Mathematical Foundations of Reinforcement Learning

Chapter 1: Basic Concepts

Minseok Seo
Artificial Intelligence Graduate School
Gwangju Institute of Science and Technology (GIST)

July 16, 2025

Overview

1. Markov Decision Process (MDPs)

Markov Decision Process (MDPs)

Sets:

- State space: the set of all states, denoted as $\mathcal{S}$.
- Action space: a set of actions, denoted as $\mathcal{A}(s)$, associated with each state $s \in \mathcal{S}$.
- Reward set: a set of rewards, denoted as $\mathcal{R}(s, a)$, associated with each state-action pair (s, a).

Markov Decision Process (MDPs)

Models:

- State transition probability:
  In state s, when taking action a, the probability of transitioning to state s' is $p(s'|s,a)$.
  It holds that $\sum_{s' \in \mathcal{S}} p(s'|s,a) = 1$ for any (s,a).
- Reward probability:
  In state s, when taking action a, the probability of obtaining reward r is $p(r|s,a)$.
  It holds that $\sum_{r \in \mathcal{R}(s,a)} p(r|s,a) = 1$ for any (s,a).

Markov Decision Process (MDPs)

Policy:

- In state s, the probability of choosing action a is $\pi(a|s)$.
- It holds that $\sum_{a \in \mathcal{A}(s)} \pi(a|s) = 1$ for any $s \in \mathcal{S}$.

Markov Decision Process (MDPs)

Markov property:
The Markov property refers to the memoryless property of a stochastic process.
Mathematically, it means that

$$\begin{aligned} p(s_{t+1}|s_t, a_t, s_{t-1}, a_{t-1}, \cdots, s_0, a_0) &= p(s_{t+1}|s_t, a_t) \\ p(r_{t+1}|s_t, a_t, s_{t-1}, a_{t-1}, \cdots, s_0, a_0) &= p(r_{t+1}|s_t, a_t) \end{aligned} \tag{1}$$

Eq. (1) indicates that the next state or reward depends merely on the current state and action and is independet of the previous ones.