Homework 3. Partially observable Markov decision problems

Consider the following guessing game. An agent is blindfolded and faced with an opponent that has one of two cards in hand: an Ace of Clubs ($A\clubsuit$) and an Ace of Diamonds ($A\diamondsuit$). The agent must guess which card the opponent is holding. For every right answer, the agent wins 1EUR, and every wrong answer costs the agent 1EUR.

Since the agent is blindfolded, it cannot know which of the two cards the opponent has. However, it can try to *peek*, in which case the agent is able to see which card the opponent holds with a probability of 0.9. However, with a 0.1 probability, the agent will see the wrong card. Every time the agent makes a guess, it receives the corresponding prize or pays the corresponding fee, depending on whether the guess is right or not. The opponent then randomly selects a new card from the two allowed, and the game restarts.

In this homework, you will model the decision of the agent 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 agent does not peek, it 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 $[0, 1]$, while respecting the value-relation between actions induced by the rules of the game.

(c) Suppose that, at some time step t, the agent believes that the opponent has the ace of clubs ($A\clubsuit$) with a probability 0.7, decides to peek and observes an ace of diamonds ($A\diamondsuit$). Compute the resulting belief.

Solution 1:

(a) The state corresponds to the relevant information for the decision of the agent—in this case, the card held by the opponent. We have

$$\mathcal{X} = \{A\clubsuit, A\diamondsuit\}.$$

The action space is $\mathcal{A} = \{A\clubsuit, A\diamondsuit, P\}$, where $A\clubsuit$ corresponds to the guess "$A\clubsuit$", $A\diamondsuit$ corresponds to the guess "$A\diamondsuit$", and P corresponds to "peeking". Finally, the observation space is $\mathcal{Z} = \{A\clubsuit, A\diamondsuit, \emptyset\}$, where $\emptyset$ corresponds to the no-observation situation.

(b) The transition probabilities come:

$$\boldsymbol{P}_{A\clubsuit} = \boldsymbol{P}_{A\diamondsuit} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}, \qquad \boldsymbol{P}_P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The observation probabilities, in turn, come:

$$\boldsymbol{O}_{A\clubsuit} = \boldsymbol{O}_{A\diamondsuit} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{O}_P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.9 & 0 \end{bmatrix}.$$

As for the cost function, the case where the agent loses 1EUR is the most costly (so we set $c = 1$), and the case where the agent wins 1EUR is the least costly (so we set $c = 0$). Finally, the peeking action does not win or lose the agent any money, so lies exactly 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:

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

where ξ is the normalization constant.