Markov Decision Processes

Policies

Recall that the reward for entering an empty cell is -1, a mountainous cell -3, the pond -50, and the goal +100. These are the rewards defined according to the environment. However, if our robot wanted to move from one cell to another, it it not guaranteed to succeed. Therefore, we must calculate the expected reward, which takes into account not just the rewards set by the environment, but the robot's transition model too.

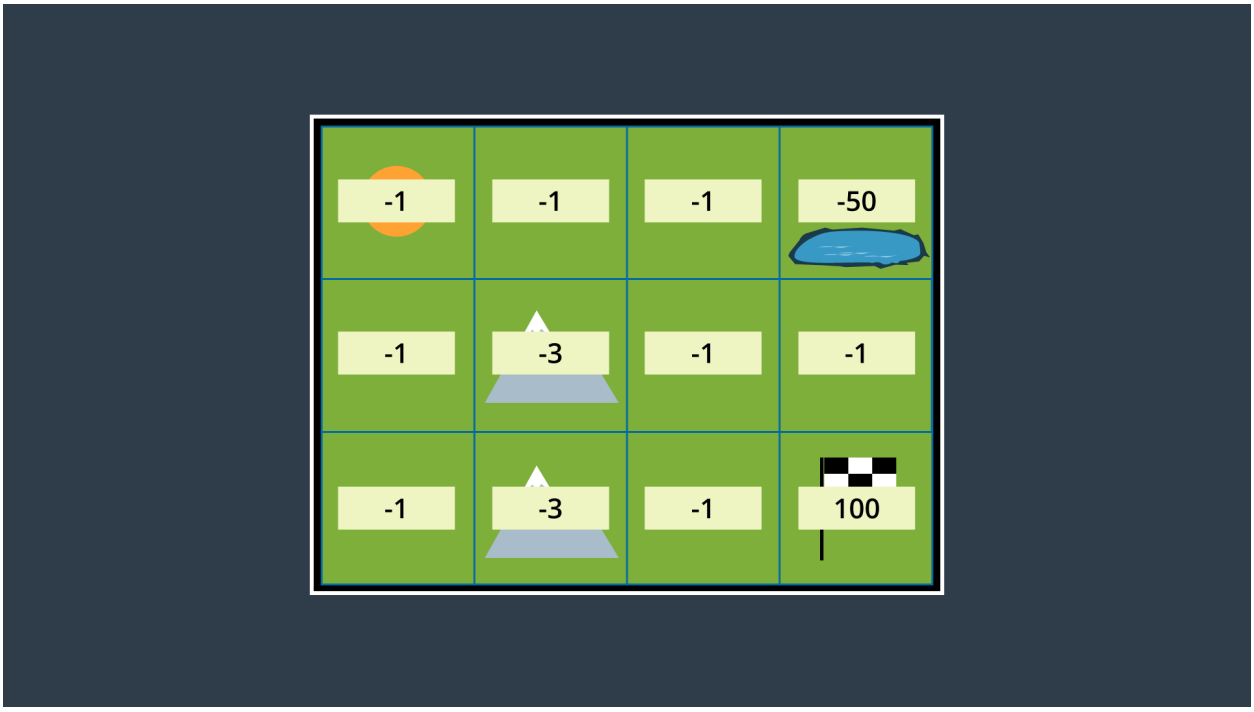

Figure 1: Rewards

1. What is the expected reward for moving from the bottom-left cell to the cell above it?

$$\begin{aligned} expected\,reward &= 0.8 * (-1) + 0.1 * (-1) + 0.1 * (-3) \\ &= -1.2 \end{aligned}$$

2. What is the expected reward for moving from the empty cell on the right to the goal,one cell below it?

$$\begin{aligned} expected\,reward &= 0.8 * (100) + 0.1 * (-1) + 0.1 * (-1) \\ &= 79.8 \end{aligned}$$

3. What is the expected reward for moving from the empty cell on the right to the cell to its left?

$$\begin{aligned} expected\,reward &= 0.8 * (-1) + 0.1 * (-50) + 0.1 * (100) \\ &= 4.2 \end{aligned}$$

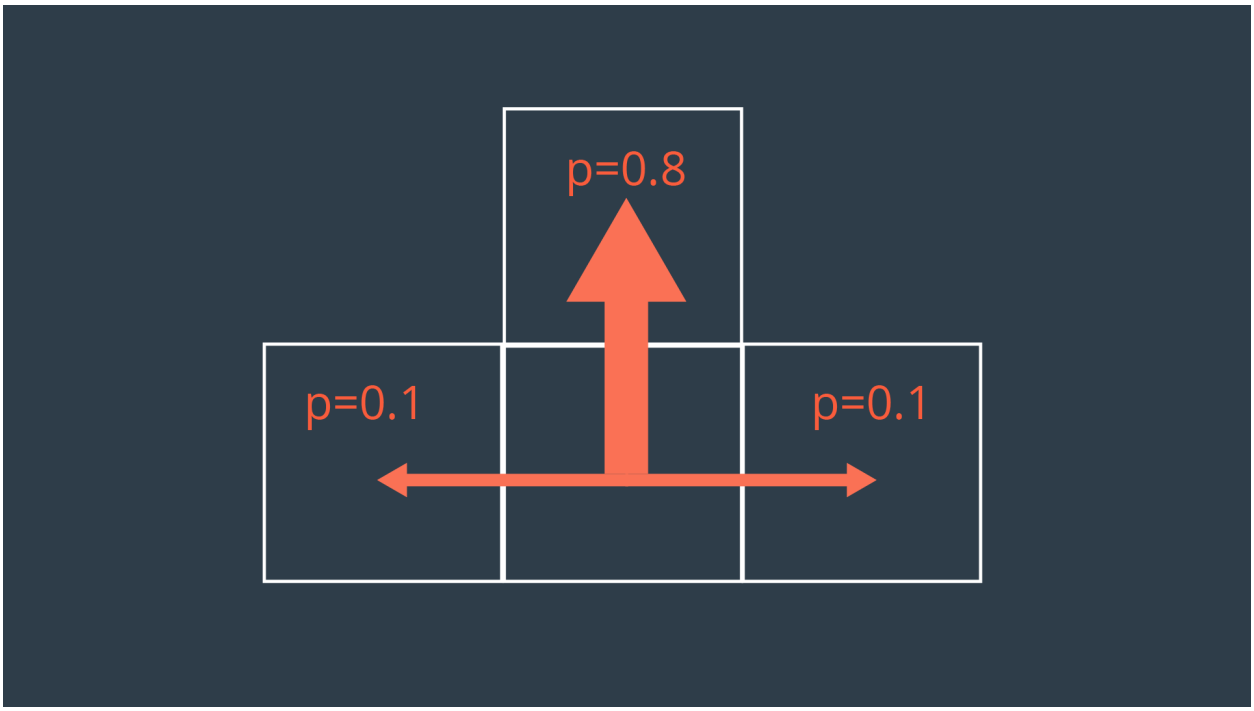

Figure 2: Robot Transition Model

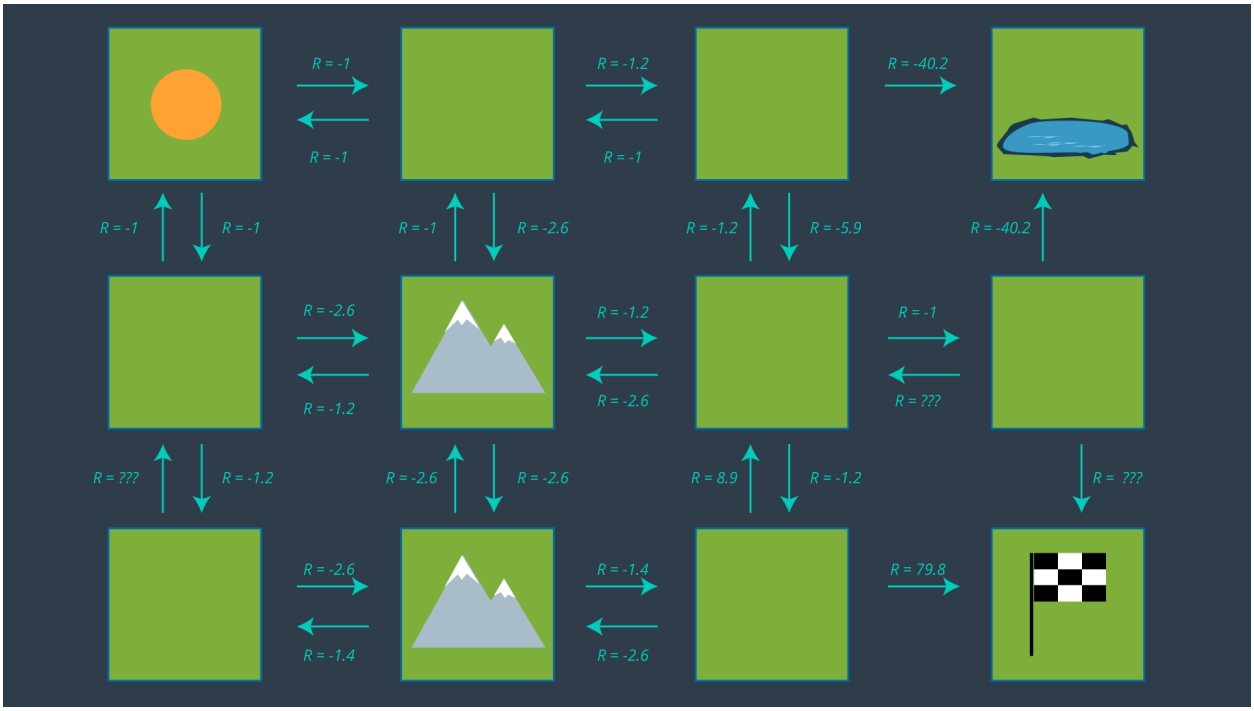

Figure 3: Expected Rewards

State Utility

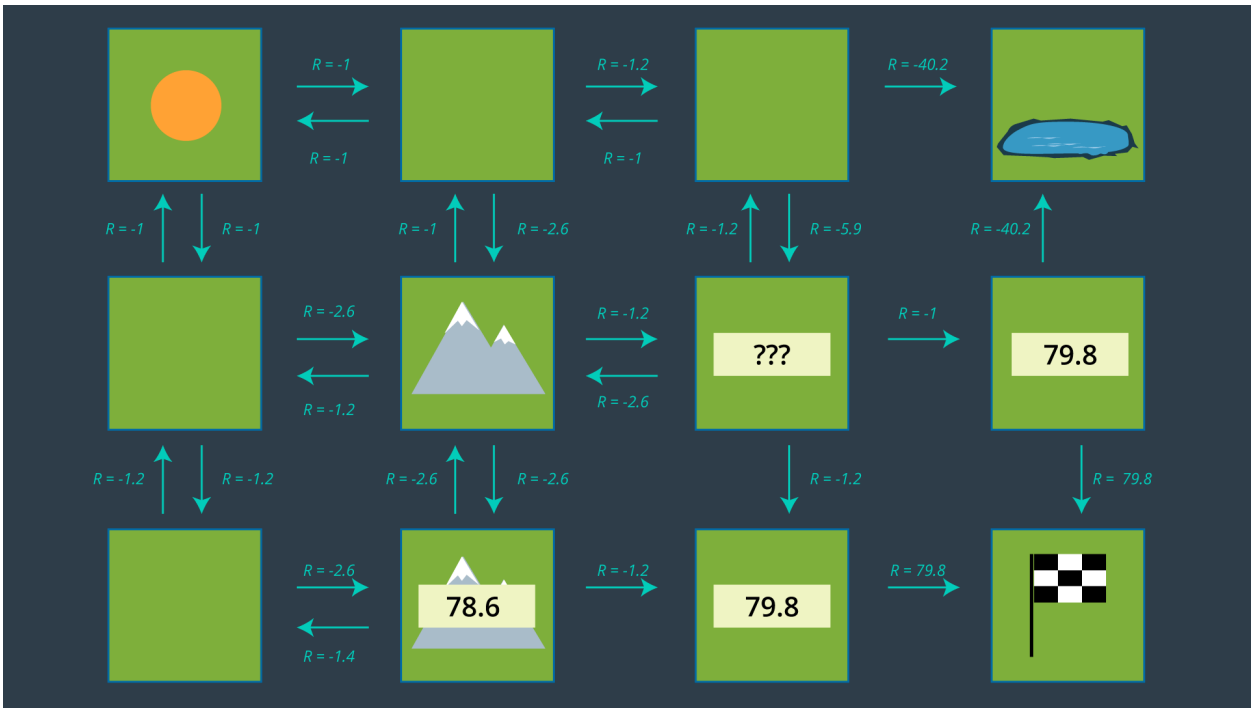

Figure 4: Utility

1. What is the utility of the state to the right of the center mountain following the optimal policy?

$$U^{\pi}(s) = 78.8$$