Planing, Learning and Intelligent Decision Making - Homework 3

99326 - Sebastião Carvalho, 99331 - Tiago Antunes

March 11, 2024

Contents

1 Question 1

1 a)

Considering $\mathcal{X}$ as our state space, $\mathcal{X} = \{1, 2a, 2b, 3, 4\}$, corresponding to each of the nodes of the graph.

Our action space is $\mathcal{A} = \{a, b, c\}$.

Since the observation at each step is the number in the state designation, our action space will be, $\mathcal{Z} = \{1, 2, 3, 4\}$,

1 b)

The transition probabilities matrices, given by the edges in the graph are

$$P_a = \begin{array}{c c} & \begin{array}{ccccc} 1 & 2a & 2b & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2a \\ 2b \\ 3 \\ 4 \end{array} & \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

$$P_b = \begin{array}{c} \\ 1 \\ 2a \\ 2b \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2a & 2b & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

$$P_c = \begin{array}{c} \\ 1 \\ 2a \\ 2b \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2a & 2b & 3 & 4 \end{array} \\ \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

The observation probabilities matrices, corresponding to the probability of an observation given the current state and the previous action, do not depend on the action selected, so

$$O_a = O_b = O_c = \begin{array}{c} \\ 1 \\ 2a \\ 2b \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Now the only thing missing is the immediate cost function, which we can obtain from the costs indicated in the edges of the graph.

The cost matrix for all actions is given by $C = \begin{array}{c} \\ 1 \\ 2a \\ 2b \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.5 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$

1 c)

Let's consider $\hat{b}_{t+1}$ as the update for the belief b_t after taking action a_t and before observing z_{t+1}.

We need to calculate $\hat{b}_{t+1} = b_t P_a$, for all actions $a \in \mathcal{A}$.

Considering we choose the action a at the time-step t, we have the following belief update:

$$\hat{b}_{t+1} = b_t P_a = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

Considering we choose the action b at the time-step t, we have the following belief update:

$$\hat{b}_{t+1} = b_t P_b = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Considering we choose the action c at the time-step t, we have the following belief update:

$$\hat{b}_{t+1} = b_t P_c = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$