Study Material

Module II: Types of Learning and Its Application [PART 2]

Approaches to Machine Learning

A **machine learning approach** is essentially the strategy we take to train a computer model to recognize patterns, make predictions, or take actions based on data. It combines **algorithms** (rules and processes), **techniques** (ways to prepare and process data), and **methods** (training procedures) to solve specific real-world problems.

Classification of Machine Learning

At its core, machine learning is simply a way of achieving AI. Machine learning is an application of artificial intelligence (AI) that enables systems to learn and advance based on experience without being clearly programmed. Machine learning focuses on the development of computer programs that can access data and use it for their ownlearning.

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

1. **Supervised Learning:** Trains models on **labeled data** to predict or classify new, unseen data.
2. **Unsupervised Learning**: Finds patterns or groups in **unlabeled data**, like clustering or dimensionality reduction.
3. **Reinforcement Learning**: Learns through trial and error to maximize rewards, ideal for decision-making tasks.

Figure: Types of Machine Learning

Additional Types -

The following are not part of the original three core types of ML, but they have become increasingly important in real-world applications, especially in deep learning.

- **Self-Supervised Learning**: Self-supervised learning is often considered a **subset of unsupervised learning**, but it has grown into its own field due to its success in training large-scale models. It generates its own labels from the data, without any manual labeling.
- **Semi-Supervised Learning:** This approach combines a small amount of labeled data with a large amount of unlabeled data. It's useful when labeling data is expensive or time-consuming.

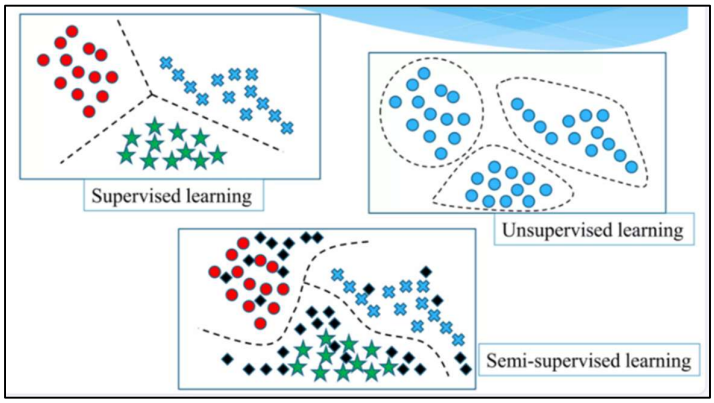


Figure: Supervised Learning vs Unsupervised Learning vs Semi-supervised Learning

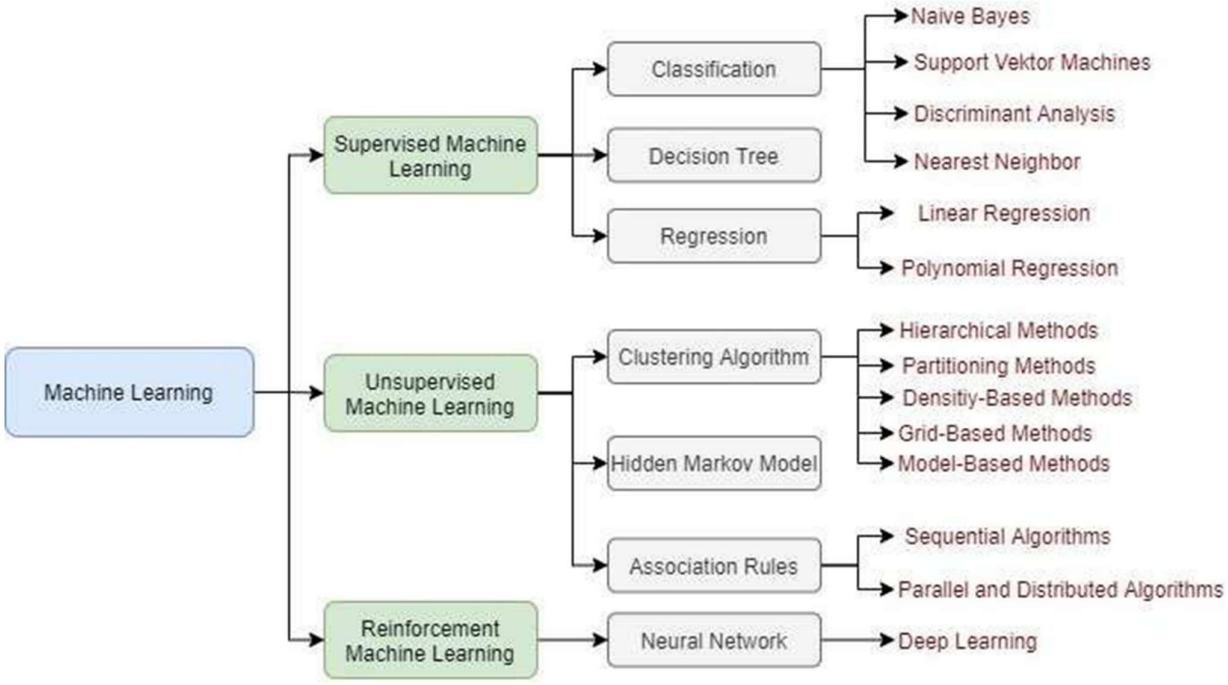


Figure: Classification of Machine Learning

Supervised Learning

Supervised machine learning can utilize what it has learned in the past to apply it to new data, using labelled examples to predict future patterns and events. It learns by explicit example. Supervised learning requires that the algorithm's possible outputs are already known and that the data used to train the algorithm is already labeled with correct answers.

It's like teaching a child that 2+2=4 or showing an image of a dog and teaching the child that it is called a dog. The approach to supervised machine learning is essentially the same – it is presented with all the information it needs to reach pre-determined conclusions. It learns how to reach the conclusion, just like a child would learn how to reach the total of '5' and the few, pre-determined ways to get there, for example, 2+3 and 1+4. If you were to present 6+3 as a way to get to 5, that would be determined as incorrect. Errors would be found and adjusted.

Supervised learning is commonly used in applications where **historical data predicts likely future events.**

Using the previous example, if 6+3 is the most common erroneous way to get to 5, the machine can predict that when someone inputs 6+3, after the correct answer of 9, 5 would be the second most commonly expected result. We can also consider an everyday example – it can foresee when credit card transactions are likely to be fraudulent or which insurance customer is more likely to put forward a claim.

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.

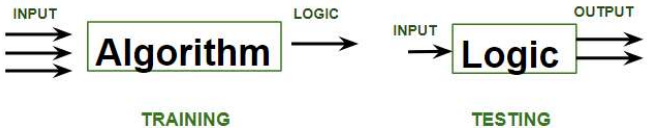
The system creates a model using labeled data to understand the datasets and learn about each data, once the training and processing are done then we test the model byproviding a sample data to check whether it is predicting the exact output or not. The goal of supervised learning is to map input data with the output data. Supervised learning is based on supervision, much like when a student learns under the guidance of a teacher.

Hence, we can say that –
A supervised learning algorithm aims to ***find amapping function to map the input variable(x) with the output variable(y)****.*

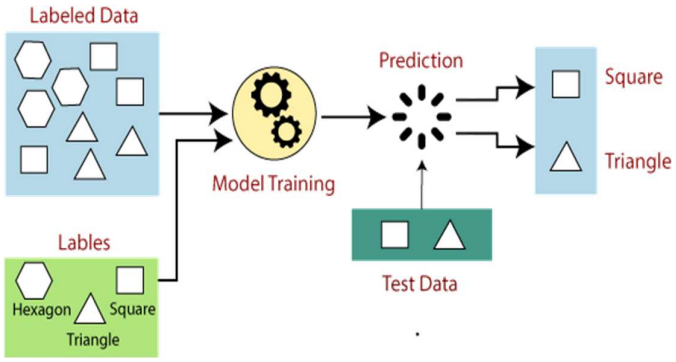
In the real-world, supervised learning can be used for **Risk Assessment, Image Classification, Fraud Detection, Spam Filtering**, etc.

How Supervised Learning Works?

In supervised learning, models are trained using labelled dataset, where the model learns about each type of data. Once the training process is completed, the model is tested on the basis of test data (a subset of the training set), and then it predicts the output.



The working of Supervised learning can be easily understood by the following diagram and example:



Suppose we have a dataset of different types of shapes, which includes squares, rectangles, triangles, and Polygons. Now, the first step is that we need to train the model for each shape, like using the following conditions -

- If the given shape has four sides, and all the sides are equal, then it will belabelled as a **Square**.
- If the given shape has three sides, then it will be labelled as a **triangle**.
- If the given shape has six equal sides, then it will be labelled as a **hexagon**.

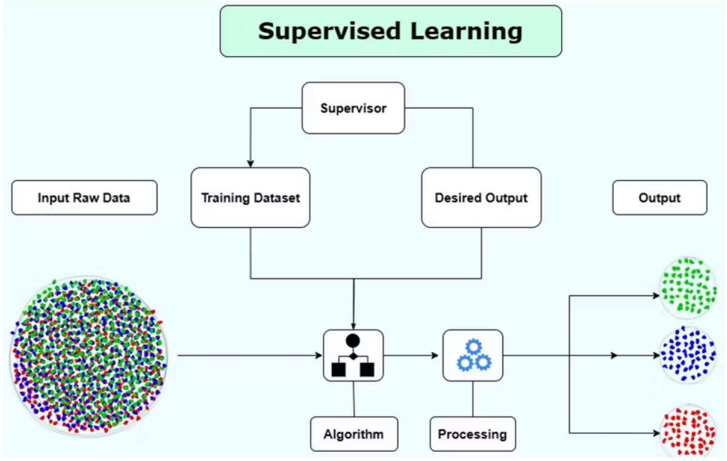
Now, after training, we test our model using the test set, and the task of the model is to identify the shape. The machine is already trained on all types of shapes, and when it finds a new shape, it classifies the shape on the basis of the number of sides and predicts the output.

Steps Involved in Supervised Learning:

The following are some of the common steps involved in supervised learning:

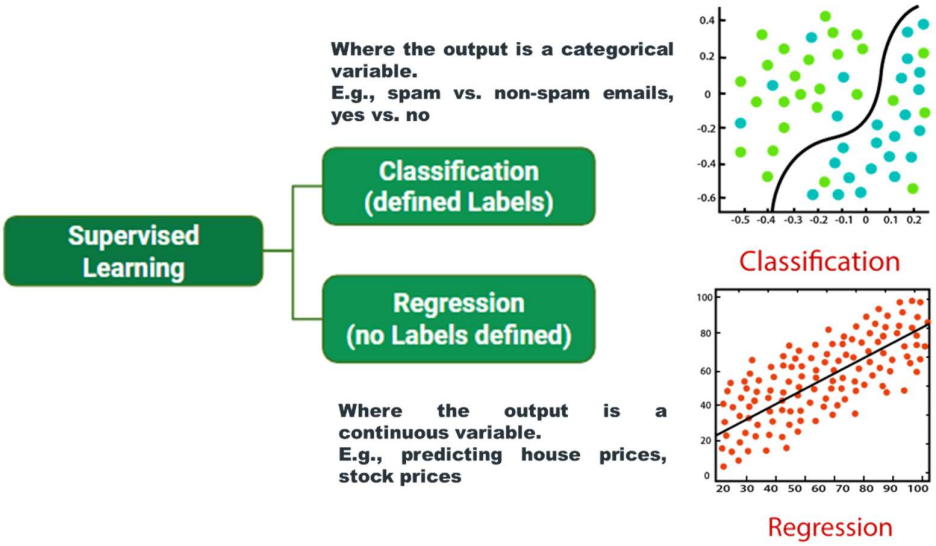
1. Gather labeled data.
2. Split the data into two sets: Training and Testing.
3. Determine the input features, which should have enough knowledge so that the model can accurately predict the output.
4. Select an appropriate algorithm for the model.
5. Execute the algorithm on the training set.
6. Analyze the algorithm's performance on the testing set. If the model gives the correct predicted output, then the model is accurate.
7. If necessary, fine-tune the model to improve performance.

8. Make predictions on new, unlabeled data using the trained model.



Types of supervised Machine learning Algorithms:

Supervised learning can be further divided into two types of problems:



1. **Regression -** Deals with predicting continuous target variables, which represent numerical values. Regression algorithms are used to determine the relationship between he input variables and the output variables. For example, predicting the price of a house based on its size, location and amenities; or forecasting the sales of a product.
   Below are some popular regression algorithms -
   - Linear Regression
   - Regression Trees
   - Non-Linear Regression
   - Bayesian Linear Regression
   - Polynomial Regression
2. **Classification** – Deals with predicting categorical target variables, which represent discrete classes or labels. For instance, classifying emails as spam or not spam, or predicting whether a patient has a high risk of heart disease. Classification algorithms learn to map the input features

to one of the predefined classes.
Below are some popular algorithms -

- Logistic Regression
- Random Forest
- Decision Trees
- Support Vector Machines
- K-Nearest Neighbors (KNN)
- Naive Bayes

Advantages of Supervised Machine Learning

1. **Supervised Learning** models can have high accuracy as they are trained on **labelled data**.
2. The process of decision-making in supervised learning models is often interpretable.
3. It can often be used in pre-trained models which saves time and resources when developing new models from scratch.

Disadvantages of Supervised Machine Learning

1. It has limitations in knowing patterns and may struggle with unseen or unexpected patterns that are not present in the training data.
2. It can be time-consuming and costly as it relies on **labeled** data only.
3. It may lead to poor generalizations based on new data.

Applications of Supervised Learning

Supervised learning is used in a wide variety of applications, including:

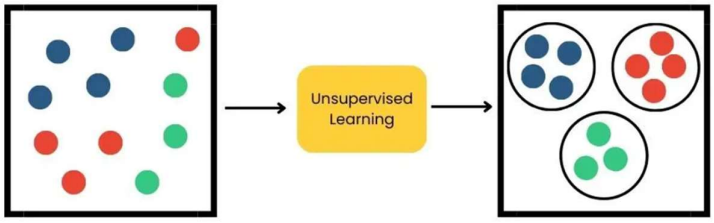
- **Image classification**: Identify objects, faces, and other features in images.
- **Natural language processing:** Extract information from text, such as sentiment, entities, and relationships.
- **Speech recognition**: Convert spoken language into text.
- **Recommendation systems**: Make personalized recommendations to users.
- **Predictive analytics**: Predict outcomes, such as sales, customer churn, and stock prices.
- **Medical diagnosis**: Detect diseases and other medical conditions.
- **Fraud detection**: Identify fraudulent transactions.
- **Autonomous vehicles**: Recognize and respond to objects in the environment.
- **Email spam detection**: Classify emails as spam or not spam.
- **Quality control in manufacturing**: Inspect products for defects.
- **Credit scoring**: Assess the risk of a borrower defaulting on a loan.
- **Gaming**: Recognize characters, analyze player behavior, and create NPCs.
- **Customer support**: Automate customer support tasks.
- **Weather forecasting**: Make predictions for temperature, precipitation, and other meteorological parameters.
- **Sports analytics**: Analyze player performance, make game predictions, and optimize strategies.

Unsupervised Learning

Supervised learning tasks find patterns where we have a dataset of "right answers" to learn from. Unsupervised learning tasks find patterns where we don't. This may bebecause the "right answers" are unobservable, or infeasible to obtain, or maybe for a given problem, there isn't even a "right answer" per se.

Unsupervised learning is a learning method in which a machine learns without any supervision. The training is provided to the machine with a set of data that has not been labeled, classified, or categorized, and the algorithm needs to act on that data without any supervision. **The goal of unsupervised learning is to restructure the input data into new features or a group of objects with similar patterns.**

Hence, here we don't have a predetermined result. The machine tries tofind useful insights from the huge amount of data. Supervised machine learning in which models are trained using labeled data under thesupervision 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.



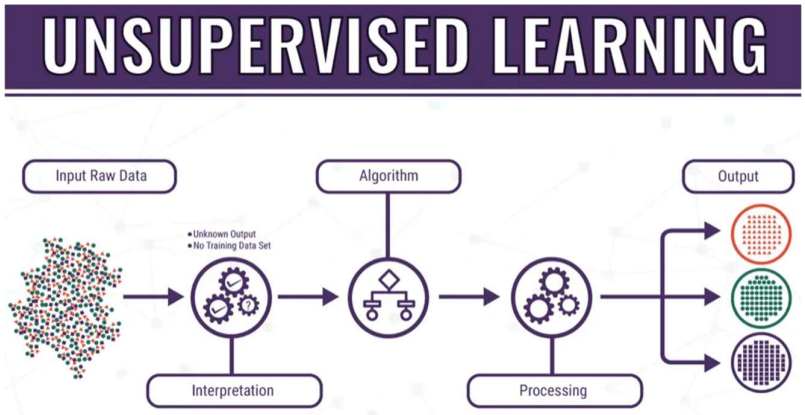
Unsupervised learning works well when the data is transactional. For example, identifying pockets of customers with similar characteristics who can then be targeted inmarketing campaigns.

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 on 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 its own. An unsupervised learning algorithm will perform this task by clustering the image dataset into groups according to similarities between images.

Why use Unsupervised Learning?

Below are some main reasons that describe the importance of Unsupervised Learning:

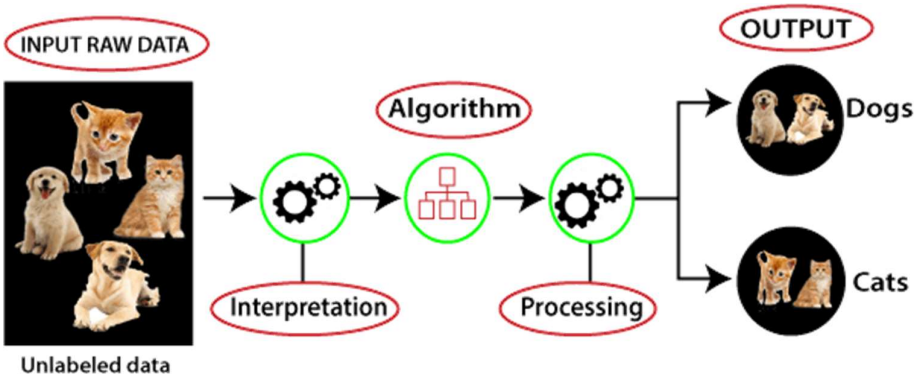
- Unsupervised learning helps find useful insights from the data.
- Unsupervised learning is much similar to a human learns to think by their ownexperiences, which makes it closer to the real AI.
- Unsupervised learning works on unlabeled and uncategorized data, which makes unsupervised learning more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



How does unsupervised learning work?

Unsupervised learning works by analyzing unlabeled data to identify patterns and relationships. The data is not labeled with any predefined categories or outcomes, so the algorithm must find these patterns and relationships on its own. This can be a challenging task, but it can also be very rewarding, as it can reveal insights into the data that would not be apparent from a labeled dataset.

Working of unsupervised learning can be understood by the below diagram:

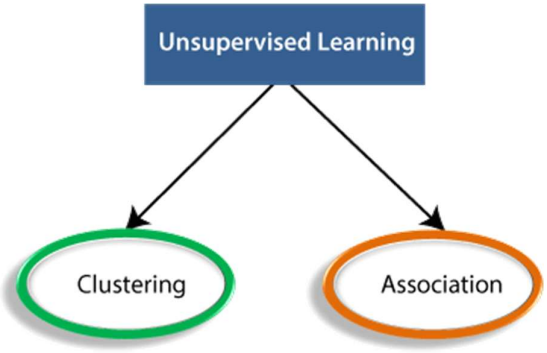


Here, we have taken unlabeled input data, which means it is not categorized, and the corresponding outputs are also not given. Now, this unlabeled input data is fed to the machine learning model to train it. Firstly, it will interpret the raw data to findthe hidden patterns from the data and then will apply suitable algorithms such as k- means clustering, Decision tree, etc. Once it applies the suitable algorithm, the algorithm divides the data objects into groups according to the similarities and

difference between the objects.

Types of Unsupervised Learning Algorithms:

The unsupervised learning algorithm can be further categorized into two types ofproblems:



- **Clustering**: Clustering is a method of grouping objects into clusters suchthat objects with most similarities remains into a group and have 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.

Some common clustering algorithms –

- K-means clustering
- Hierarchal clustering
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
- Gaussian Mixture Models (GMMs)

- **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 rulemakes 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.

Some common Association rule learning algorithms include:

- Apriori Algorithm
- Eclat Algorithm
- FP-Growth Algorithm

Advantages of Unsupervised Learning

1. It doesn't need labeled data so we can start working with large datasets more easily and quickly.

2. This handles large amounts of data and reduce it into simpler forms without losing important patterns which makes it manageable and efficient.
3. It discovers patterns and relationships in the data that were previously unknown which offers valuable insights.
4. By analyzing unlabeled data, it shows meaningful trends and groups that help us to understand our data deeply.

Disadvantages of Unsupervised Learning

1. Without labeled answers, it's difficult to tell how accurate or effective the model is.
2. Lack of clear guidance can lead to less precise results for complex problems.
3. After grouping the data, we may needs to check and label these groupings which can be time-consuming.
4. Missing data, outliers or noise in the data can easily affect the quality of the results.

Application of Unsupervised learning

Unsupervised learning can be used to solve a variety of problems which includes:

1. **Anomaly detection**: It can identify unusual patterns or behaviors in data helps in the detection of fraud, security breaches or system problems.
2. **Scientific discovery**: It can show hidden relationships and patterns in scientific data which leads to new insights and ideas.
3. **Recommendation systems**: It finds similarities in user behavior and preferences to recommend products, movies or music that align with their interests.
4. **Customer segmentation**: It can identify groups of customers with similar characteristics which allows businesses to target marketing campaigns and improve customer service more effectively.

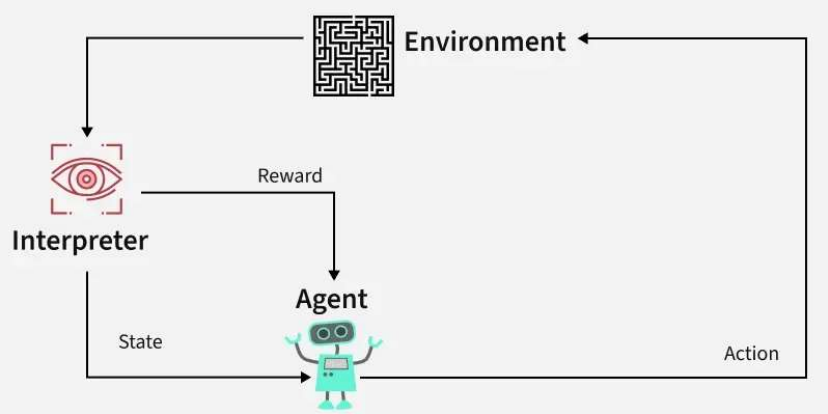
Difference between Supervised and Unsupervised Learning

PARAMETERS	SUPERVISED LEARNING	UNSUPERVISED LEARNING
Input Data	They are trained on labeled data.	They are trained on unlabeled data.
Computational Complexity	Simpler method	Computationally complex
Accuracy	Highly accurate	Less accurate
No. of classes	No. of classes is known	No. of classes is not known
Data Analysis	Uses offline analysis	Uses real-time analysis of data
Algorithms used	Linear and Logistic regression, KNN, Random Forest, multi-class classification, Decision tree, Support Vector Machine, Neural Network.	K-Means clustering, Hierarchical clustering, Apriori algorithm.
Output	Desired output is given.	Desired output is not given.
Training data	Use training data to infer model.	No training data is used.
Complex model	It is not possible to learn larger and more complex models with supervised learning.	It is possible to learn larger and more complex models with unsupervised learning.
Model	We can test our model.	We cannot test our model.
Supervision	Supervised learning needs supervision to train the model.	Unsupervised learning does not need any supervision to train the model.
Classification	Divided into two types: Regression Classification	Divided into two types: Clustering Association
Feedback	It has a feedback mechanism.	It has no feedback mechanism.
Time Consumption	It's more time consuming.	It's less time consuming.
Example	Optical character recognition.	Find a face in an image.

Reinforcement Learning

Reinforcement learning is a type of dynamic programming that trains algorithms using **a system of reward and punishment.** A reinforcement learning algorithm, or agent, learns by interacting with its environment. It receives rewards for performing correctly and penalties for doing so incorrectly. Therefore, it learns without having to be directly taught by a human; **it learns by seeking the greatest reward and minimizing penalty.** This learning is tied to a context because what may lead to maximum reward in one situation may be directlyassociated with a penalty in another.

Reinforcement Learning (RL) is a branch of machine learning that focuses on how agents can learn to make decisions through trial and error to maximize cumulative rewards.



The key components -

- AGENT – The decision maker that performs the action.
- ENVIRONMENT - The world or system in which the agent operates.
- STATE: The situation or condition the agent is currently in.
- ACTIONS - The possible moves or decisions the agent can make.
- REWARD - The feedback or result from the environment based on the agent's action.

Machines and software agents learn to determine the perfect behavior within a specific context, to maximize their performance and reward. Learning occurs via reward feedback which is known as the reinforcement signal.

An everyday example of training pets to relieve themselves outside is a simple way to illustrate this. The goal is to get the pet into the habit of going outside rather than in the house. The training then involves rewards and punishments intended for the pet's learning. It gets a treat for going outsideor has its nose rubbed in its mess if it fails to do so.

Facebook's News Feed is an example most of us will be able to understand. Facebook uses machine learning to personalize people's feeds. If you frequently read or "like" a particular friend's activity, the News Feed will begin to bring up more of that friend's activity more often and nearer to the top. Should you stop interacting with this friend's activity in the same way, the data set will be updated and the News Feed will consequently adjust.

Key Features of Reinforcement Learning

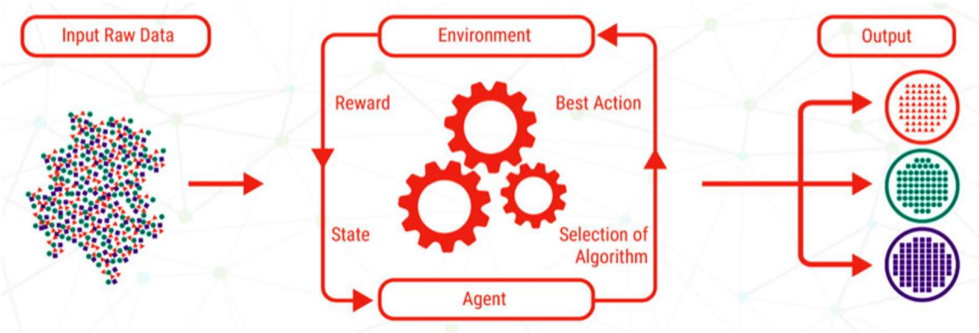
- In RL, the agent is not instructed about the environment and what actions needto be taken.
- It is based on the hit and trial process.
- The agent takes the next action and changes states according to the feedback ofthe previous action.
- The agent may get a delayed reward.
- The environment is stochastic, and the agent needs to explore it to reach to getthe maximum positive rewards.

How Reinforcement Learning Works?

The RL process involves an agent performing actions in an environment, receiving rewards or penalties based on those actions, and adjusting its behavior accordingly. This loop helps the agent improve its decision-making over time to maximize the **cumulative reward**.

Here's a breakdown of RL components:

- **Policy**: A strategy that the agent uses to determine the next action based on the current state.
- **Reward Function**: A function that provides feedback on the actions taken, guiding the agent towards its goal.
- **Value Function**: Estimates the future cumulative rewards the agent will receive from a given state.
- **Model of the Environment**: A representation of the environment that predicts future states and rewards, aiding in planning.

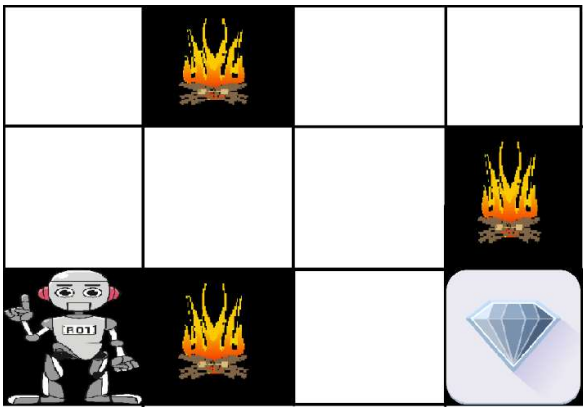


Reinforcement Learning Example: Navigating a Maze

Imagine a robot navigating a maze to reach a diamond while avoiding fire hazards. The goal is to find the optimal path with the least number of hazards while maximizing the reward:

- Each time the robot moves correctly, it receives a reward.
- If the robot takes the wrong path, it loses points.

The robot learns by exploring different paths in the maze. By trying various moves, it evaluates the rewards and penalties for each path. Over time, the robot determines the best route by selecting the actions that lead to the highest cumulative reward



The robot's learning process can be summarized as follows:

1. **Exploration**: The robot starts by exploring all possible paths in the maze, taking different actions at each step (e.g., move left, right, up, or down).

2. **Feedback**: After each move, the robot receives feedback from the environment:
   - A positive reward for moving closer to the diamond.
   - A penalty for moving into a fire hazard.

3. **Adjusting Behavior**: Based on this feedback, the robot adjusts its behavior to maximize the cumulative reward, favoring paths that avoid hazards and bring it closer to the diamond.

4. **Optimal Path**: Eventually, the robot discovers the optimal path with the least number of hazards and the highest reward by selecting the right actions based on past experiences.

Types of Reinforcement Learning

There are two main types of reinforcement learning: value-based and policy-based.

- Value-based Reinforcement learning algorithms learn a value function, which maps states to their expected rewards. The agent then chooses the action that leads to the state with the highest value.

- Policy-based Reinforcement learning algorithms learn a policy, which maps states to actions. The agent then chooses the action that is prescribed by the policy.

Application of Reinforcement Learning

1. **Robotics:** RL is used to automate tasks in structured environments such as manufacturing, where robots learn to optimize movements and improve efficiency.

2. **Game Playing:** Advanced RL algorithms have been used to develop strategies for complex games like chess, Go, and video games, outperforming human players in many instances.

3. **Industrial Control:** RL helps in real-time adjustments and optimization of industrial operations, such as refining processes in the oil and gas industry.

4. **Personalized Training Systems:** RL enables the customization of instructional content based on an individual's learning patterns, improving engagement and effectiveness.

Advantages of Reinforcement Learning

1. **Solving Complex Problems:** RL is capable of solving highly complex problems that cannot be addressed by conventional techniques.

2. **Error Correction:** The model continuously learns from its environment and can correct errors that occur during the training process.

3. **Direct Interaction with the Environment:** RL agents learn from real-time interactions with their environment, allowing adaptive learning.

4. **Handling Non-Deterministic Environments:** RL is effective in environments where outcomes are uncertain or change over time, making it highly useful for real-world applications.

Disadvantages of Reinforcement Learning

1. **Not Suitable for Simple Problems**: RL is often an overkill for straightforward tasks where simpler algorithms would be more efficient.

2. **High Computational Requirements**: Training RL models requires a significant amount of data and computational power, making it resource-intensive.

3. **Dependency on Reward Function**: The effectiveness of RL depends heavily on the design of the reward function. Poorly designed rewards can lead to suboptimal or undesired behaviors.

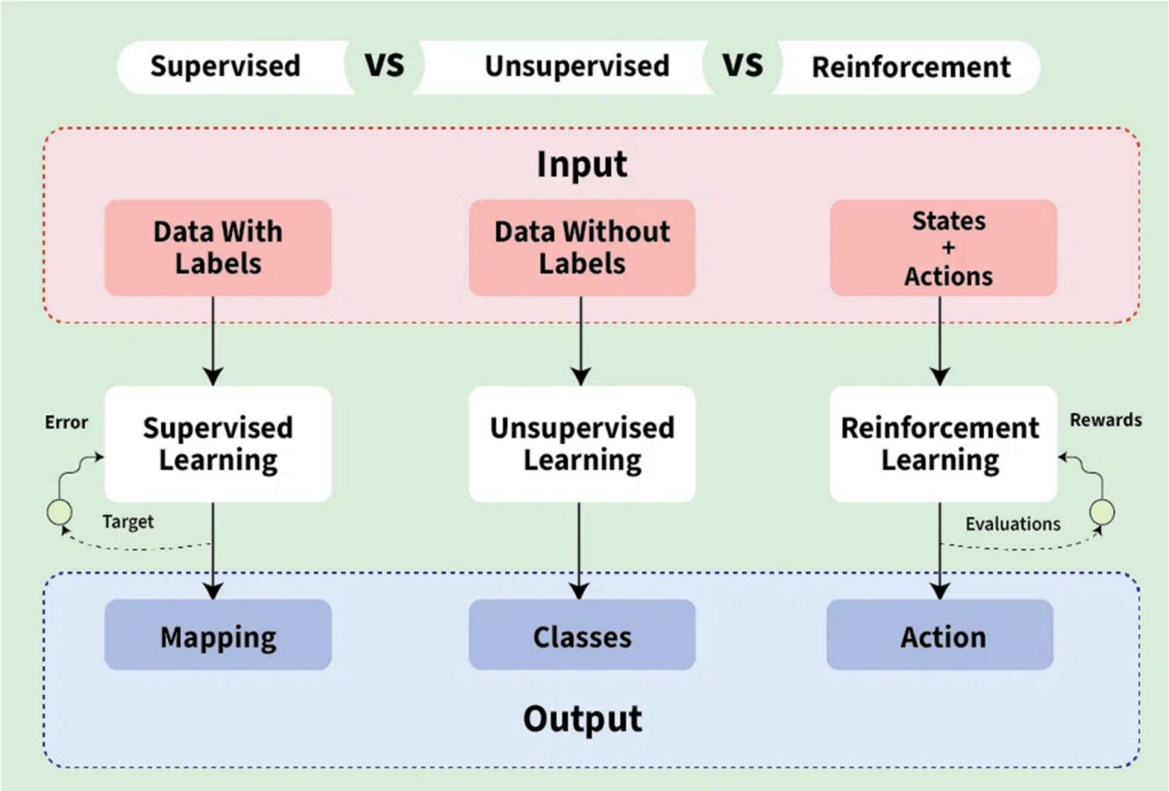
4. **Difficulty in Debugging and Interpretation**: Understanding why an RL agent makes certain decisions can be challenging, making debugging and troubleshooting complex.

Difference between Reinforcement Learning and Supervised Learning

Reinforcement Learning	Supervised Learning
RL works by interacting with the environment.	Supervised learning works on the existing dataset.
The RL algorithm works like the human brain works when makingsome decisions.	Supervised Learning works as when a human learns things in thesupervision of a guide.
There is no labeled dataset ispresent	The labeled dataset is present.
No previous training is provided tothe learning agent.	Training is provided to the algorithm so that it can predict the output.
RL helps to take decisions sequentially.	In Supervised learning, decisions are made when input is given.

Supervised Learning VS Unsupervised Learning VS Reinforcement Learning

Criteria	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Definition	Learns from labeled data	Identifies patterns in unlabeled data	Learns through interaction with environment
Type of Data	Labeled data	Unlabeled data	No predefined data learn from environment
Type of Problems	Classification, Regression	Clustering, Association	Sequential decision-making
Supervision	Requires external supervision	No supervision	No supervision, learns from feedback
Algorithms	SVM, Decision Trees, Neural Networks	K-Means, PCA, Autoencoders	Q-learning, DQN, SARSA
Goal	Predict outcomes accurately	Discover hidden patterns	Optimize actions for maximum rewards
Applications	Medical diagnosis, fraud detection	Customer segmentation, anomaly detection	Self-driving cars, robotics, gaming



Real-World Applications

Machine Learning Type	Domain	Examples
Supervised Learning	Healthcare	Disease diagnosis like cancer detection
	Finance	Loan approval, credit risk assessment
	NLP	Sentiment analysis, text classification
Unsupervised Learning	E-commerce	Product recommendation, customer segmentation
	Cybersecurity	Fraud detection, intrusion detection
	Biology	Gene classification, dimensionality reduction
Reinforcement Learning	Autonomous Driving	Self-driving cars learning optimal driving behavior
	Robotics	Training robots for automated assembly tasks
	Gaming	AI-driven strategy games like AlphaGo

Choosing the Right Learning Approach

- **Supervised Learning:** When labeled data is available for prediction tasks like spam filtering, stock price forecasting.
- **Unsupervised Learning:** When exploring data structures without predefined labels like customer segmentation, anomaly detection.
- **Reinforcement Learning:** When decision-making is required in a dynamic environment like game AI, robotics, self-driving cars.

Each learning paradigm has unique applications and the choice depends on the nature of the problem and the data available.