Reinforcement Learning

Introduction

Dr. Alireza Aghamohammadi

- **Reinforcement Learning (RL)** is a framework where an agent learns to make decisions in an environment, aiming to maximize accumulated rewards.
- **Example (Finance):** In a financial application, an RL agent could act as a virtual trader. The agent (trader) decides when to buy or sell assets (actions) on a market platform (environment) to maximize profits (reward).
- **Agent:** The decision maker that selects actions based on the information available from the environment.
- **Environment:** Everything external to the agent, including factors the agent cannot directly control.
- **Example (Robotics):** When training a robot to pick up objects, the objects, the tray, and external conditions (e.g., wind) are all part of the environment.
- **State Space:** A set of variables that represent the environment and their possible values.
- **State:** A specific configuration of the environment's variables, representing a particular scenario or situation.

- **Partial Observability:** Agents often do not have access to the full state of the environment.
- **Observation:** The information an agent can perceive from the environment. It is derived from the actual state but may not provide complete details.
- **Example (Robotics):** The agent might only receive a camera image. While the exact position of objects exists in the environment, the agent only perceives what the camera shows.
- **Actions:** The choices made by the agent that influence the environment. Actions can alter the state of the environment.
- **State Transitions and Rewards:** The environment may change its state in response to the agent's actions and provide a reward signal as feedback.
- **Experience:** At each time step, the agent observes the environment, takes an action, and receives a new observation and reward. This cycle forms an **experience**, represented as (state, action, reward, next state).
- Each experience provides a chance for the agent to learn and improve its decision-making strategy.

- The problem the agent aims to solve may have a natural ending (**episodic tasks**) or continue indefinitely (**continuing tasks**).
  - **Episodic Tasks:** Have a clear endpoint, such as games or specific goal-oriented tasks.
  - **Continuing Tasks:** Do not have a defined ending, like maintaining balance or learning continuous motion.
- **Episode:** A sequence of time steps from the start to the end of an episodic task is called an episode. Agents may need multiple episodes to learn effectively.
- The agent's actions may not have an immediate impact on the environment.
- Rewards may be infrequent or delayed, requiring the agent to perform a series of actions before receiving feedback.
- **Sequential Feedback:** The agent must learn from a sequence of experiences, considering the entire sequence of actions over time to improve its performance.

- **Temporal Credit Assignment Problem:** The challenge of determining which specific actions or states in a sequence contributed to the final reward.
- When rewards are delayed, it is difficult to know if the reward should be attributed to the most recent action or to an earlier decision made by the agent.
- A reward indicates the quality of the outcome (e.g., a high score is "good" and a low score is "bad"), but it does not provide direct information about what actions should have been taken to achieve a better outcome.
- **Exploration vs. Exploitation Trade-off:** The agent must decide between:
  - **Exploration:** Trying new actions to discover potentially better rewards.
  - **Exploitation:** Leveraging its existing knowledge by choosing actions it already knows yield good results.