HT 2020

CS234 - Reinforcement Learning

Jiaming (George) Yu
jiaming.yu@jesus.ox.ac.uk

August 1, 2022

Contents

Part I

Lecture 1

1 Reinforcement Learning

Part II

Given a Model of the World

2 Markov Processes

For now, we only consider Markov processes with finite states. We first consider Markov reward processes[1], which do not involve actions

[1] we will only consider finite Markov processes as of right now

Definition 2.1

A *Markov Reward Process* (MRP) is a 4-tuple (S, P, R, γ) where

(i) S is the (finite) set of possible states

(ii) P is the transition model

(iii) R is the reward function such that $R(s_t = s) = \mathbb{E}[r_t \mid s_t = s]$

(iv) $\gamma \in [0, 1]$ is the discount factor

Specifically, the transition (a.k.a. dynamics) model specifies $\mathbb{P}(s_{t+1} = s' \mid s_t = s)$ and can be deterministic or stochastic.

1. auhdiaush daisuhd iaus dhia usdiasidhu as
2. vbhnasdfiuhasid uhai sduh asudhaius hdaisu
3. iuhadis uhasid asuidhasid

we only consider finite cases

Definition 2.2

The *horizon* of an MRP is the number of timesteps in each episode (could potentially be infinite). When an MRP has a finite horizon, we call it a *finite MRP*.

Definition 2.3

The *return* G_t of an MRP from timestep t is the discounted sum of rewards from that timestep to horizon:

$$G_t = \sum_{k=0} \gamma^k r_{t+k} = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots$$

Definition 2.4

The *state value function* $V(S)$ of an MRP gives the expected return starting in state s:

$$V(s) = \mathbb{E}[G_t \mid s_t = s]$$

Obtaining $V(S)$ For a finite-state MRP

Next, we define the Markov decision process, which are essentially the MRP with actions. Formally:

Definition 2.5

A *Markov Decision Process* (MDP) is a 5-tuple (S, A, P, R, γ) where

(i) S is the (finite) set of possible states

(ii) A is the (finite) set of possible actions

(iii) P is the transition model

(iv) R is the reward function such that $R(s_t = s) = \mathbb{E}[r_t \mid s_t = s]$

(v) $\gamma \in [0, 1]$ is the discount factor

Similar to MRP, P specifies $\mathbb{P}(s_{t+1} = s' \mid s_t = s, a_t = a)$, and $R : S \times A \to \mathbb{R}$ is defined by $R(s_t = s, a_t = a) = \mathbb{E}[r_t \mid s_t = s, a_t = a]$