Note for Reinforcement Learning

Chenmi'en Tan
E-mail: chenmientan@outlook.com
Actively Updating (Last update: June 23, 2021)

Contents

1 Markov Decision Process

In a Markov decision process, the state and reward in the next episode is determined (stochastically) by the current state and action. By denoting $\mathcal{S}, \mathcal{R}, \mathcal{A}$ as the sets of states, rewards, and actions, the dynamic $p : \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$ is defined as

$$p(s', r|s, a) = \mathbb{P}(S_t = s', R_t = r|S_{t-1} = s, A_{t-1} = a)$$

The purpose of our agent is to maximize the accumulative reward through adopting proper policy. Denote n as the number of episodes, the objective function can be expressed as $G(a) = \mathbb{E}[\sum_{t=1}^{n} R_t]$