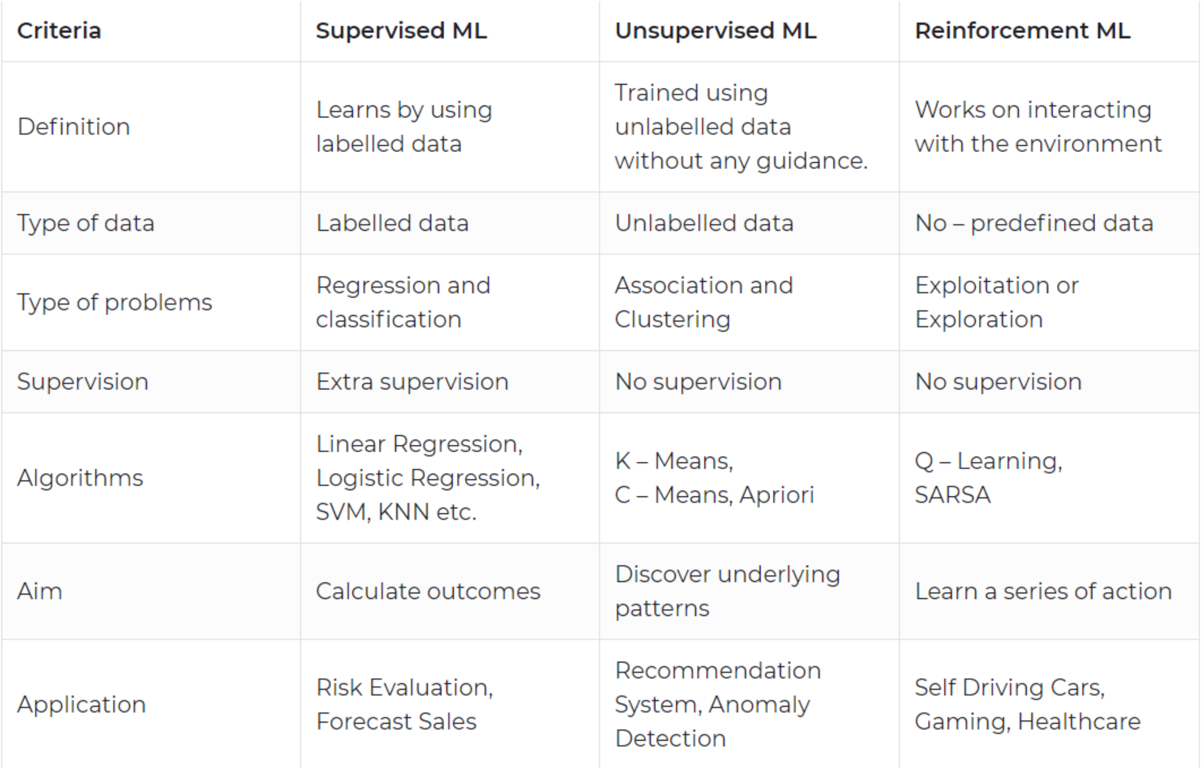
1. What is Reinforcement Learning?

Reinforcement Learning (RL) is the science of decision-making. It is about learning the optimal behavior in an environment to obtain the maximum reward. In RL, the data is accumulated from machine learning systems that use a trial-and-error method. Data is not part of the input that we would find in supervised or unsupervised machine learning.

Reinforcement learning uses algorithms that learn from outcomes and decide which action to take next. After each action, the algorithm receives feedback that helps it determine whether the choice it made was correct, neutral, or incorrect. It is a good technique to use for automated systems that have to make a lot of small decisions without human guidance.

It performs actions with the aim of maximizing rewards, or in other words, it is learning by doing in order to achieve the best outcomes.

2. Differentiate between supervised, unsupervised and reinforcement learning.

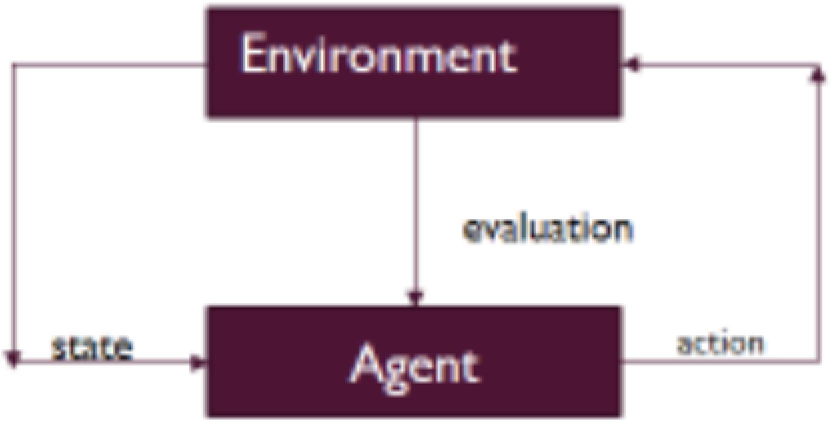


Criteria	Supervised ML	Unsupervised ML	Reinforcement ML
Definition	Learns by using labelled data	Trained using unlabelled data without any guidance.	Works on interacting with the environment
Type of data	Labelled data	Unlabelled data	No – predefined data
Type of problems	Regression and classification	Association and Clustering	Exploitation or Exploration
Supervision	Extra supervision	No supervision	No supervision
Algorithms	Linear Regression, Logistic Regression, SVM, KNN etc.	K – Means, C – Means, Apriori	Q – Learning, SARSA
Aim	Calculate outcomes	Discover underlying patterns	Learn a series of action
Application	Risk Evaluation, Forecast Sales	Recommendation System, Anomaly Detection	Self Driving Cars, Gaming, Healthcare

3. Explain any one application of RL.

RL enables robots to learn complex tasks by trial and error, which is especially useful in scenarios where it is difficult or impossible to pre-program all the possible actions. For example, consider a robot tasked with picking and placing objects in a warehouse. RL can be used to train the robot to learn how to pick and place objects on its own, interact with the environment, take actions, receive feedback, and learn from its mistakes. Over time, the robot improves its performance and its actions become more efficient and accurate. RL has been used to train robots for a variety of tasks, such as locomotion, manipulation, and navigation, and to perform tasks in environments with changing conditions.

4. Explain RL framework with example



The RL framework consists of an agent, an environment, actions, states, rewards, and policies.
Agent: The entity that interacts with the environment, learns, and takes actions
Environment: The surrounding where the agent interacts and receives feedback (reward)
Actions: The decisions are taken by the agent in response to the environment
States: The current situation or context of the environment at a given time step
Rewards: The feedback signal received by the agent indicating how well it did in the environment
Policies: A mapping of states to actions that guide the agent's behavior

Example: Let's consider the game of chess. In this case, the agent is the computer program that plays chess, the environment is the chessboard, the actions are the moves that the agent can make, the states are the positions of the chess pieces on the board, the rewards are the points received by the agent for winning or losing the game, and the policy is the set of rules the agent follows to make decisions. The chess-playing agent will explore the board by making moves and evaluating the resulting state. The agent learns by trial and error and adjusts its policy to increase its chances of winning the game. The reward function would give a positive reward for winning, a negative reward for losing, and a neutral reward for a draw.

5. Describe the elements of RL with an example

Beyond the agent and the environment, the four main sub-elements of a reinforcement learning system are:

Policies: A mapping of states to actions that guide the agent's behavior
Rewards: The feedback signal received by the agent indicating how well it did in the environment
A value function is the total amount of reward an agent can expect to accumulate over the future, starting from that state (in the long run)
A model of the environment mimics the behavior of the environment, or more generally, allows inferences to be made about how the environment will behave

Let's consider the game of chess. In this case, the agent is the computer program that plays chess, the environment is the chessboard, the actions are the moves that the agent can make, the states are the positions of the chess pieces on the board, the rewards are the points received by the agent for winning or losing the game, and the policy is the set of rules the agent follows to make decisions. The chess-playing agent will explore the board by making moves and evaluating the resulting state. The agent learns by trial and error and adjusts its policy to increase its chances of winning the game. The reward function would give a positive reward for winning, a negative reward for losing, and a neutral reward for a draw.

6. Explain deterministic and stochastic policy with examples.

In Reinforcement Learning (RL), the policy is a function that maps states to actions. A policy can be deterministic or stochastic.

Deterministic Policy: A deterministic policy is a mapping from states to a single action. In other words, given a state, the policy will always choose the same action. A deterministic policy is represented as follows:

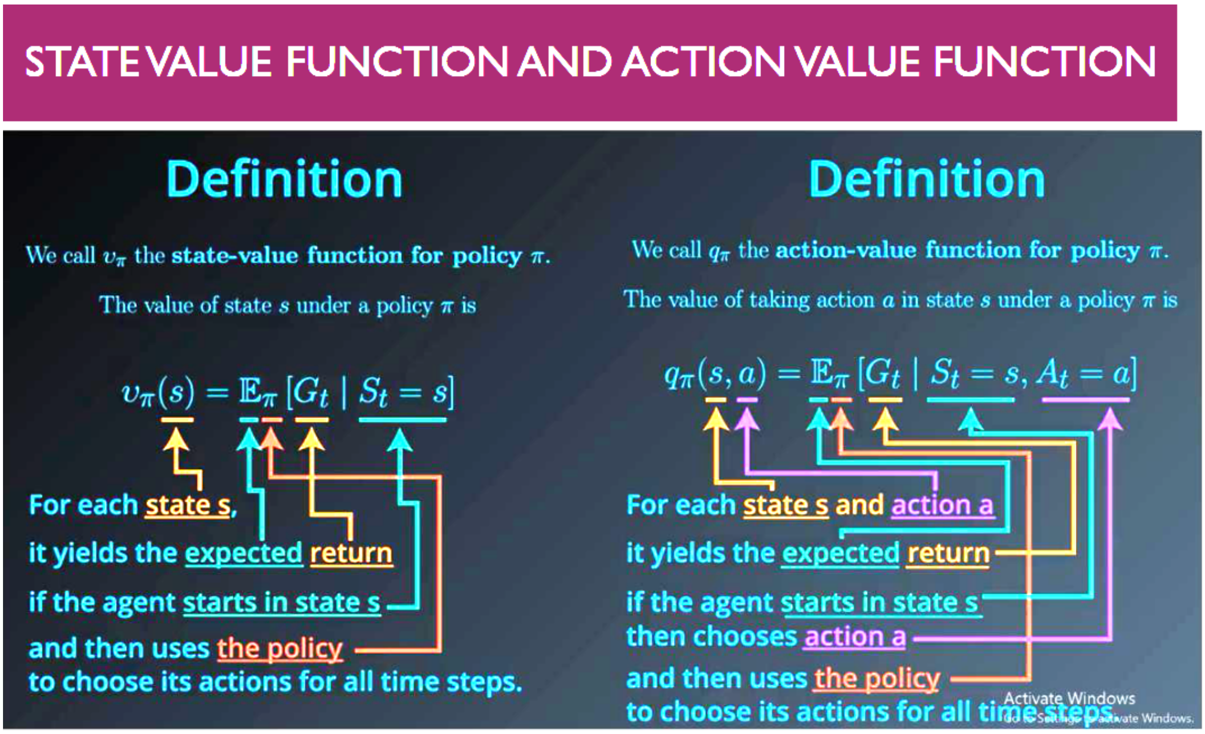
$\pi:(s) \rightarrow a$
Where π is the policy function, s is the state, and a is the action chosen by the policy in state s.

For example, consider a game of chess where the agent is the player. A deterministic policy in chess would always choose a particular move for a given position. If the agent observes the current state of the chessboard as a state, the deterministic policy would always suggest the same move for that state.

Stochastic Policy: If an agent follows policy π at time t, then $\pi(a|s)$ is the probability that $A_t = a$ if $S_t = s$. This means that at time t, under policy π, the probability of taking action a in state s is $\pi(a|s)$. For each state $s \in S$, π is a probability distribution over $a \in A(s)$

For example, consider an autonomous car learning to navigate a busy street. A stochastic policy in this case would choose the next action probabilistically based on the traffic condition, pedestrian activity, and other factors in the environment. The policy may suggest a different action with different probabilities for the same state based on the conditions at the time.

7. Discuss state-value function and action-value function for policy π with their mathematical definition



STATE VALUE FUNCTION AND ACTION VALUE FUNCTION

Definition

We call v_π the **state-value function for policy** π.

The value of state s under a policy π is

$$v_\pi(s) = \mathbb{E}_\pi\left[G_t \mid S_t = s\right]$$

For each state s,
it yields the expected return
if the agent starts in state s
and then uses the policy
to choose its actions for all time steps.

Definition

We call q_π the **action-value function for policy** π.

The value of taking action a in state s under a policy π is

$$q_\pi(s,a) = \mathbb{E}_\pi\left[G_t \mid S_t = s, A_t = a\right]$$

For each state s and action a
it yields the expected return
if the agent starts in state s
then chooses action a
and then uses the policy
to choose its actions for all time steps.

8. Explain the concept of exploration and exploitation in RL

Exploitation is defined as a greedy approach in which agents try to get more rewards by using estimated value but not the actual value. So, in this technique, agents make the best decision based on current information.

Unlike exploitation, in exploration techniques, agents primarily focus on improving their knowledge about each action instead of getting more rewards so that they can get long-term benefits. So, in this technique, agents work on gathering more information to make the best overall decision.

9. What do you mean by exploration and exploitation dilemma in RL? Explain with an example.

The dilemma is between choosing what you know and getting something close to what you expect ('exploitation') and choosing something you aren't sure about and possibly learning more ('exploration'). The reinforcement learning agent will be in a dilemma on whether to exploit the partial knowledge to receive some rewards or it should explore unknown actions which could result in many rewards.

In the example of a mobile robot navigating a grid-world environment, the exploration-exploitation dilemma arises when the robot must decide whether to explore new actions or exploit its existing knowledge. Exploration involves trying different actions to gather information about the environment, while exploitation entails using current knowledge to select actions that yield immediate rewards. The robot needs to strike a balance between exploring to discover potentially better actions and exploiting its current knowledge to maximize long-term rewards. By exploring, the robot gathers information about obstacles and the target location, but it incurs short-term penalties. By exploiting, the robot prioritizes actions expected to yield higher rewards based on its current knowledge, but it risks missing out on undiscovered rewards. Balancing exploration and exploitation is crucial for the robot to navigate the environment effectively and reach its target.

10. Compare evolutionary methods and RL

An evolutionary algorithm is considered a component of evolutionary computation in artificial intelligence. An evolutionary algorithm functions through the selection process in which the least fit members of the population set are eliminated, whereas the fit members are allowed to survive and continue until better solutions are determined. Evolutionary algorithms are a heuristic-based approach to solving problems that cannot be easily solved in polynomial time, such as classically NP-Hard problems, and anything else that would take far too long to exhaustively process.

Reinforcement learning uses the concept of one agent, and the agent learns by interacting with the environment in different ways. In evolutionary algorithms, they usually start with many "agents" and only the "strong ones survive". Reinforcement learning agent(s) learns both positive and negative actions, but evolutionary algorithms only learn the optimal, and the negative or suboptimal solution information is discarded and lost.

11. Describe how RL and Evolutionary methods will approach the scenario of changing room temperature from 15 ° to 23 °

Using Reinforcement learning, the agent will try a bunch of different actions to increase and decrease the temperature. Eventually, it learns that increasing the temperature yields a good reward. But it also learns that reducing the temperature will yield a bad reward.

For evolutionary algorithms, it initiates with a bunch of random agents that all have a preprogrammed set of actions it is going to do. Then the agents that have the "increase temperature" action survive and move on to the next generation. Eventually, only agents that increase the temperature survive and are deemed the best solution. However, the algorithm does not know what happens if you decrease the temperature.

12. What is immediate RL? Give example

Immediate Reinforcement Learning (RL) is a type of RL where the reward signal is received immediately after each action. In Immediate RL, the agent learns by interacting with the environment in a trial-and-error fashion, receiving a reward or penalty immediately after each action.

In robot control tasks, the agent needs to take actions based on the immediate state of the environment to achieve a specific objective, such as moving to a target location.

13. Define Agent, Action, Environment, reward, policy

Agent: The entity that interacts with the environment, learns, and takes actions
Actions: The decisions are taken by the agent in response to the environment
Environment: The surrounding where the agent interacts and receives feedback (reward)
Rewards: The feedback signal received by the agent indicating how well it did in the environment
Policies: A mapping of states to actions that guide the agent's behavior

14. Classify the following applications under Supervised, Unsupervised and Reinforcement Learning techniques.

SL - Risk evaluation system, weather forecast
UL - Recommendation system, Cyber fraud detection
RL - Self-driving car, game of chess

15. You have a bank credit dataset and want to take a decision whether to approve a loan of the applicant based on his profile. Which learning technique will be used?

Supervised Learning

16. You have to establish a mathematical equation for distance as a function of speed. So that you can predict the distance when only speed is known. Which learning technique can be used to implement this?

Supervised Learning