

UNIT 4: LEARNING

[4 Hrs.]

4.1 Concepts of machine learning

4.2 Rote learning, learning by analogy, inductive learning, Explanation based learning, Supervised and unsupervised learning, learning by evolution (genetic algorithm)

UNIT 5: NEURAL NETWORKS AND NATURAL LANGUAGE PROCESSING [7Hrs.]

- 5.1 Introduction to artificial neural network, Mathematical model of neural network, types of neural network: feed-forward, feed-back, Gate realization using neural network, Learning in neural networks: Back propagation algorithm, Hopfield network, Boltzmann machines
- 5.2 Concepts of natural language understanding and natural language generation, Steps in natural language processing, Syntax analysis, Semantic analysis, Pragmatic analysis

Yuba Raj Devkota

Kathmandu Shiksha Campus

Concepts of Learning

- Learning is the process of converting experience into expertise or knowledge.
- **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.

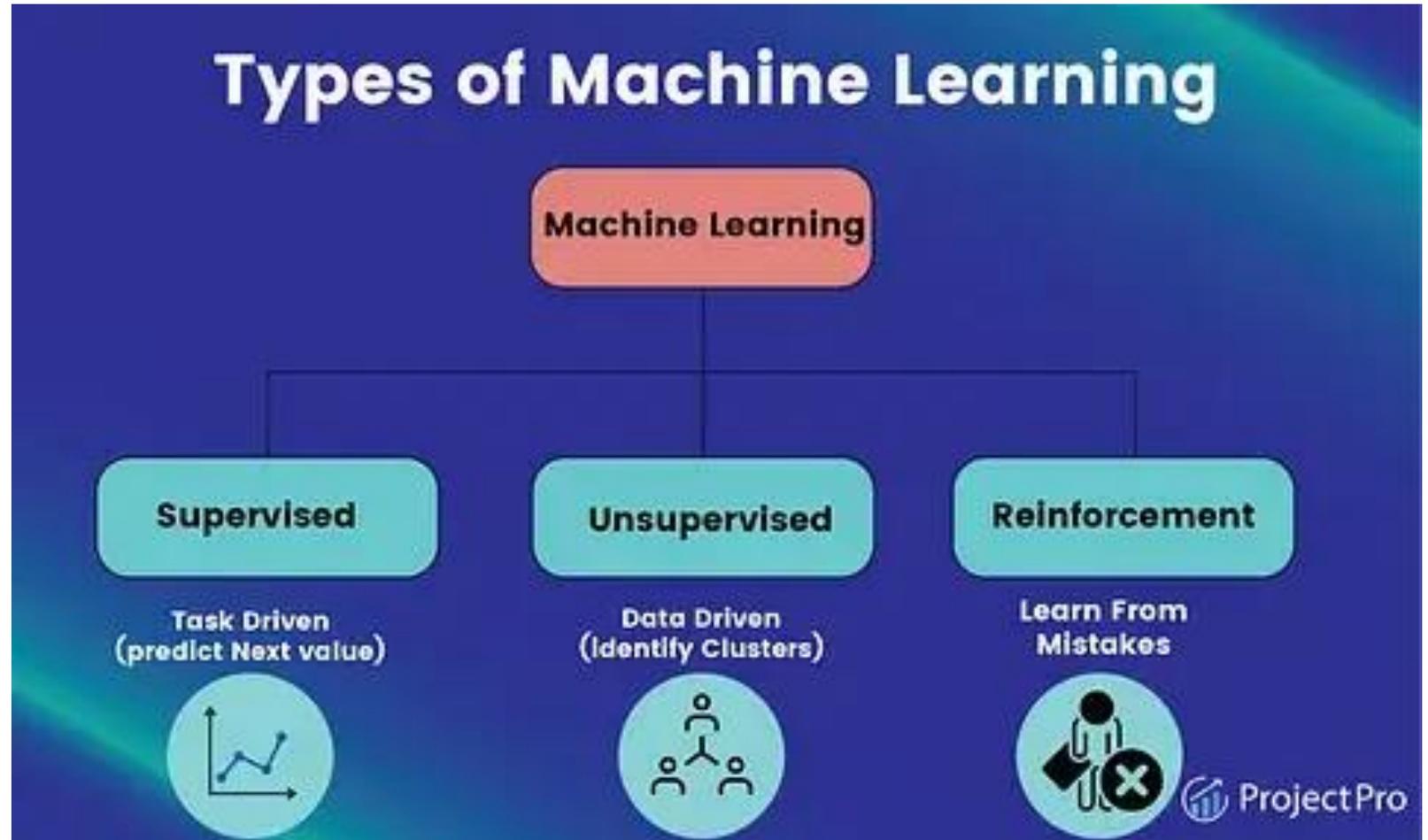
Purpose of Machine Learning

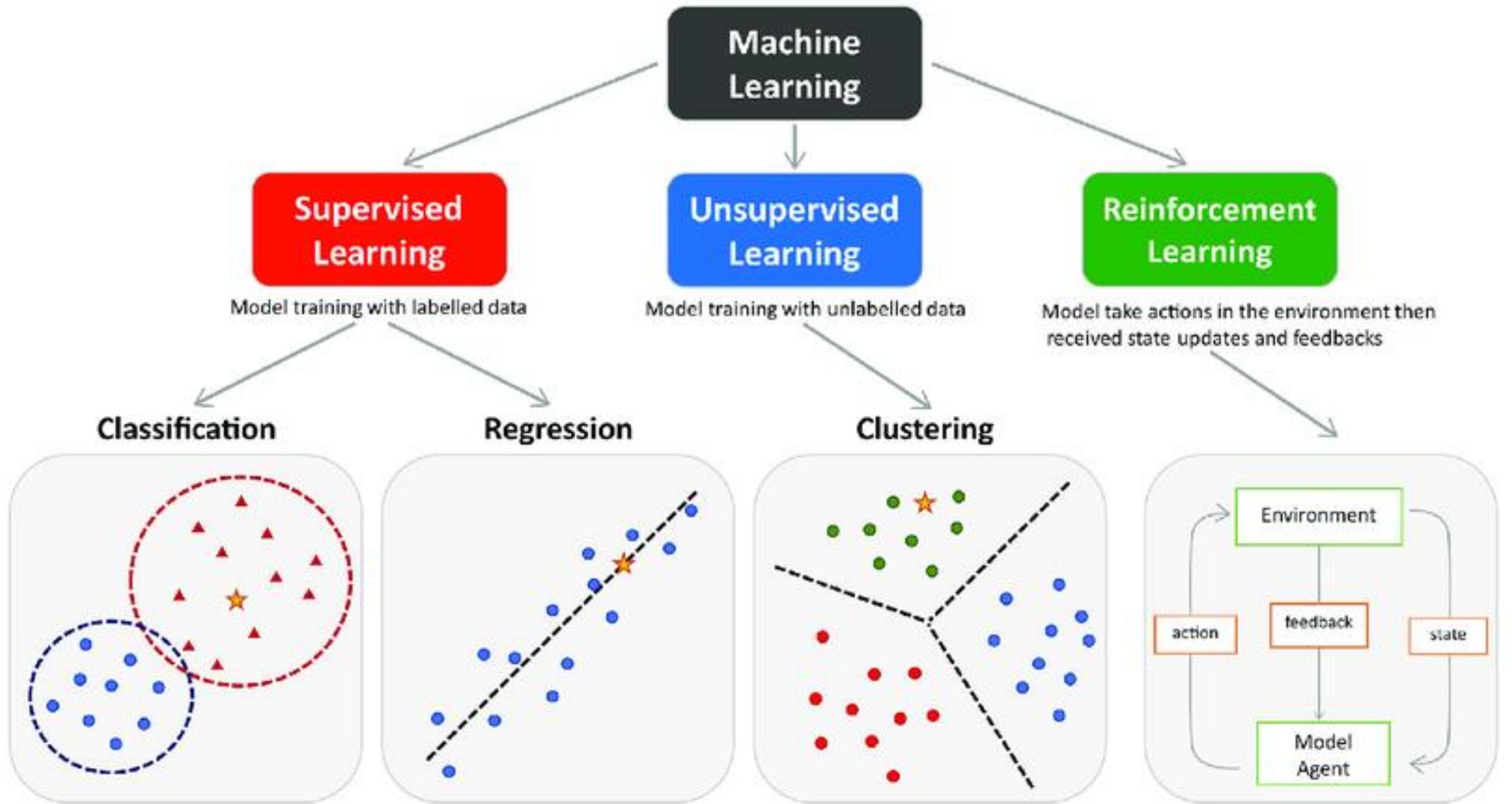
- Machine learning can be seen as a branch of AI or Artificial Intelligence, since, the ability to change experience into expertise or to detect patterns in complex data is a mark of human or animal intelligence.
- As a field of science, machine learning shares common concepts with other disciplines such as statistics, information theory, game theory, and optimization.
- As a subfield of information technology, its objective is to program machines so that they will learn.
- However, it is to be seen that, the purpose of machine learning is not building an automated duplication of intelligent behavior, but using the power of computers to complement and supplement human intelligence. For example, machine learning programs can scan and process huge databases detecting patterns that are beyond the scope of human perception.

Categories of Learning

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.

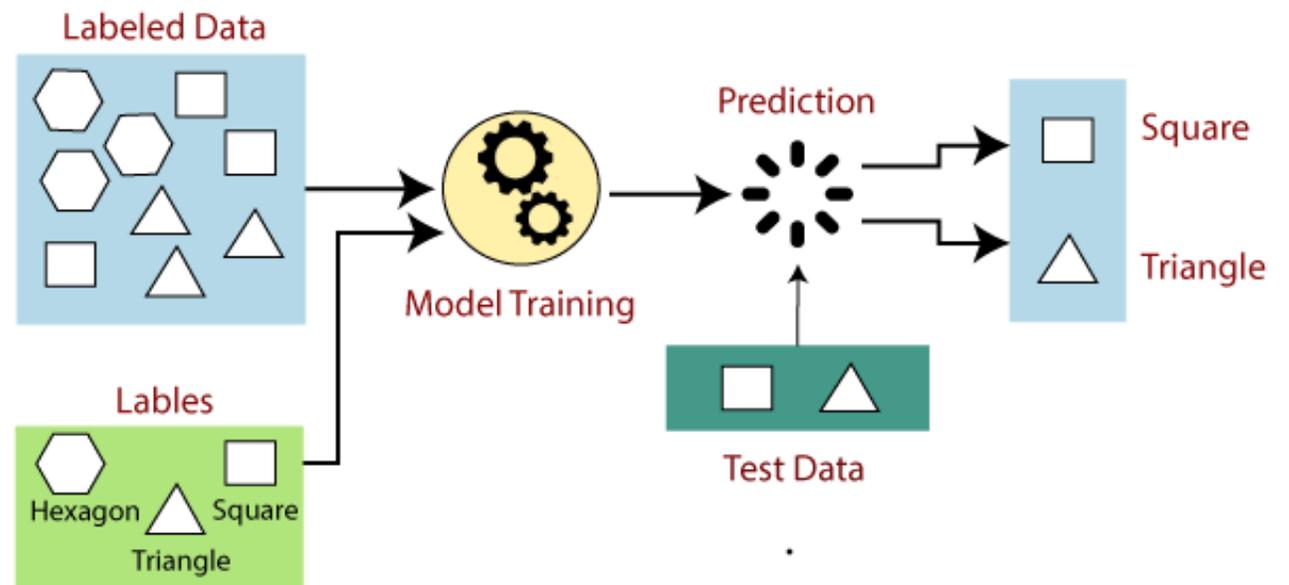
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning





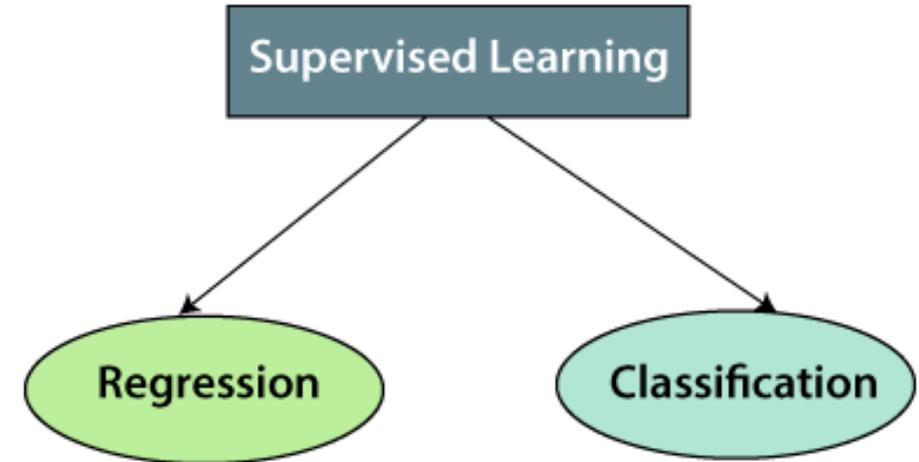
Supervised Learning in AI

- Supervised learning is the types of machine learning in which machines are trained using well "labelled" training data, and on basis of that data, machines predict the output. The labelled data means some input data is already tagged with the correct output.
- In supervised learning, the training data provided to the machines work as the supervisor that teaches the machines to predict the output correctly. It applies the same concept as a student learns in the supervision of the teacher.
- In the real-world, supervised learning can be used for **Risk Assessment, Image classification, Fraud Detection, spam filtering**, etc.



Steps Involved in Supervised Learning:

- First Determine the type of training dataset
- Collect/Gather the labelled training data.
- Split the training dataset into training **dataset**, **test dataset**, and **validation dataset**.
- Determine the input features of the training dataset, which should have enough knowledge so that the model can accurately predict the output.
- Determine the suitable algorithm for the model, such as support vector machine, decision tree, etc.
- Execute the algorithm on the training dataset. Sometimes we need validation sets as the control parameters, which are the subset of training datasets.
- Evaluate the accuracy of the model by providing the test set. If the model predicts the correct output, which means our model is accurate.



A. Regression

- Regression algorithms are used if there is a relationship between the input variable and the output variable. It is used for the prediction of continuous variables, such as Weather forecasting, Market Trends, etc. Below are some popular Regression algorithms which come under supervised learning:

- Linear Regression
- Regression Trees
- Non-Linear Regression
- Bayesian Linear Regression
- Polynomial Regression



Regression

What is the temperature going to be tomorrow?



B. Classification

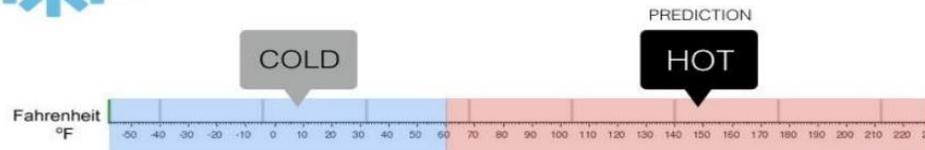
- Classification algorithms are used when the output variable is categorical, which means there are two classes such as Yes-No, Male-Female, True-false, etc.

- Spam Filtering,
- Random Forest
- Decision Trees
- Logistic Regression
- Support vector Machines



Classification

Will it be Cold or Hot tomorrow?



Advantages of Supervised learning:

- With the help of supervised learning, the model can predict the output on the basis of prior experiences.
- In supervised learning, we can have an exact idea about the classes of objects.
- Supervised learning model helps us to solve various real-world problems such as **fraud detection, spam filtering**, etc.

Disadvantages of supervised learning:

- Supervised learning models are not suitable for handling the complex tasks.
- Supervised learning cannot predict the correct output if the test data is different from the training dataset.
- Training required lots of computation times.
- In supervised learning, we need enough knowledge about the classes of object.

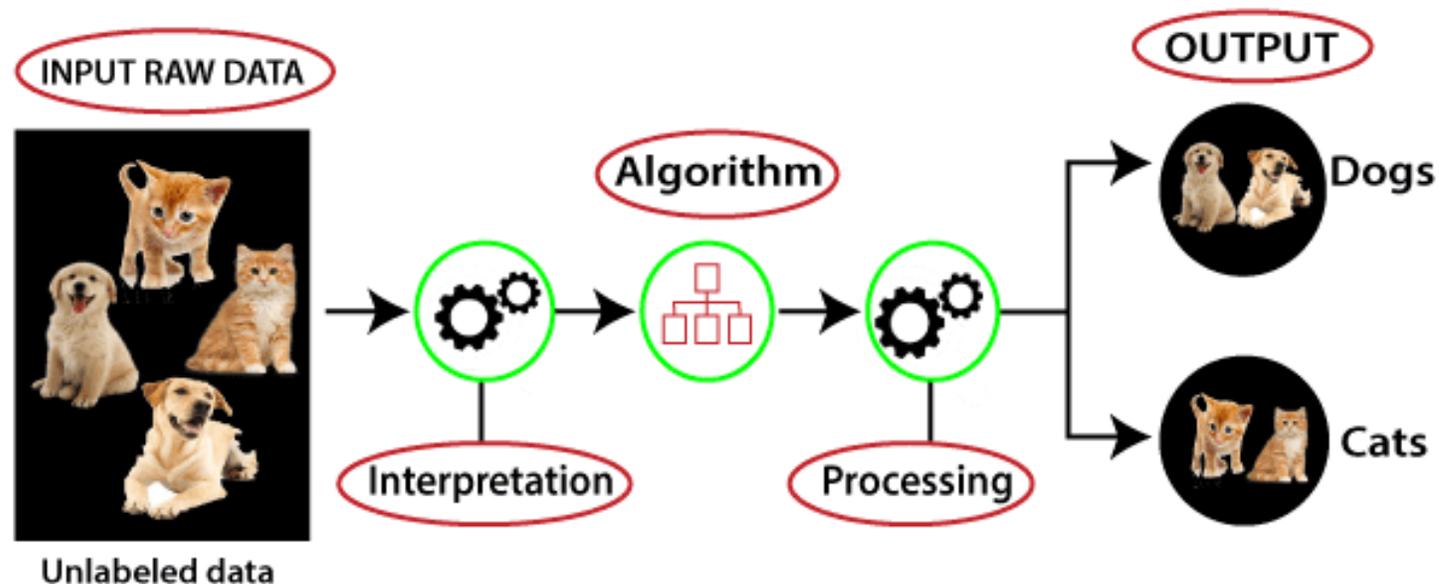
Unsupervised Machine Learning

- supervised machine learning in which models are trained using labeled data under the supervision of training data. But there may be many cases in which we do not have labeled data and need to find the hidden patterns from the given dataset. So, to solve such types of cases in machine learning, we need unsupervised learning techniques.
- *Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision.*
- The goal of unsupervised learning is to **find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.**
- **Example:** Suppose the unsupervised learning algorithm is given an input dataset containing images of different types of cats and dogs. The algorithm is never trained upon the given dataset, which means it does not have any idea about the features of the dataset. The task of the unsupervised learning algorithm is to identify the image features on their own. Unsupervised learning algorithm will perform this task by clustering the image dataset into the groups according to similarities between images.

Why use Unsupervised Learning?

- Unsupervised learning is helpful for finding useful insights from the data.
- Unsupervised learning is much similar as a human learns to think by their own experiences, which makes it closer to the real AI.
- Unsupervised learning works on unlabeled and uncategorized data which make unsupervised learning more important.
- In real-world, we do not always have input data with the corresponding output so to solve such cases, we need unsupervised learning.

Working of Unsupervised Learning



Unsupervised Learning

Clustering

Association

- **Clustering:** Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group. Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.
- **Association:** An association rule is an unsupervised learning method which is used for finding the relationships between variables in the large database. It determines the set of items that occurs together in the dataset. Association rule makes marketing strategy more effective. Such as people who buy X item (suppose a bread) are also tend to purchase Y (Butter/Jam) item. A typical example of Association rule is Market Basket Analysis.

Advantages of Unsupervised Learning

Unsupervised Learning algorithms:

- K-means clustering
- KNN (k-nearest neighbors)
- Hierarchical clustering
- Anomaly detection
- Neural Networks
- Principle Component Analysis
- Independent Component Analysis
- Apriori algorithm
- Singular value decomposition

- Unsupervised learning is used for more complex tasks as compared to supervised learning because, in unsupervised learning, we don't have labeled input data.
- Unsupervised learning is preferable as it is easy to get unlabeled data in comparison to labeled data.

Disadvantages of Unsupervised Learning

- Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.
- The result of the unsupervised learning algorithm might be less accurate as input data is not labeled, and algorithms do not know the exact output in advance.

Aspect	Supervised Learning	Unsupervised Learning
Objective	Predict a target variable based on input data.	Discover patterns, structure, or relationships in data without labeled outcomes.
Labeled Data Required	Yes, training data includes input-output pairs.	No, unlabeled data or data without explicit output labels.
Learning Process	The model is trained to map inputs to outputs.	The model identifies hidden structures or clusters in the data.
Examples	Classification, regression, object detection.	Clustering, dimensionality reduction, association rules.
Evaluation	Performance is measured using metrics like accuracy, RMSE.	Evaluation is often more subjective, based on data exploration and visualization.
Feedback	Error or loss is used to adjust model parameters.	No explicit feedback from labeled data; feedback may come from within the data itself.
Use Cases	Recognizing objects, spam detection, sentiment analysis.	Customer segmentation, anomaly detection, topic modeling.
Supervision Level	High supervision, guidance from labeled data.	Low or no supervision, self-discovery of patterns.
Algorithm Examples	Decision trees, neural networks, support vector machines.	K-means clustering, Principal Component Analysis (PCA).

Genetic Algorithms

- *A genetic algorithm is an adaptive heuristic search algorithm inspired by "Darwin's theory of evolution in Nature."* It is used to solve optimization problems in machine learning. It is one of the important algorithms as it helps solve complex problems that would take a long time to solve.
- Genetic Algorithms are being widely used in different real-world applications, for example, **Designing electronic circuits, code-breaking, image processing, and artificial creativity.**
- Basic terminologies
 - **Population:** Population is the subset of all possible or probable solutions, which can solve the given problem.
 - **Chromosomes:** A chromosome is one of the solutions in the population for the given problem, and the collection of gene generate a chromosome.
 - **Gene:** A chromosome is divided into a different gene, or it is an element of the chromosome.
 - **Allele:** Allele is the value provided to the gene within a particular chromosome.
 - **Fitness Function:** The fitness function is used to determine the individual's fitness level in the population. It means the ability of an individual to compete with other individuals. In every iteration, individuals are evaluated based on their fitness function.
 - **Genetic Operators:** In a genetic algorithm, the best individual mate to regenerate offspring better than parents. Here genetic operators play a role in changing the genetic composition of the next generation.

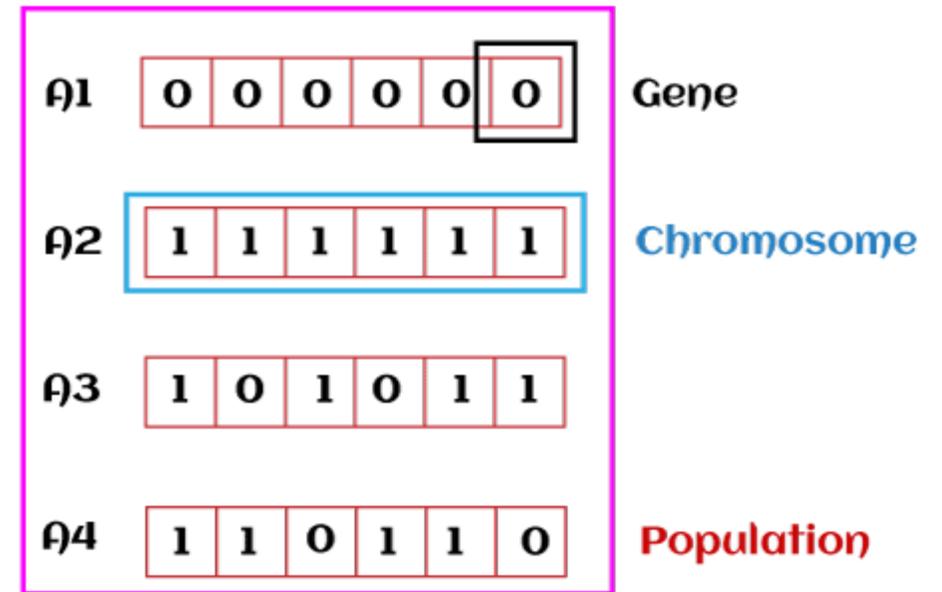
How Genetic Algorithms Works (Different Operators)

- It basically involves five phases to solve the complex optimization problems, which are given as below:

- **Initialization**
- **Fitness Assignment**
- **Selection**
- **Reproduction**
- **Termination**

- **1. Initialization**

The process of a genetic algorithm starts by generating the set of individuals, which is called population. Here each individual is the solution for the given problem. An individual contains or is characterized by a set of parameters called Genes. Genes are combined into a string and generate chromosomes, which is the solution to the problem. One of the most popular techniques for initialization is the use of random binary strings.



2. Fitness Assignment

- Fitness function is used to determine how fit an individual is? It means the ability of an individual to compete with other individuals. In every iteration, individuals are evaluated based on their fitness function. The fitness function provides a fitness score to each individual. This score further determines the probability of being selected for reproduction. The high the fitness score, the more chances of getting selected for reproduction.

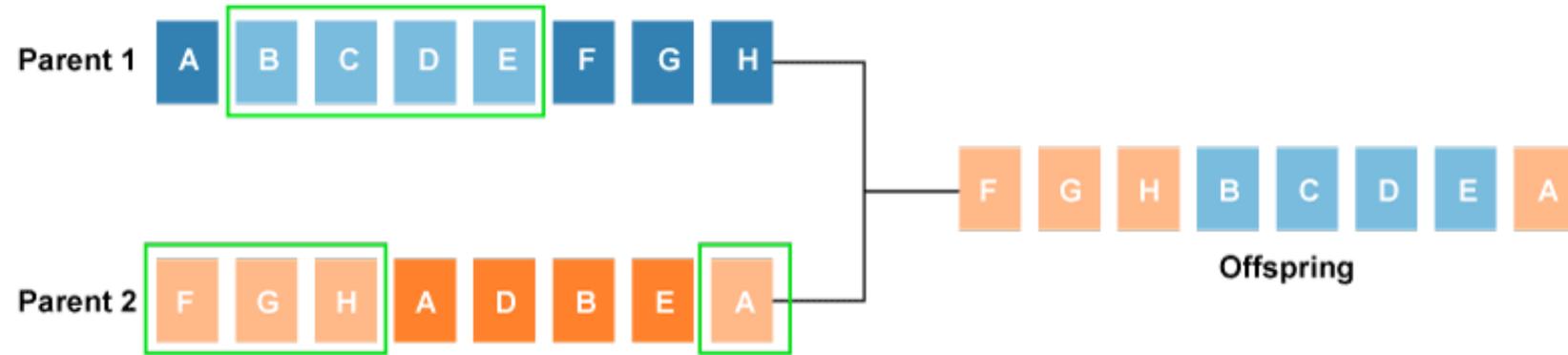
3. Selection

- The selection phase involves the selection of individuals for the reproduction of offspring. All the selected individuals are then arranged in a pair of two to increase reproduction. Then these individuals transfer their genes to the next generation.

4. Reproduction

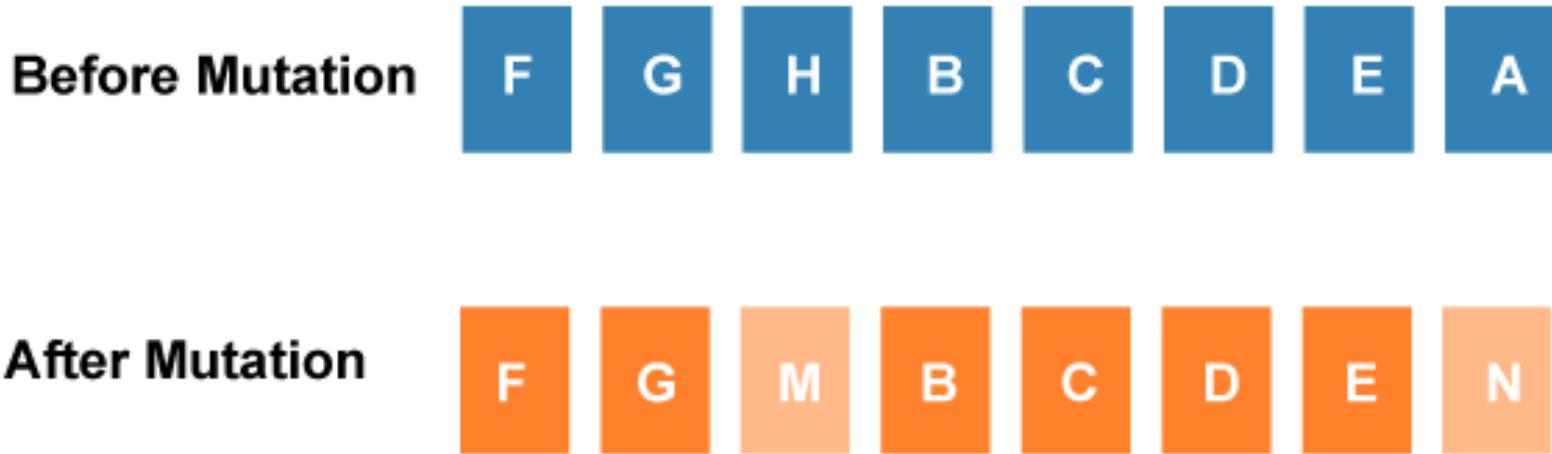
- After the selection process, the creation of a child occurs in the reproduction step. In this step, the genetic algorithm uses two variation operators that are applied to the parent population. The two operators involved in the reproduction phase are given below:
 - Crossover Operator
 - Mutation Operator

a) Crossover Operator



- The crossover plays a most significant role in the reproduction phase of the genetic algorithm. In this process, a crossover point is selected at random within the genes. Then the crossover operator swaps genetic information of two parents from the current generation to produce a new individual representing the offspring.
- The genes of parents are exchanged among themselves until the crossover point is met. These newly generated offspring are added to the population. This process is also called or crossover. Types of crossover styles available:
 - One point crossover
 - Two-point crossover
 - Livery crossover
 - Inheritable Algorithms crossover

b) Mutation Operator



- The mutation operator inserts random genes in the offspring (new child) to maintain the diversity in the population. It can be done by flipping some bits in the chromosomes.
- Mutation helps in solving the issue of premature convergence and enhances diversification.
- The image shows the mutation process. Types of mutation styles available,
 - **Flip bit mutation**
 - **Gaussian mutation**
 - **Exchange/Swap mutation**

5. Termination

After the reproduction phase, a stopping criterion is applied as a base for termination. The algorithm terminates after the threshold fitness solution is reached. It will identify the final solution as the best solution in the population.

Advantages of Genetic Algorithm

- The parallel capabilities of genetic algorithms are best.
- It helps in optimizing various problems such as discrete functions, multi-objective problems, and continuous functions.
- It provides a solution for a problem that improves over time.
- A genetic algorithm does not need derivative information.

Limitations of Genetic Algorithms

- Genetic algorithms are not efficient algorithms for solving simple problems.
- It does not guarantee the quality of the final solution to a problem.
- Repetitive calculation of fitness values may generate some computational challenges.

