Mario Exercise

Design a Super Mario Bros MDP
- Reward function R
- Discount factor γ

Your states are: eat mushroom, collect coins, die, game over

Compute discounted return for:
- Eat mushroom at $t = 10$
- Collect coins at $t = 11, 12$
- Die to bowser at $t = 20$
- Game over screen at $t = 21...\infty$
- $r = 0$ for other timesteps

Answer

We have the states:

$$S = \{\text{mushroom}, \text{coin}, \text{die}, \text{game over}, 0\}$$

You should define some scalar reward for each, this is up to you to decide, but you probably want game over to be zero so your return is finite

$$R(\text{mushroom}) = 2, R(\text{coin}) = 1, R(\text{die}) = -5, R(\text{game over}) = 0$$

You can choose any γ between 0 and 1

$$\gamma = 0.9$$

Then, the return is

$$\begin{aligned} G &= \gamma^{10} R_{\text{mushroom}} + \gamma^{11} R_{\text{coin}} + \gamma^{12} R_{\text{coin}} + \gamma^{20} R_{\text{die}} \\ G &= 0.9^{10} \cdot 2 + 0.9^{11} \cdot 1 + 0.9^{12} \cdot 1 + 0.9^{20} \cdot -5 \\ G &\approx 0.686 \end{aligned}$$

Markov Process Exercise

Design an Markov process about a problem you care about

- 4 states
- State transition function $\mathrm{Tr} = \Pr(s_{t+1} \mid s_t)$ for all $s_t, s_{t+1} \in S$
- Create a terminal state
- Given a starting state s_0, what will your state distribution be for s_2?

$$\Pr(s_n \mid s_0) = \sum_{s_1, s_2, \ldots s_{n-1} \in S} \prod_{t=0}^{n-1} \Pr(s_{t+1} \mid s_t)$$

Answer

States:

$$S = \{S_a, S_b, S_c, S_d\}$$

S_d is a terminal state

Transition Function (Tr):

You can come up with whatever state transition function you want, as long as each section sums to one

From s_a:

$$\Pr(s_a \mid s_a) = 0.5$$
$$\Pr(s_b \mid s_a) = 0.3$$
$$\Pr(s_c \mid s_a) = 0.1$$
$$\Pr(s_d \mid s_a) = 0.1$$

From s_b:

$$\Pr(s_a \mid s_b) = 0.2$$
$$\Pr(s_b \mid s_b) = 0.6$$
$$\Pr(s_c \mid s_b) = 0.1$$
$$\Pr(s_d \mid s_b) = 0.1$$

From s_c:

$$\Pr(s_a \mid s_c) = 0.1$$
$$\Pr(s_b \mid s_c) = 0.1$$
$$\Pr(s_c \mid s_c) = 0.7$$
$$\Pr(s_d \mid s_c) = 0.1$$

From s_d (terminal):

$$\Pr(s_a \mid s_d) = 0$$
$$\Pr(s_b \mid s_d) = 0$$
$$\Pr(s_c \mid s_d) = 0$$
$$\Pr(s_d \mid s_d) = 1.0$$

Roll the Process Forward

Compute $\Pr(s_2 \mid s_0 = s_a)$ by summing over all paths through intermediate states s_1. You can either compute all conditional probabilities at each timestep, or you can just enumerate all possible paths from s_0 to s_2. In this case, I choose the latter.

For $s_2 = s_a$:

$$\begin{aligned}
&\Pr(s_a \to s_a \to s_a) + \Pr(s_a \to s_b \to s_a) + \Pr(s_a \to s_c \to s_a) + \Pr(s_a \to s_d \to s_a) \\
&= (0.5 * 0.5) + (0.3 * 0.2) + (0.1 * 0.1) + (0.1 * 0) \\
&= 0.25 + 0.06 + 0.01 + 0 = 0.32
\end{aligned}$$

For $s_2 = s_b$:

$$\begin{aligned}
&\Pr(s_a \to s_a \to s_b) + \Pr(s_a \to s_b \to s_b) + \Pr(s_a \to s_c \to s_b) + \Pr(s_a \to s_d \to s_b) \\
&= (0.5 * 0.3) + (0.3 * 0.6) + (0.1 * 0.1) + (0.1 * 0) \\
&= 0.15 + 0.18 + 0.01 + 0 = 0.34
\end{aligned}$$

For $s_2 = s_c$:

$$\begin{aligned}&\Pr(s_a \to s_a \to s_c) + \Pr(s_a \to s_b \to s_c) + \Pr(s_a \to s_c \to s_c) + \Pr(s_a \to s_d \to s_c)\\&= (0.5 * 0.1) + (0.3 * 0.1) + (0.1 * 0.7) + (0.1 * 0)\\&= 0.05 + 0.03 + 0.07 + 0 = 0.15\end{aligned}$$

For $s_2 = s_d$:

$$\begin{aligned}&\Pr(s_a \to s_a \to s_d) + \Pr(s_a \to s_b \to s_d) + \Pr(s_a \to s_c \to s_d) + \Pr(s_a \to s_d \to s_d)\\&= (0.5 * 0.1) + (0.3 * 0.1) + (0.1 * 0.1) + (0.1 * 1.0)\\&= 0.05 + 0.03 + 0.01 + 0.10 = 0.19\end{aligned}$$

Final Distribution for s_2:

Make sure this sums to one

$$\Pr(s_2 = s_a \mid s_0 = s_a) = 0.32$$
$$\Pr(s_2 = s_b \mid s_0 = s_a) = 0.34$$
$$\Pr(s_2 = s_c \mid s_0 = s_a) = 0.15$$
$$\Pr(s_2 = s_d \mid s_0 = s_a) = 0.19$$