Homework 2. Markov decision problems

Consider as in HW1 the kids' game Insey-Winsey-Spider.

The spiders have several levels to climb (corresponding to steps in a ladder) and want to reach the top level.

At each turn, the player can decide to go and they throw a die. After that, the player has the possibility to climb a number of steps. But they will only go up if it is a sunny day, if it is a rainy day then they go back to level 0. For this, they turn an arrow and see where it lands (let's assume that the rainy part corresponds to 20% of the area).

To simplify, let's consider that there are only 6 steps in the ladder (0-5) and we are using a two-sided dice (so you can go up 1 or 2 steps).

After reaching level 5, the game continues, but it is no longer possible to move up. For level 4, both either a 1 or 2 on the die will make it jump to level 5 (no need to get the correct value).

We have the transition matrix for action play according to these rules,

$$\boldsymbol{P}^{play} = \begin{bmatrix} .2 & .4 & .4 & 0 & 0 & 0 \\ .2 & 0 & .4 & .4 & 0 & 0 \\ .2 & 0 & 0 & .4 & .4 & 0 \\ .2 & 0 & 0 & 0 & .4 & .4 \\ .2 & 0 & 0 & 0 & 0 & .8 \\ .2 & 0 & 0 & 0 & 0 & .8 \end{bmatrix}.$$

We will make a game expansion inspired in the game *Can't Stop*. Now, the spiders have another action "Stop" that allows them to build a web, but they can only build it at state 2, the weather does not have an impact. So the action only works at state 2. If they do the action at another level, they will fall. If there is a web, they will fall to the web, or to level 0 if below the web. If there is no web, they fall to the ground. After reaching the last level (level 5 corresponding to state 5 if no web, or state 11 with web), any action will give cost 0 and the spider will not move (the game stops). Any other play action has a cost of 1. Action "stop" has no cost except at the ground or at the web. Building the web has zero cost as it has the potential to get yummy insects. Consider a $\gamma = 0.9$.

Exercise

(a) Compute the new transition matrices, and define all the components of the MDP. If you need to define more states, the new states should be included after the previous ones.

(b) Compute the cost-to-go for the two deterministic policies in state 2, assuming that the policy is optimal in the remaining states.

(c) What is the optimal policy in state 2? Justify and comment.

Solution:

Note that the questions were a bit ambiguous and the last level could be interpreted as stopping the game (and the transition would always be 1 as indicated), but could also be interpreted as continuing. We will accept both solutions.

(a) The MDP is specified as a tuple $(\mathcal{X}, \mathcal{A}, \boldsymbol{P}, C, \gamma)$. We have the state space

$$\mathcal{X} = \{0, 1, 2, 3, 4, 5, 0w, 1w, 2w, 3w, 4w, 5w\},$$

and the action space

$$\mathcal{A} = \{play, stop\}.$$

The new states correspond to the situation where there is a web at level 2. The transition matrices associated with each of the actions are

$$\boldsymbol{P}^{play} = \begin{bmatrix}
.2 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.2 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.2 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.2 & 0 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\
.2 & 0 & 0 & 0 & 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & .8 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

$$\boldsymbol{P}^{stop} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The cost function is

$$C^{play} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$
$$C^{stop} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

The discount factor $\gamma = 0.9$.

Alternative solution transition matrices:

$$\boldsymbol{P}^{play-alt} = \begin{bmatrix} .2 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & .4 & .4 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\ .2 & 0 & 0 & 0 & 0 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & .4 & .4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & .4 & .4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & .8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .2 & 0 & 0 & .8 \end{bmatrix}$$

$$\boldsymbol{P}^{stop-alt} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

(b) We note that the action "Stop" is only useful in state 2, after that it just makes the spider fall. So we just need to consider what happens if we "Stop" to make a web in state 2 and then always "Play", or do not make a web and always "Play".

Since the game stops when the spider is at the top and the cost is zero, we know that $V(5) = V(11) = 0$

The case we did "Stop" and so we have a web,

$$\begin{aligned} V(10) &= 1 + \gamma(.8V(11) + .2V(8)) \\ V(9) &= 1 + \gamma(.4V(10) + .4V(11) + .2V(8)) \\ V(8) &= 1 + \gamma(.4V(9) + .4V(10) + .2V(8)) \end{aligned}$$

we can solve this using a system of equations.

$$\begin{aligned} V(10) - \gamma(.8V(11) + .2V(8)) &= 1 \\ V(9) - \gamma(.4V(10) + .4V(11) + .2V(8)) &= 1 \\ V(8) - \gamma(.4V(9) + .4V(10) + .2V(8)) &= 1 \end{aligned}$$

$$
\begin{aligned}
V(10) &= 1.50 \\
V(9) &= 2.04 \\
V(8) &= 2.77 \\
Q(2, stop) &= 0 + \gamma V(8) = 2.49
\end{aligned}
$$

The case where we didn't do the action "Stop" is more difficult, as we will fall to the ground,

$$
\begin{aligned}
V(4) &= 1 + \gamma(.8V(5) + .2V(0)) \\
V(3) &= 1 + \gamma(.4V(4) + .4V(5) + .2V(0)) \\
V(2) &= 1 + \gamma(.4V(3) + .4V(4) + .2V(0)) \\
V(1) &= 1 + \gamma(.4V(2) + .4V(3) + .2V(0)) \\
V(0) &= 1 + \gamma(.4V(1) + .4V(2) + .2V(0))
\end{aligned}
$$

$$
\begin{aligned}
V(4) &= 1.78 \\
V(3) &= 2.43 \\
Q(2, play) &= 3.30 \\
V(1) &= 3.85 \\
V(0) &= 4.36
\end{aligned}
$$

Answer:

$$
\begin{aligned}
Q(2, stop) &= 2.49 \\
Q(2, play) &= 3.30
\end{aligned}
$$

Alternative solution (with alternative transition matrices):

In this interpretation of the text, the game no longer stops when at the last level, which means that for states in the last level we have:

$$
\begin{aligned}
V(5) &= 0 + \gamma(.2V(0) + .8V(5)) \\
V(11) &= 0 + \gamma(.2V(8) + .8V(11))
\end{aligned}
$$

For the case where we did "Stop" and so we have a web:

$$
\begin{aligned}
V(10) &= 1 + \gamma(.2V(8) + .8V(11)) \\
V(9) &= 1 + \gamma(.2V(8) + .4V(10) + .4V(11)) \\
V(8) &= 1 + \gamma(.2V(8) + .4V(9) + .4V(10))
\end{aligned}
$$

we can solve this using a system of equations.

$$
\begin{aligned}
V(11) - \gamma(.2V(8) + .8V(11)) &= 0 \\
V(10) - \gamma(.2V(8) + .8V(11)) &= 1 \\
V(9) - \gamma(.2V(8) + .4V(10) + .4V(11)) &= 1 \\
V(8) - \gamma(.2V(8) + .4V(9) + .4V(10)) &= 1
\end{aligned}
$$

$$
\begin{aligned}
V(11) &= 3.33 \\
V(10) &= 4.33 \\
V(9) &= 4.69 \\
V(8) &= 5.18 \\
Q(2, stop) &= 0 + \gamma V(8) = 4.66
\end{aligned}
$$

For the case where we didn't do the action "Stop" is more difficult, as we will fall to the ground:

$$
\begin{aligned}
V(4) &= 1 + \gamma(.2V(0) + .8V(5)) \\
V(3) &= 1 + \gamma(.2V(0) + .4V(4) + .4V(5)) \\
V(2) &= 1 + \gamma(.2V(0) + .4V(3) + .4V(4)) \\
V(1) &= 1 + \gamma(.2V(0) + .4V(2) + .4V(3)) \\
V(0) &= 1 + \gamma(.2V(0) + .4V(1) + .4V(2))
\end{aligned}
$$

we can solve this using a system of equations.

$$
\begin{aligned}
V(5) - \gamma(.2V(0) + .8V(5)) &= 0 \\
V(4) - \gamma(.2V(0) + .8V(5)) &= 1 \\
V(3) - \gamma(.2V(0) + .4V(4) + .4V(5)) &= 1 \\
V(2) - \gamma(.2V(0) + .4V(3) + .4V(4)) &= 1 \\
V(1) - \gamma(.2V(0) + .4V(2) + .4V(3)) &= 1 \\
V(0) - \gamma(.2V(0) + .4V(1) + .4V(2)) &= 1
\end{aligned}
$$

$$
\begin{aligned}
V(5) &= 4.40 \\
V(4) &= 5.40 \\
V(3) &= 5.76 \\
Q(2, play) &= V(2) = 6.24 \\
V(1) &= 6.55 \\
V(0) &= 6.84
\end{aligned}
$$

Answer:

$$
\begin{aligned}
Q(2, stop) &= 4.66 \\
Q(2, play) &= 6.24
\end{aligned}
$$

(c) The best choice is to build the web. Of course, if the cost were also 1, then the conclusions would be different. If there were a higher probability of rain, then building the web would also be better.