algorithms, the concept of Semi-supervised learning is introduced.
- The main aim of semi-supervised learning is to effectively use all the available data, rather than only labelled data like in supervised learning.
- Initially, similar data is clustered along with an unsupervised learning algorithm, and further, it helps to label the unlabelled data into labelled data.
- It is because labelled data is a comparatively more expensive acquisition than unlabelled data.
- We can imagine these algorithms with an example.
- Supervised learning is where a student is under the supervision of an instructor at home and college.
- Further, if that student is self-analysing the same concept without any help from the instructor, it comes under unsupervised learning.
- Under semi-supervised learning, the student has to revise himself after analysing the same concept under the guidance of an instructor at college



Advantages and disadvantages of Semi-supervised Learning

Advantages:

- It is simple and easy to understand the algorithm.
- It is highly efficient.
- It is used to solve drawbacks of Supervised and Unsupervised Learning algorithms.

Disadvantages:

- Iterations results may not be stable.
- We cannot apply these algorithms to network-level data.
- Accuracy is low.

4. Reinforcement Learning:

- Reinforcement learning works on a feedback-based process, in which an AI agent (A software component) automatically explore its surrounding by hitting & trail, taking action, learning from experiences, and improving its performance.
- Agent gets rewarded for each good action and get punished for each bad action; hence the goal of reinforcement learning agent is to maximize the rewards.
- In reinforcement learning, there is no labelled data like supervised learning, and agents learn from their experiences only.
- The reinforcement learning process is similar to a human being; for example, a child learns various things by experiences in his day-to-day life.
- An example of reinforcement learning is to play a game, where the Game is the environment, moves of an agent at each step define states, and the goal of the agent is to get a high score. Agent receives feedback in terms of punishment and rewards.
- Due to its way of working, reinforcement learning is employed in different fields such as Game theory, Operation Research, Information theory, multi-agent systems.



- A reinforcement learning problem can be formalized using Markov Decision Process(MDP).
- In MDP, the agent constantly interacts with the environment and performs actions; at each action, the environment responds and generates a new state.

Categories of Reinforcement Learning:

Reinforcement learning is categorized mainly into two types of methods/algorithms:

Positive Reinforcement Learning: Positive reinforcement learning specifies increasing the tendency that the required behaviour would occur again by adding something. It enhances the strength of the behaviour of the agent and positively impacts it.

Negative Reinforcement Learning: Negative reinforcement learning works exactly opposite to the positive RL. It increases the tendency that the specific behaviour would occur again by avoiding the negative condition.

Real-world Use cases of Reinforcement Learning:

Video Games:

- RL algorithms are much popular in gaming applications.
- It is used to gain super-human performance.
- Some popular games that use RL algorithms are AlphaGO and AlphaGO Zero.

Resource Management:

- The "Resource Management with Deep Reinforcement Learning" paper showed that how to use RL in computer to automatically learn and schedule resources to wait for different jobs in order to minimize average job slowdown.



Robotics:

- RL is widely being used in Robotics applications. Robots are used in the industrial and manufacturing area, and these robots are made more powerful with reinforcement learning.
- There are different industries that have their vision of building intelligent robots using AI and Machine learning technology.

Text Mining:

- Text-mining, one of the great applications of NLP, is now being implemented with the help of Reinforcement Learning by Salesforce company.

Advantages and Disadvantages of Reinforcement Learning

Advantages

- It helps in solving complex real-world problems which are difficult to be solved by general techniques.
- The learning model of RL is similar to the learning of human beings; hence most accurate results can be found.
- Helps in achieving long term results.

Disadvantage

- RL algorithms are not preferred for simple problems.
- RL algorithms require huge data and computations.
- Too much reinforcement learning can lead to an overload of states which can weaken the results.

The curse of dimensionality limits reinforcement learning for real physical systems.