TÉCNICO LISBOA

Planning, Learning and Decision Making

MSc in Computer Science and Engineering

First test – November 12, 2020

Instructions

- You have 90 minutes to complete the test.
- Make sure that your test has a total of 9 pages and is not missing any sheets, then write your full name and student n. on this page (and your number in all others).
- The test has a total of 5 questions, with a maximum score of 20 points. The questions have different levels of difficulty. The point value of each question is provided next to the question number.
- *If you get stuck in a question, move on.* You should start with the easier questions to secure those points, before moving on to the harder questions.
- *No interaction with the faculty is allowed during the exam.* If you are unclear about a question, clearly indicate it and answer to the best of your ability.
- Please provide your answer in the space below each question. If you make a mess, clearly indicate your answer.
- The exam is open book and open notes. You may use a calculator, but any other type of electronic or communication equipment is not allowed.
- Good luck.

Question 1. (**3 pts.**)

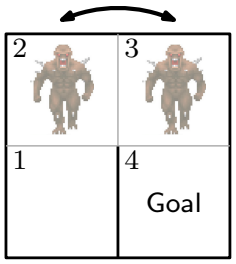


Figure 1: An agent moves in an environment where an Imp also exists. The agent must try to reach the Goal location while avoiding the Imp.

Consider the following problem. An agent moves in the environment depicted in Fig. 1. Moving in the same environment there is an *Imp*—a monster that, if it stands in the same cell as the agent, will inflict a large amount of damage to the agent. The Imp moves between cells 2 and 3. At each step, the Imp moves to the adjacent cell with probability 0.5, and remains in the same cell with probability 0.5.

At each step, the agent may move in any of the four directions—up, down, left, and right—or, in alternative, *shoot* the Imp. Movement actions (up, down, etc.) across a *grey* cell division succeed with probability 1.0, and across black divisions fail with probability 1.0 (in which case, the agent remains in the same cell). Once the agent reaches cell 4, it remains there forever (independently of which action it selects).

The action "Shoot", always keeps the position of the agent unchanged, and *kills* the Imp with probability 1.0 if the agent is in the same cell as or an adjacent cell to that of the Imp. Otherwise, it has no effect. Once the Imp is killed, it disappears from the environment.

The agent is able to see the Imp with probability 1.0 if it stands in the same cell as the Imp. If it stands in a cell adjacent to that of the Imp, it will see the Imp with probability 0.5, and with probability 0.5 it will not see the Imp. If the Imp is dead, the agent does not see it. At each step, the agent sees its own position with probability 1.0.

The agent pays a cost of 1.0 if it stands in the same cell as the Imp. It pays a cost of 0.0 for standing in the Goal cell. In all other cells, the agent pays a cost of 0.3.

Describe the decision problem faced by the agent using the adequate type of model. In particular, you should indicate:

- The type of model needed to describe the decision problem of the agent;
- The state space;
- The action space;
- The observation space (if relevant);
- The transition probabilities for the action shoot;
- The observation probabilities for the action shoot (if relevant);
- The immediate cost function.

Solution 1.

- In order for the agent to reach the goal while avoiding the Imp, the agent should—ideally—know its own position and that of the Imp at each time step. As such the states should include this information, leading to the state space

$$\mathcal{X} = \{(1, D), (1, 2), (1, 3), (2, D), (2, 2), (2, 3), (3, D), (3, 2), (3, 3), 4\}\,.$$

  In a state (x, y), x represents the agent's position, and y represents the cell of the Imp (or D if the Imp is dead). Since as soon as the agent reaches 4 it remains there forever, we do not need to track the Imp at this point, so we consider a single state 4.

- The agent has 5 actions available: "up", "down", "left", "right", "shoot", which we represent as

$$\mathcal{A} = \{u, d, l, r, s\}\,.$$

- At each moment, the agent can observe its own position and, in some circumstances, the position of the Imp. This leads to the observation space

$$\mathcal{Z} = \{(1, \emptyset), (1, 2), (2, \emptyset), (2, 2), (2, 3), (3, \emptyset), (3, 2), (3, 3), 4\}\,.$$

- The action "shoot" maintains the position of the agent unchanged. If the Imp is next to or in the same cell as the agent, the Imp dies; otherwise, the position of the Imp changes with probability 0.5. This corresponds to the transition probabilities:

$$\mathbf{P}_s = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}.$$

- The observation probabilities are action-independent. The agent always observes its own position. Additionally, when in the same cell as the Imp, the agent observes it with probability 1.0. When in a cell adjacent to the Imp, the agent observes it with probability 0.5. Finally, if the Imp is dead, the agent obviously cannot see it. This leads to the observation probabilities

$$\mathbf{O}_s = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.5 & 0.0 & 0.5 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.5 & 0.5 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}.$$

- Finally, the cost function comes, directly,

$$\mathbf{C} = \begin{bmatrix} 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 0.3 & 0.3 & 0.3 & 0.3 & 0.3 \\ 1.0 & 1.0 & 1.0 & 1.0 & 1.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{bmatrix}.$$

Question 2. (**2 pts.**)

Given an arbitrary HMM $(\mathcal{X}, \mathcal{Z}, \mathbf{P}, \mathbf{O})$ with initial distribution μ_0 and a sequence of observations, $\boldsymbol{z}_{0:T}$, show that the probability distribution $\mu_{T|0:T}$ computed using the forward algorithm remains unchanged if, at each step t of the algorithm, the vectors α_t are normalized.

Solution 2.

Recall that, according to the definition of forward mapping,

$$\alpha_t(x) = \mathbb{P}\left[\mathrm{x}_t = x, \mathrm{z}_{0:t} = \boldsymbol{z}_{0:t}\right].$$

Let $\hat{\alpha}$ denote the normalized version of α_t. We get

$$\begin{aligned}\hat{\alpha}_t(x) &= \frac{\alpha_t(x)}{\sum_{x' \in \mathcal{X}} \alpha_t(x')} = \frac{\mathbb{P}\left[\mathrm{x}_t = x, \mathrm{z}_{0:t} = \boldsymbol{z}_{0:t}\right]}{\sum_{x' \in \mathcal{X}} \mathbb{P}\left[\mathrm{x}_t = x', \mathrm{z}_{0:t} = \boldsymbol{z}_{0:t}\right]} \\ &= \frac{\mathbb{P}\left[\mathrm{x}_t = x, \mathrm{z}_{0:t} = \boldsymbol{z}_{0:t}\right]}{\sum_{x' \in \mathcal{X}} \mathbb{P}\left[\mathrm{z}_{0:t} = \boldsymbol{z}_{0:t}\right]} = \mathbb{P}\left[\mathrm{x}_t = x \mid \mathrm{z}_{0:t} = \boldsymbol{z}_{0:t}\right] = \mu_{t|0:t}(x).\end{aligned}$$

Suppose now that we perform a forward update on $\hat{\alpha}_t$. We get

$$\begin{aligned}\hat{\alpha}_{\mathrm{upd}}(x) &= \mathbf{O}(z_{t+1} \mid x) \sum_{x' \in \mathcal{X}} \mathbf{P}(x \mid x') \hat{\alpha}(x') \\ &= \mathbb{P}\left[\mathrm{z}_{t+1} = z_{t+1} \mid \mathrm{x}_{t+1} = x\right] \sum_{x' \in \mathcal{X}} \mathbb{P}\left[\mathrm{x}_{t+1} = x \mid \mathrm{x}_t = x'\right] \mathbb{P}\left[\mathrm{x}_t = x' \mid \mathrm{z}_{0:t} = \mathbf{z}_{0:t}\right] \\ &= \mathbb{P}\left[\mathrm{z}_{t+1} = z_{t+1} \mid \mathrm{x}_{t+1} = x\right] \mathbb{P}\left[\mathrm{x}_{t+1} = x \mid \mathrm{z}_{0:t} = \mathbf{z}_{0:t}\right] \\ &= \mathbb{P}\left[\mathrm{z}_{t+1} = z_{t+1} \mid \mathrm{x}_{t+1} = x, \mathrm{z}_{0:t} = \mathbf{z}_{0:t}\right] \mathbb{P}\left[\mathrm{x}_{t+1} = x \mid \mathrm{z}_{0:t} = \mathbf{z}_{0:t}\right],\end{aligned}$$

where the last equality follows from the fact that the observation probabilities are independent of the past observations. Moreover,

$$\begin{aligned}\sum_{x \in \mathcal{X}} \hat{\alpha}_{\mathrm{upd}}(x) &= \sum_{x \in \mathcal{X}} \mathbb{P}\left[\mathrm{z}_{t+1} = z_{t+1} \mid \mathrm{x}_{t+1} = x\right] \mathbb{P}\left[\mathrm{x}_{t+1} = x \mid \mathrm{z}_{0:t} = \mathbf{z}_{0:t}\right] \\ &= \mathbb{P}\left[\mathbf{z}_{t+1} = \boldsymbol{z}_{t+1} \mid \mathbf{z}_{0:t} = \boldsymbol{z}_{0:t}\right].\end{aligned}$$

Finally,

$$\frac{\hat{\alpha}_{\mathrm{upd}}(x)}{\sum_{x\in\mathcal{X}}\hat{\alpha}_{\mathrm{upd}}(x)} = \frac{\mathbb{P}\left[\mathrm{z}_{t+1}=z_{t+1}\mid \mathrm{x}_{t+1}=x, \mathrm{z}_{0:t}=\mathbf{z}_{0:t}\right]\mathbb{P}\left[\mathrm{x}_{t+1}=x\mid \mathrm{z}_{0:t}=\mathbf{z}_{0:t}\right]}{\mathbb{P}\left[\mathbf{z}_{0:t+1}=\boldsymbol{z}_{0:t+1}\right]}$$
$$= \mathbb{P}\left[\mathrm{x}_{t+1}=x_{t+1}\mid \mathrm{z}_{0:t+1}=\mathbf{z}_{0:t+1}\right],$$

where the last equality follows from Bayes rule. The conclusion follows.

In the remainder of the test, consider the POMDP $\mathcal{M}=(\mathcal{X},\mathcal{A},\mathcal{Z},\{\mathbf{P}_a\},\{\mathbf{O}_a\},c,\gamma)$ where

- $\mathcal{X}=\{1,2,3\}$;
- $\mathcal{A}=\{a,b,c\}$;
- $\mathcal{Z}=\{u,v\}$;
- The transition probabilities are

$$\mathbf{P}_a=\begin{bmatrix}1.0 & 0.0 & 0.0\\ 0.0 & 0.2 & 0.8\\ 0.0 & 0.8 & 0.2\end{bmatrix};\qquad \mathbf{P}_b=\begin{bmatrix}0.2 & 0.8 & 0.0\\ 0.8 & 0.2 & 0.0\\ 0.0 & 0.0 & 1.0\end{bmatrix};\qquad \mathbf{P}_c=\begin{bmatrix}1.0 & 0.0 & 0.0\\ 0.0 & 1.0 & 0.0\\ 0.0 & 0.0 & 1.0\end{bmatrix}.$$

- The observation probabilities are

$$\mathbf{O}_a=\begin{bmatrix}1.0 & 0.0\\ 0.3 & 0.7\\ 0.7 & 0.3\end{bmatrix};\qquad \mathbf{O}_b=\begin{bmatrix}0.3 & 0.7\\ 0.7 & 0.3\\ 0.0 & 1.0\end{bmatrix};\qquad \mathbf{O}_c=\begin{bmatrix}1.0 & 0.0\\ 0.5 & 0.5\\ 0.0 & 1.0\end{bmatrix}.$$

- The cost function c is given by

$$\mathbf{C}=\begin{bmatrix}0.2 & 0.2 & 0.0\\ 1.0 & 1.0 & 0.8\\ 1.0 & 1.0 & 0.8\end{bmatrix}.$$

- Finally, the discount is given by $\gamma=0.9$.

Question 3. (**2 pts.**)

Consider the initial distribution for the POMDP $\mathcal{M}$ given by

$$\boldsymbol{b}_0=\begin{bmatrix}0.1 & 0.3 & 0.6\end{bmatrix}.$$

Suppose that the agent performs actions $\boldsymbol{a}_{0:1}=\{a,b\}$ and makes the observations $\boldsymbol{z}_{1:2}=\{u,u\}$. Determine the most likely sequence of states $\boldsymbol{x}^*_{0:2}$.

Solution 3.

We use the Viterbi algorithm. The initial distribution is

$$\boldsymbol{b}_0 = \begin{bmatrix} 0.1 & 0.3 & 0.6 \end{bmatrix},$$

and since there is no initial observation, we have that

$$\boldsymbol{m}_0 = \begin{bmatrix} 0.1 & 0.3 & 0.6 \end{bmatrix}.$$

At time step $t = 0$ the agent performs action $a_0 = a$ and then makes, at time step $t = 1$, the observation $z_1 = u$. Then,

$$\begin{aligned} \boldsymbol{m}_1 &= (\max \operatorname{diag}(\boldsymbol{m}_0)\mathbf{P}_a) \operatorname{diag}(\mathbf{O}_{a,u}) \\ &= \max \begin{bmatrix} 0.1 & 0.0 & 0.0 \\ 0.0 & 0.06 & 0.24 \\ 0.0 & 0.48 & 0.12 \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.7 \end{bmatrix} \\ &= \begin{bmatrix} 0.1 & 0.48 & 0.24 \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 0.3 & 0.0 \\ 0.0 & 0.0 & 0.7 \end{bmatrix} \\ &= \begin{bmatrix} 0.1 & 0.144 & 0.168 \end{bmatrix}, \end{aligned}$$

with

$$\boldsymbol{i}_1 = \operatorname{argmax} \operatorname{diag}(\boldsymbol{m}_0)\mathbf{P}_a = \begin{bmatrix} 1 & 3 & 2 \end{bmatrix}.$$

Similarly, at time step $t = 1$ the agent performs action $a_1 = b$ and makes, at time step $t = 2$, the observation $z_2 = u$. Then,

$$\boldsymbol{m}_2 = (\max \operatorname{diag}(\boldsymbol{m}_1)\mathbf{P}_b) \operatorname{diag}(\mathbf{O}_{b,u}) = \begin{bmatrix} 0.035 & 0.056 & 0.0 \end{bmatrix},$$

with

$$\boldsymbol{i}_2 = \operatorname{argmax} \operatorname{diag}(\boldsymbol{m}_1)\mathbf{P}_b = \begin{bmatrix} 2 & 1 & 3 \end{bmatrix}.$$

Finally, we can conclude that the most likely state at $t = 2$ is $x_2^* = 2$. Backtracking, we successively have: $x_1^* = 1$ and $x_0^* = 1$. The most likely sequence is, therefore, $\boldsymbol{x}_{0:2}^* = \{1, 1, 2\}$.

Question 4. (**11 pts.**)

Consider the MDP obtained from $\mathcal{M}$ by ignoring partial observability. Consider also an agent that, in this MDP, follows the policy

$$\boldsymbol{\pi} = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 \end{bmatrix}.$$

(a) **(1 pt.)** Compute the transition probabilities for the Markov chain induced by the policy π.

(b) **(2 pt.)** Suppose that the agent's initial state is drawn from the distribution

$$\boldsymbol{b}_0 = \begin{bmatrix} 0.1 & 0.3 & 0.6 \end{bmatrix}.$$

Compute the state distribution after 2 steps if the agent follows policy π.

(c) **(3 pt.)** Compute the cost-to-go function for the policy π.

(d) **(3 pt.)** Show that the greedy policy with respect to J^π is

$$\pi_g^{J^\pi} = \begin{bmatrix} 0.0 & 0.0 & 1.0 \\ 0.0 & 1.0 & 0.0 \\ 1.0 & 0.0 & 0.0 \end{bmatrix}.$$

(e) **(1 pt.)** Suppose that the Q-function for policy $\pi_g^{J^\pi}$ is given by

$$Q^{\pi_g} = \begin{bmatrix} 0.2 & 1.08 & 0.0 \\ 2.87 & 1.22 & 1.9 \\ 2.29 & 3.06 & 2.86 \end{bmatrix}.$$

Is the policy $\pi_g^{J^\pi}$ optimal? Justify your answer.

(f) **(1 pt.)** Using the policy $\pi_g^{J^\pi}$, compute the action prescribed by the AV heuristic for the POMDP, given the belief

$$\boldsymbol{b} = \begin{bmatrix} 0.1 & 0.3 & 0.6 \end{bmatrix}.$$

Solution 4.

(a) The transition probabilities come, directly,

$$\mathbf{P}_\pi = \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.8 & 0.2 & 0.0 \\ 0.0 & 0.8 & 0.2 \end{bmatrix}.$$

(b) To compute the state distribution after two steps, we can compute

$$\boldsymbol{b}_0\mathbf{P}_\pi^2 = \begin{bmatrix} 0.1 & 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.96 & 0.04 & 0.0 \\ 0.64 & 0.32 & 0.04 \end{bmatrix} = \begin{bmatrix} 0.772 & 0.204 & 0.024 \end{bmatrix}$$

(c) The cost to go J^π can be computed as

$$\boldsymbol{J}^\pi = (\boldsymbol{I} - \gamma\mathbf{P}_\pi)^{-1}\boldsymbol{c}_\pi,$$

where $\mathbf{P}_\pi$ was computed in (a), and

$$\boldsymbol{c}_\pi = \begin{bmatrix} 0.2 & 1.0 & 1.0 \end{bmatrix}^\top.$$

We thus get

$$\begin{aligned} \boldsymbol{J}^\pi &= (\boldsymbol{I} - \gamma\mathbf{P}_\pi)^{-1}\boldsymbol{c}_\pi \\ &= \left(\begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} - 0.9 \begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.8 & 0.2 & 0.0 \\ 0.0 & 0.8 & 0.2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0.2 \\ 1.0 \\ 1.0 \end{bmatrix} \\ &= \begin{bmatrix} 10.0 & 0.0 & 0.0 \\ 8.78 & 1.22 & 0.0 \\ 7.71 & 1.07 & 1.22 \end{bmatrix} \begin{bmatrix} 0.2 \\ 1.0 \\ 1.0 \end{bmatrix} = \begin{bmatrix} 2.0 \\ 2.98 \\ 3.83 \end{bmatrix}. \end{aligned}$$

(d) To compute the greedy policy, we start by computing the Q-function Q^π. We have, for action a,

$$\begin{aligned}\boldsymbol{Q}^\pi_{:,a} &= \boldsymbol{C}_{:,a} + \gamma \mathbf{P}_a \boldsymbol{J}^\pi \\ &= \begin{bmatrix}0.2\\1.0\\1.0\end{bmatrix} + 0.9 \begin{bmatrix}1.0 & 0.0 & 0.0\\0.0 & 0.2 & 0.8\\0.0 & 0.8 & 0.2\end{bmatrix} \begin{bmatrix}2.0\\2.98\\3.83\end{bmatrix} = \begin{bmatrix}2.0\\4.29\\3.83\end{bmatrix}.\end{aligned}$$

Similarly, for actions b and c,

$$\boldsymbol{Q}^\pi_{:,b} = \boldsymbol{C}_{:,b} + \gamma \mathbf{P}_b \boldsymbol{J}^\pi = \begin{bmatrix}2.7\\2.98\\4.45\end{bmatrix} \qquad \boldsymbol{Q}^\pi_{:,c} = \boldsymbol{C}_{:,c} + \gamma \mathbf{P}_c \boldsymbol{J}^\pi = \begin{bmatrix}1.8\\3.48\\4.25\end{bmatrix},$$

leading to the Q-function

$$\boldsymbol{Q}^\pi = \begin{bmatrix}2.0 & 2.70 & 1.8\\4.29 & 2.98 & 3.48\\3.83 & 4.45 & 4.25\end{bmatrix}.$$

The conclusion follows.

(e) Computing the greedy policy with respect to Q^{π_g}, we recover the policy

$$\pi_g^{J^\pi} = \begin{bmatrix}0.0 & 0.0 & 1.0\\0.0 & 1.0 & 0.0\\1.0 & 0.0 & 0.0\end{bmatrix},$$

indicating that—in a policy iteration algorithm—we have reached convergence. Then, policy $\pi_g^{J^\pi}$ must be optimal.

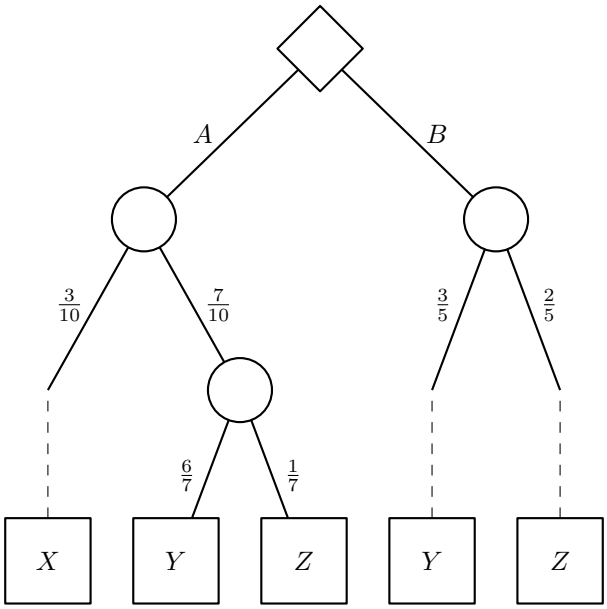
(f) Using the AV heuristic, each state "votes" in the corresponding action proportionally to its probability in $\boldsymbol{b}$. This means that

- State 1 votes for action c with 0.1 votes;
- State 2 votes for action b with 0.3 votes;
- State 3 votes for action a with 0.6 votes.

The agent will thus select action a.

Question 5. (**2 pts.**)

Consider the decision problem described by the following decision tree.



Identify a valid preference relation between outcomes X, Y and Z that ensures that action A is optimal.

Solution 5.

Let u denote the utility associated with the sought preference. Then,

$$Q(A) = 0.3u(X) + 0.6u(Y) + 0.1u(Z)$$
$$Q(B) = 0.6u(Y) + 0.4u(Z).$$

In order for $Q(A) > Q(B)$ we must have

$$0.3u(X) + 0.6u(Y) + 0.1u(Z) > 0.6u(Y) + 0.4u(Z),$$

or, equivalently,

$$u(X) > u(Z).$$

Therefore, any preference relation such that $u(X) > u(Z)$ ensures the desired result (e.g., $X \succ Y \succ Z$).