Homework 8

Submission: You **do not** need to submit your solutions. This homework is not marked. It is for your own practice. But you need to feel comfortable with its content.

Spend or Save

You are facing a small crisis where you have to decide whether you should spend your money or save it in your bank account. If the balance of your bank account is low, you can only buy something that gives you a little bit of joy. If you save your money, your account's balance increases, but you actually feel a bit sad. Later in the future, however, you can buy something more expensive that gives you more joy. Your dilemma is what the optimal decision should be depending on your current financial state (low or high).

This problem can be formulated as a discounted MDP where you have two states s_1, which corresponds to low amount of money in your bank account, and s_2, which corresponds to having a lot of money in your bank account. At each state you have two actions:

- a_1: Save money
- a_2: Spend money

Depending on the current state and the selected action, your financial state in the next time step might change. You also receive some amount of reward, which is an indicator of your level of joy.

To formulate this problem as a discounted MDP, we have to define transition probabilities, reward function, and the discount factor. The discount factor $0 \leq \gamma < 1$ determines how myopic/farsighted you are.[1] Figure 1 describes the dynamics and the reward. In this figure, all the transitions are deterministic, e.g., if you are at state s_1 and choose action a_1, you will definitely move to state s_2. In mathematical term, $\mathcal{P}(s_2|s_1, a_1) = 1$.

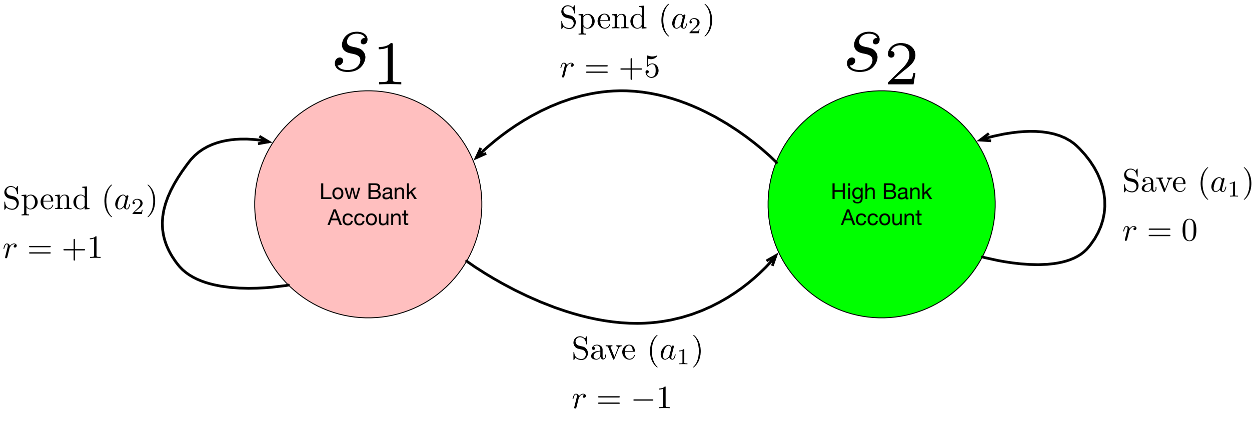

Figure 1: Spend or Save

The purpose of this question is to make you feel comfortable solving simple MDPs. Answer the following questions:

[1] This is a very crude modelling of a real-world situation. This is not an exercise in modelling, but an exercise in solving simple planning problems.

1. Write down $\mathcal{P}(s'|s,a)$ for all $s, s' = \{s_1, s_2\}$ and $a \in \{a_1, a_2\}$.

2. Consider the following two policies:

   - $\pi_{\text{save}}(s) = a_1$
   - $\pi_{\text{spend}}(s) = a_2$

   Write down the Bellman equations for $Q^{\pi_{\text{save}}}(s,a)$ and $Q^{\pi_{\text{spend}}}(s,a)$ for all $(s,a) \in \mathcal{S} \times \mathcal{A}$.

3. Bellman equation defines a linear system of equations. Solve them in order to find $Q^{\pi_{\text{save}}}$ and $Q^{\pi_{\text{spend}}}$. Your answer depends on γ.

4. Write a simple program that computes the *optimal* action-value function and *optimal* policy for a given γ.

5. How does the choice of γ affect your optimal policy?

Solutions

The Transition Probability $\mathcal{P}(s'|s,a)$.

$$\begin{aligned} \mathcal{P}(s_1|s_1,a_1) &= 0 \qquad & \mathcal{P}(s_2|s_1,a_1) &= 1 \\ \mathcal{P}(s_1|s_2,a_1) &= 0 \qquad & \mathcal{P}(s_2|s_2,a_1) &= 1 \\ \mathcal{P}(s_1|s_2,a_2) &= 1 \qquad & \mathcal{P}(s_2|s_2,a_2) &= 0 \\ \mathcal{P}(s_1|s_1,a_2) &= 1 \qquad & \mathcal{P}(s_2|s_1,a_2) &= 0 \end{aligned}$$

The Bellman Equation for $Q^{\pi_{\textbf{save}}}$ and its Solution. We have the following Bellman equations:

$$\begin{aligned} Q^{\pi_{\text{save}}}(s_1,a_1) &= -1 + \gamma Q^{\pi_{\text{save}}}(s_2,a_1) \\ Q^{\pi_{\text{save}}}(s_1,a_2) &= +1 + \gamma Q^{\pi_{\text{save}}}(s_1,a_1) \\ Q^{\pi_{\text{save}}}(s_2,a_1) &= 0 + \gamma Q^{\pi_{\text{save}}}(s_2,a_1) \\ Q^{\pi_{\text{save}}}(s_2,a_2) &= +5 + \gamma Q^{\pi_{\text{save}}}(s_1,a_1) \end{aligned}$$

We could use a computer to solve these. But by close inspection of equations and some re-substitutions, we see that $Q^{\pi_{\text{save}}}(s_2,a_1) = 0$. So $Q^{\pi_{\text{save}}}(s_1,a_1) = -1$. As a result, $Q^{\pi_{\text{save}}}(s_1,a_2) = 1 - \gamma$. And finally, $Q^{\pi_{\text{save}}}(s_2,a_2) = 5 - \gamma$.

The Bellman Equation for $Q^{\pi_{\textbf{spend}}}$ and its Solution. We have the following Bellman equations:

$$\begin{aligned} Q^{\pi_{\text{spend}}}(s_1,a_1) &= -1 + \gamma Q^{\pi_{\text{spend}}}(s_2,a_2) \\ Q^{\pi_{\text{spend}}}(s_1,a_2) &= +1 + \gamma Q^{\pi_{\text{spend}}}(s_1,a_2) \\ Q^{\pi_{\text{spend}}}(s_2,a_1) &= 0 + \gamma Q^{\pi_{\text{spend}}}(s_2,a_2) \\ Q^{\pi_{\text{spend}}}(s_2,a_2) &= +5 + \gamma Q^{\pi_{\text{spend}}}(s_1,a_2) \end{aligned}$$

We can solve these linear system of equations similarly. We observe that

$$Q^{\pi_{\text{spend}}}(s_1,a_2) = \frac{1}{1-\gamma}.$$

So

$$Q^{\pi_{\text{spend}}}(s_2,a_2) = 5 + \frac{\gamma}{1-\gamma},$$

and

$$Q^{\pi_{\text{spend}}}(s_1,a_1) = -1 + \gamma \left[5 + \frac{\gamma}{1-\gamma} \right].$$

Finally,

$$Q^{\pi_{\text{spend}}}(s_2,a_1) = \gamma \left[5 + \frac{\gamma}{1-\gamma} \right].$$

Optimal Value Function and the Effect of Discount Factor. Refer to the code.