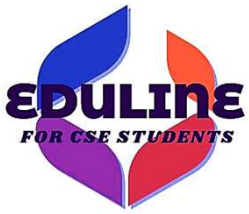
MODULE 5

MACHINE LEARNING



MACHINE LEARNING

- Machine Learning (ML) is an automated learning with little or no human intervention.
- It involves programming computers so that they learn from the available inputs.
- The main purpose of machine learning is to explore and construct algorithms that can learn from the previous data and make predictions on new input data.
- The input to a learning algorithm is training data, representing experience, and the output is any expertise, which usually takes the form of another algorithm that can perform a task.

- The input data to a machine learning system can be numerical, textual, audio, visual, or multimedia.
- The corresponding output data of the system can be a floating-point number, for instance, the velocity of a rocket, an integer representing a category or a class, for example, a pigeon or a sunflower from image recognition.

❖**Concepts of Learning**

- Learning is the process of converting experience into expertise or knowledge.
- Learning can be broadly classified into three categories, as mentioned below, based on the nature of the learning data and interaction between the learner and the environment.

➢Supervised Learning

➢Unsupervised Learning

➢Semi-supervised Learning

- Similarly, there are four categories of machine learning algorithms as shown below −

➢Supervised learning algorithm

➢Unsupervised learning algorithm

➢Semi-supervised learning algorithm

➢Reinforcement learning algorithm

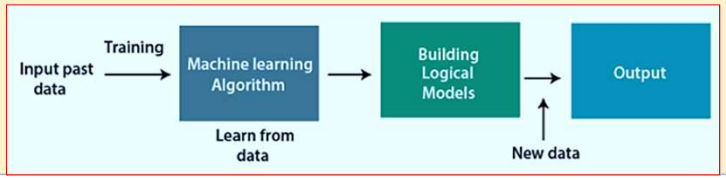
- However, the most commonly used ones are supervised and unsupervised learning.

❖What is Machine Learning

- In the real world, we are surrounded by humans who can learn everything from their experiences with their learning capability, and we have computers or machines which work on our instructions. But can a machine also learn from experiences or past data like a human does? So here comes the role of Machine Learning.

- Machine Learning is said as a subset of artificial intelligence that is mainly concerned with the development of algorithms which allow a computer to learn from the data and past experiences on their own.
- The term machine learning was first introduced by Arthur Samuel in 1959.
- We can define it in a summarized way as: Machine learning enables a machine to automatically learn from data, improve performance from experiences, and predict things without being explicitly programmed.

- With the help of sample historical data, which is known as training data, machine learning algorithms build a mathematical model that helps in making predictions or decisions without being explicitly programmed.
- Machine learning brings computer science and statistics together for creating predictive models.
- Machine learning constructs or uses the algorithms that learn from historical data. The more we will provide the information, the higher will be the performance.



❖Features of Machine Learning:

- Machine learning uses data to detect various patterns in a given dataset.
- It can learn from past data and improve automatically.
- It is a data-driven technology.
- Machine learning is much similar to data mining as it also deals with the huge amount of the data.

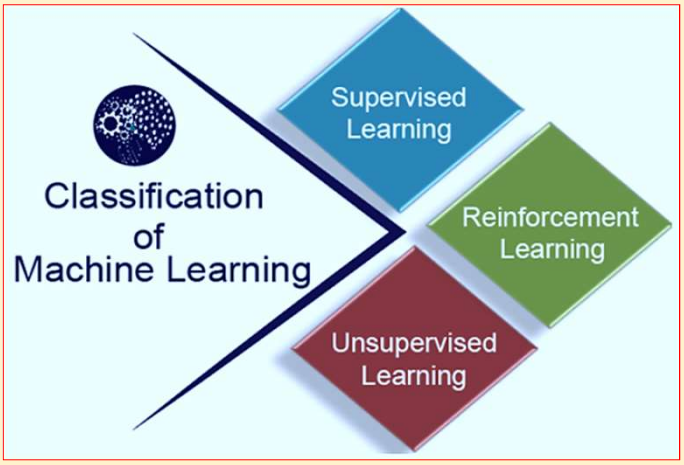
❖Classification of Machine Learning

- At a broad level, machine learning can be classified into three types:

➢Supervised learning

➢Unsupervised learning

➢Reinforcement learning



SUPERVISED LEARNING

- Supervised learning is a type of machine learning method in which we provide sample labeled data to the machine learning system in order to train it, and on that basis, it predicts the output.
- The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model by providing a sample data to check whether it is predicting the exact output or not.
- The goal of supervised learning is to map input data with the output data.

- The supervised learning is based on supervision, and it is the same as when a student learns things in the supervision of the teacher.
- The example of supervised learning is spam filtering.
- Supervised learning can be grouped further in two categories of algorithms:

➢ **Classification**

➢ **Regression**

❑ 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

❑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 Regression Algorithm
- Lasso Regression
- Random Forest

❖Advantages and Disadvantages of Supervised Learning

➢Advantages:

- 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:

- 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.

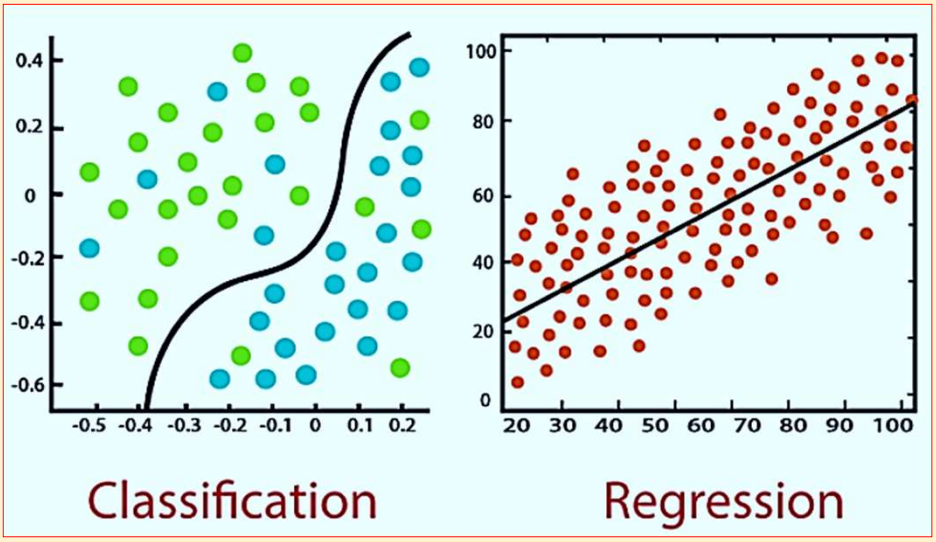
➢**Spam detection** - In spam detection & filtering, classification algorithms are used. These algorithms classify an email as spam or not spam. The spam emails are sent to the spam folder.

➢**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.

REGRESSION VS. CLASSIFICATION IN MACHINE LEARNING

- Regression and Classification algorithms are Supervised Learning algorithms. Both the algorithms are used for prediction in Machine learning and work with the labeled datasets.
- But the difference between both is how they are used for different machine learning problems.
- The main difference between Regression and Classification algorithms that Regression algorithms are used to predict the continuous values such as price, salary, age, etc.
- Classification algorithms are used to predict/Classify the discrete values such as Male or Female, True or False, Spam or Not Spam, etc.

❖Consider the below diagram:



➢Classification:

- Classification is a process of finding a function which helps in dividing the dataset into classes based on different parameters. In Classification, a computer program is trained on the training dataset and based on that training, it categorizes the data into different classes.
- The task of the classification algorithm is to find the mapping function to map the input(x) to the discrete output(y).
- Example: The best example to understand the Classification problem is Email Spam Detection. The model is trained on the basis of millions of emails on different parameters, and whenever it receives a new email, it identifies whether the email is spam or not. If the email is spam, then it is moved to the Spam folder.

❖Types of ML Classification Algorithms:

- Classification Algorithms can be further divided into the following types:
  - Logistic Regression
  - K-Nearest Neighbours
  - Support Vector Machines
  - Kernel SVM
  - Naïve Bayes
  - Decision Tree Classification
  - Random Forest Classification

➢Regression:

- Regression is a process of finding the correlations between dependent and independent variables. It helps in predicting the continuous variables such as prediction of Market Trends, prediction of House prices, etc.
- The task of the Regression algorithm is to find the mapping function to map the input variable(x) to the continuous output variable(y).
- Example: Suppose we want to do weather forecasting, so for this, we will use the Regression algorithm. In weather prediction, the model is trained on the past data, and once the training is completed, it can easily predict the weather for future days.

❖Types of Regression Algorithm:

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression

❖Difference between Regression and Classification:

Regression Algorithm	Classification Algorithm
In Regression, the output variable must be of continuous nature or real value.	In Classification, the output variable must be a discrete value.
The task of the regression algorithm is to map the input value (x) with the continuous output variable(y).	The task of the classification algorithm is to map the input value(x) with the discrete output variable(y).
Regression Algorithms are used with continuous data.	Classification Algorithms are used with discrete data.
In Regression, we try to find the best fit line, which can predict the output more accurately.	In Classification, we try to find the decision boundary, which can divide the dataset into different classes.
Regression algorithms can be used to solve the regression problems such as Weather Prediction, House price prediction, etc.	Classification Algorithms can be used to solve classification problems such as Identification of spam emails, Speech Recognition, Identification of cancer cells, etc.
The regression Algorithm can be further divided into Linear and Non-linear Regression.	The Classification algorithms can be divided into Binary Classifier and Multi-class Classifier.

Prepared By Mr. EBIN PM, Chandigarh University, Punjab EDULINE 23