TÉCNICO LISBOA

MSc in Computer Science and Engineering

Learning and Decision Making 2016-2017

Homework 3. Partially observable Markov decision problems

Consider once again the problem of the knight that must save the captive princess.[1] After arriving at the castle where the princess is held captive, the knight realizes that the princess may be in any one of two towers (towers A and B). The knight must thus decide which tower to invade. However, he does not know which tower the princess is at. He can try to peer at the towers with a scope, in which case he is able to detect the princess with a probability of 0.9. However, with a 0.1 probability, the knight will see the princess in the wrong tower.

When the knight invades a tower, he either saves the princess or is expelled from those lands. For the purposes of this homework, whichever the outcome of the knight's invasion, we assume that the world resets—i.e., the princess is captured again, placed randomly in one of the two towers, and the knight must again rescue her.

In this homework, you will model the decision of the knight as a partially observable MDP (POMDP).

Exercise 1.

(a) Identify the state space, $\mathcal{X}$, the action space $\mathcal{A}$, and the observation space, $\mathcal{Z}$. You should explicitly model the fact that, when the knight does not peer into the towers with his scope, he sees *nothing*.

(b) Write down the transition probabilities, the observation probabilities and the cost function for this problem. Make sure that the values in your cost function all lie in the interval $[0, 1]$.

(c) Suppose that, at some time step t, the knight believes that the princess is in Tower A with a probability 0.7, decides to peer at the tower, and observes the princess in Tower B. Compute the resulting belief.

[1] Aside from the common theme, there is no other relation between this homework and the previous one, where the knight must navigate the grid.

Solution 1:

(a) The state corresponds to the relevant information for the decision of the knight—in this case, the position of the princess. We have

$$\mathcal{X} = \{A, B\}\,.$$

The knight has 3 actions available: invading Tower A, invading tower B, and peering at the towers with the scope. We thus represent the action space as the set $\mathcal{A} = \{I_A, I_B, P\}$, where I_A corresponds to invading Tower A, I_B corresponds to invading Tower B, and P corresponds to peering at the towers. Finally, the observation space is $\mathcal{Z} = \{A, B, \emptyset\}$, where $\emptyset$ corresponds to the no-observation situation.

(b) The transition probabilities come:

$$\boldsymbol{P}_{I_A} = \boldsymbol{P}_{I_B} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}, \qquad \boldsymbol{P}_P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The observation probabilities, in turn, come:

$$\boldsymbol{O}_{I_A} = \boldsymbol{O}_{I_B} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{O}_P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.9 & 0 \end{bmatrix}.$$

As for the cost function, the case where the knight is captured is the most costly (so we set $c = 1$), and the case where the knight rescues the princess is the least costly (so we set $c = 0$). The peeking action should lie somewhere in the middle. We thus set

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 0.5 \\ 1 & 0 & 0.5 \end{bmatrix}.$$

(c) Using the belief update rule, we get:

$$\boldsymbol{b}_{\text{new}} = \xi \boldsymbol{b}_{\text{old}} \boldsymbol{P}_P \,\text{diag}(\boldsymbol{O}_P) = \xi \begin{bmatrix} 0.7 & 0.3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.1 & 0 \\ 0 & 0.9 \end{bmatrix} = \begin{bmatrix} 0.2059 & 0.7941 \end{bmatrix},$$

where ξ is the normalization constant.