Introduction to Deep Reinforcement Learning

**Reinforcement Learning:**

Reinforcement Learning is a computational approach of learning from actions. We build an agent that learns from the environment by interacting with it through trial and error and receiving rewards (negative or positive) as feedback.

The RL process is a loop that outputs a sequence of state, action, reward, and next state.

**Agent:**

An agent learns to make decisions by trial and error, with rewards and punishments from the surroundings.

**Environment:**

An environment is a simulated world where an agent can learn by interacting with it.

**Markov Property:**

It implies that the action taken by our agent is conditional solely on the present state and independent of the past states and actions.

**Observations/State:**

State: Complete description of the state of the world.

Observation: Partial description of the state of the environment/world.

**Actions:**

Discrete Actions: Finite number of actions, such as left, right, up, and down.

Continuous Actions: Infinite possibility of actions; for example, in the case of self-driving cars, the driving scenario has an infinite possibility of actions occurring.

**Rewards and Discounting:**

Rewards: Fundamental factor in RL. Tells the agent whether the action taken is good/bad.

RL algorithms are focused on maximizing the cumulative reward.

Reward Hypothesis: RL problems can be formulated as a maximisation of (cumulative) return.

Discounting is performed because rewards obtained at the start are more likely to happen as they are more predictable than long-term rewards.

**Tasks:**

Episodic: Has a starting point and an ending point.

Continuous: Has a starting point but no ending point.

**Exploration v/s Exploitation Trade-Off:**

Exploration: It is all about exploring the environment by trying random actions and receiving feedback/returns/rewards from the environment.

Exploitation: It is about exploiting what we know about the environment to gain maximum rewards.

Exploration-Exploitation Trade-Off: It balances how much we want to explore the environment and how much we want to exploit what we know about the environment.

**Policy:**

Policy: It is called the agent’s brain. It tells us what action to take, given the state.

Optimal Policy: Policy that maximizes the expected return when an agent acts according to it. It is learned through training.

Policy-based Methods:

* An approach to solving RL problems.
* In this method, the Policy is learned directly.
* Will map each state to the best corresponding action at that state. Or a probability distribution over the set of possible actions at that state.

Value-based Methods:

* Another approach to solving RL problems.
* Here, instead of training a policy, we train a value function that maps each state to the expected value of being in that state.