6/18/2021

**1/ MDP – Markov Decision Processes:**

**-** Elements of MDP:

+ Environment

+ Agent

+ States of environment

+ Actions that Agent can take

+ Rewards that Agents can receive

=> Goals: Agents wants to maximize the cumulative rewards

**Diagram

Description automatically generated**

- Expected Return of the reward of a given time step: It is the agent’s goal to maximize the expected discounted return of rewards

- Policy and Value function:

+ Policy = What is the probability that an agent will select a specific action from a specific state? We say that the agent follow a policy meaning that at time t, under policy pi, the probability of taking action a in stat s is pi(s|a).

+ Value function = How good is a specific action or a specific state for the agent? Function of state or state action pair that measure how good it is for an agent to be in a given state or how good it is for agent to take action in a given state. => Return the “expected return”

**Text

Description automatically generated**

**Text

Description automatically generated**

**2/ What do the RL Algorithms Learn?** **- Optimal Policies**