1.041/1.200 Spring 2024: Recitation 8

Date: Apr 8, 2:00 PM

1 Problem 1 : A Simple Chain

We define an infinite horizon discounted MDP in the following manner. There are three states s_0, s_1, s_2 and one action a. The MDP dynamics are independent of the action a as shown below:

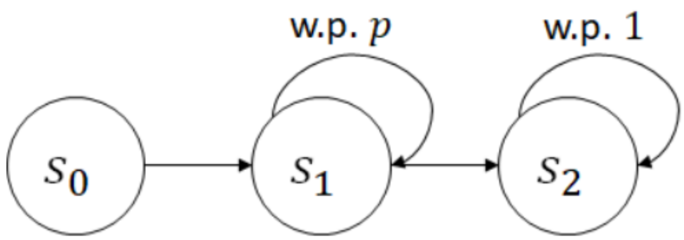

Figure 1

At state s_0, with probability 1 the state transits to s_1, i.e.,

$$p(s_1 \mid s_0) = 1$$

Then at state s_1, we have

$$p(s_1 \mid s_1) = p, \quad p(s_2 \mid s_1) = 1 - p$$

which says there is probability p we stay in s_1 and probability $1 - p$ the state transits to s_2. Finally, state s_2 is the absorbing state so that

$$p(s_2 \mid s_2) = 1$$

The instant reward is set to 1 for staying at state s_1 and 0 elsewhere: (the reward only depends on the current state, and does not depend on the action)

$$r(s_1) = 1, \quad r(s_0) = r(s_2) = 0$$

The discount factor γ satisfies $0 < \gamma < 1$

1. Using the optimal Bellman equation, compute $V^*(s_1)$.
2. Compute $Q^*(s_0, a)$.

2 Go through the CL3 codes