INFO8006 Introduction to Artificial Intelligence

Exam of January 2022

Instructions

- *Duration: 4 hours.*
- *Provide a separate sheet for each question (including those you do not answer), labelled with the question number, your first name, last name and student id.*
- *Answer in English or French.*
- *You are allowed to use a calculator during the exam, but documents of any kind are forbidden.*

Question 1 [4 points]

Multiple choice questions. Choose one of the four choices. Correct answers are graded $+\frac{4}{10}$*, wrong answers are graded* $-\frac{2}{15}$ *and the absence of answers is graded* 0*. The total of your grade for Question 1 is bounded below at* $0/4$*.*

1. Which of the following is true? An environment is...
   (a) fully observable whenever the agent perceives images of the environment.
   (b) deterministic whenever the next state of the environment is fully determined by the current state and the action of the agent.
   (c) is episodic whenever the environment is modeled as a discrete series of time steps.
   (d) is unknown whenever the agent sensors do not give access to the complete state of the environment, at each point in time.

2. Consider the game of Pacman in the absence of ghosts and with a single food dot in the maze. In state s, Pacman is located at position (i_s, j_s) while the food dot is located at (x_s, y_s). At timestep t, the score of the game is $500 - t$. The game ends when the food is eaten. Which of the following heuristics is the best to use?
   (a) $h_1(s) = |i_s - x_s| + |j_s - y_s|$
   (b) $h_2(s) = \sqrt{(i_s - x_s)^2 + (j_s - y_s)^2}$
   (c) $h_3(s) = \max(|i_s - x_s|, |j_s - y_s|)$
   (d) $h_4(s) = \min(h_1(s), h_2(s), h_3(s))$

3. Which of the following is wrong? In turn-taking stochastic games, ...
   (a) randomness can be modeled with chance nodes that map a q-state to the set of possible outcomes, along with their respective probability.
   (b) randomness can be modeled by adding an extra random agent that acts once all other players have acted.
   (c) EXPECTIMINIMAX is optimal if the heuristic evaluation function is a positive linear transformation of the expected utility of the state.
   (d) Monte Carlo Tree Search does not guarantee that the optimal action is selected.

4. In probability theory, the chain rule implies that
   (a) $P(a, b, c) = P(a)P(b)P(c)$.
   (b) $P(a, b, c) = P(a, b)P(b, c)$.
   (c) $P(a, b, c) = P(a)P(b|a)P(c|a, b)$.
   (d) $(f \circ P)'(x) = P'(f(x))f'(x)$.

5. Sampling from the distribution of a random variable X with k discrete outcomes $x_1, \ldots, x_k$ can be implemented by...
   (a) dividing the interval $[0, 1]$ into k regions of equal size and then mapping a random outcome $u \sim \mathcal{U}(0, 1)$ to the value x_i associated with the region i in which u falls.
   (b) dividing the interval $[0, 1]$ into k regions of equal size and then mapping a random outcome $u \sim \mathcal{N}(0, 1)$ to the value x_i associated with the region i in which u falls.

(c) dividing the interval $[0,1]$ into k regions, with region i having a size $P(x_i)$, and then mapping a random outcome $u \sim \mathcal{U}(0,1)$ to the value x_i associated with the region i in which u falls.

(d) dividing the interval $[0,1]$ into k regions, with region i having a size $P(x_i)$, and then mapping a random outcome $u \sim \mathcal{N}(0,1)$ to the value x_i associated with the region i in which u falls.

6. Let us assume a Markov process with hidden state variables $\mathbf{X}_t$, evidence variables $\mathbf{E}_t$, transition model $\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{X}_t)$ and sensor model $\mathbf{P}(\mathbf{E}_t|\mathbf{X}_t)$. Which of the following is true?

   (a) The Bayes filter estimates $\mathbf{P}(\mathbf{X}_{1:t}|\mathbf{e}_{1:t})$.

   (b) Smoothing consists in computing $\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{e}_{1:t+1})$ by pushing the belief state $\mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t})$ forward through the transition model and then updating this new belief state with the evidence $\mathbf{e}_{t+1}$.

   (c) The Kalman filter computes $\mathbf{P}(\mathbf{X}_{t+k}|\mathbf{e}_{1:t})$ by pushing the belief state $\mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t})$ forward k times through the transition model.

   (d) The Viterbi algorithm computes the most likely sequence of state values $\mathbf{x}_{1:t}$ given the evidence $\mathbf{e}_{1:t}$.

7. Let us consider a robot wandering around at the Montefiore Institute. Which of the following is true?

   (a) The robot and its environment can be modeled as a partially observable MDP.

   (b) The Kalman filter can be used for determining accurately its past trajectory in the building.

   (c) A convolutional neural network would be too large to fit in the robot's memory.

   (d) Value iteration (with an admissible heuristic) can be used for decoding the speech of its visitors.

8. The Bellman equations $Q(s,a) = R(s) + \gamma \sum_{s'} P(s'|s,a) \max_{a'} Q(s',a')$ form a system of n non-linear equations with as many unknowns. Which of the following is true?

   (a) n is the size of the action space $\mathcal{A}$.

   (b) n is the size of the state space $\mathcal{S}$.

   (c) n is the size of the state-action space $\mathcal{S} \times \mathcal{A}$.

   (d) n is the sum of the sizes of the action and state spaces.

9. In DQN (Mnih et al, 2015), a reinforcement learning agent is trained to ...

   (a) to classify images.

   (b) to control a robot in a simulated environment.

   (c) to play Atari video games.

   (d) to play the game of Go.

10. In DQN (Mnih et al, 2015), the Q-table is approximated with ...

    (a) a hash table.

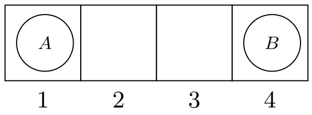
    (b) a transposition table.

    (c) a linear regression model.

    (d) a convolutional neural network.

Question 2 [4 points]

Consider the following two-player turn-taking game which initial configuration is shown in the figure below. Player A moves first. Each player must move their token to an open adjacent cell in either direction. If the opponent occupies an adjacent cell, then a player may jump over the opponent to the next open cell, if any. (For example, if A is on 3 and B is on 2, then A may move back to 1.) The game ends when a player reaches the opposite end of the board. If player A reaches cell 4 first, then the value of the game to A is $+1$; if player B reaches cell 1 first, then the value of the game to A is -1.



(a) Define the search problem associated with this game.

(b) Draw the complete game tree, using the following conventions:

- Put each terminal state in a square box and write its game value in a circle.
- Put loop states (states that already appear on the path to the root) in double square boxes. Since their value is unclear, annotate each with a '?' in a circle.

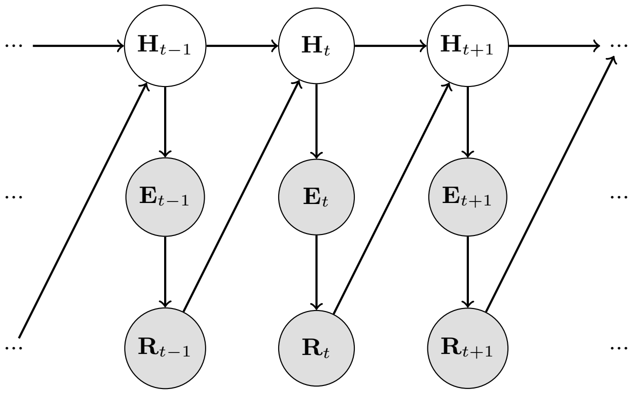
(c) Mark each node with its backed-up minimax value (also in a circle). Explain how you handled the '?' values and why.

(d) Explain why the standard minimax algorithm would fail on this game tree and briefly sketch how you might fix it.

(e) This 4-cell game can be generalized to n cells for any $n > 2$. Prove that A wins if n is even and loses if n is odd. (Hint: Use induction.)

Question 3 [4 points]

In the near future, autonomous robots will live among us. Therefore, the robots need to know how to act appropriately in the presence of humans. In this question, we explore a simplified model of this interaction. We assume that we can observe the robot's actions at time t, $\mathbf{R}_t$, and an evidence observation, $\mathbf{E}_t$, directly caused by the (unobserved) human action, $\mathbf{H}_t$. Human actions and robot actions from the past time-step affect the human and the robot actions in the next-time step, as illustrated in the dynamic Bayesian network given below.



Assuming discrete variables and given $\mathbf{P}(\mathbf{H}_0)$, $\mathbf{P}(\mathbf{E}_t|\mathbf{H}_t)$, $\mathbf{P}(\mathbf{R}_t|\mathbf{E}_t)$ and $\mathbf{P}(\mathbf{H}_t|\mathbf{H}_{t-1}, \mathbf{R}_{t-1})$, our goal is to derive a procedure to maintain a belief state $\mathbf{P}(\mathbf{H}_t|\mathbf{e}_{0:t}, \mathbf{r}_{0:t})$ about the state of the human at time t.

(a) Derive an update equation for incorporating a pair of observations $(\mathbf{e}_t, \mathbf{r}_t)$ to a given belief state $\mathbf{P}(\mathbf{H}_t)$.

(b) Derive an update equation for predicting the future state $\mathbf{H}_{t+1}$ of the human at time $t+1$ given a belief state $\mathbf{P}(\mathbf{H}_t|\mathbf{e}_{0:t}, \mathbf{r}_{0:t})$.

(c) Combine both equations to define a recursive update equation of the belief state $\mathbf{P}(\mathbf{H}_t|\mathbf{e}_{0:t}, \mathbf{r}_{0:t})$, as observations are collected and time passes.

(d) Let us now assume that all variables are continuous. Discuss how you would compute or approximate the belief state on a computer.

Question 4 [4 points]

A new virtual escape game came out, and you decide to play it. You arrive in a 5×5 grid world where each cell (i, j) is a room with doors leading to the adjacent rooms. The game's goal is to reach the exit room as fast as possible, but its position is unknown. Furthermore, some regions of the world are full of riddles, and crossing rooms in these regions takes longer. Fortunately, a leaderboard with the players' best times is provided, starting from a few different rooms. Due to rounding errors, you assume that the best times reported in the leaderboard are measurements affected by Gaussian noise of unit variance.

Starting room	Measured best time
$(4,5)$	2.0
$(5,3)$	3.5
$(3,3)$	4.5
$(4,1)$	7.0
$(1,2)$	8.5

From the leaderboard, you wish to learn a heuristic $h(i, j)$ approximating the optimal time to get to the exit from room (i, j). You decide to use a small neural network as approximator, described by the following parametric function,

$$\begin{aligned} h(i,j) &= \mathrm{ReLU}(iw_{1,i} + jw_{1,j} + b_1) + \mathrm{ReLU}(iw_{2,i} + jw_{2,j} + b_2) \\ \mathrm{ReLU}(x) &= \max(x, 0). \end{aligned}$$

(a) Among the following sets of parameters (A, B or C), which one would you use? Justify your answer.

	$w_{1,i}$	$w_{1,j}$	b_1	$w_{2,i}$	$w_{2,j}$	b_2
$A:$	-1.5	1	4	1	-1.5	6
$B:$	-1	1.5	3	0	-1	4
$C:$	-2	0.5	4.5	1.5	0	5

(b) You now assume a Gaussian prior $\mathcal{N}(0, 1)$ on each parameter. Which set of parameters in the table above would you now choose? Justify your answer.

(c) Discuss the procedure you would implement on a computer to find the optimal set of parameters, had the table above not been provided?

(d) Using the parameters $(-2, 1, 5, 0.5, -2, 7)$, apply 5 iterations of the greedy search algorithm, starting from room $(1, 1)$.

Question 5 [4 points]

(a) Define Markov decision processes (MDPs), formally.

(b) Define the optimal policy of an MDP, formally.

(c) Describe the Value iteration algorithm, with pseudo-code.

(d) In Micro-Blackjack, you repeatedly draw a card (with replacement) that is likely to be a 2, 3 or 4. At each step, if the total score of the cards is lower than 6, you can either draw again (d) or cash (c). Otherwise, you can only cash. When you cash, the game stops and your utility is equal to your total score (up to 5) plus 1, or zero if you get a total of 6 or higher. Until you cash and after it, you receive no reward. There is no discount ($\gamma = 1$).

   (i) Formalize Micro-Blackjack as an MDP.

   (ii) Derive the optimal policy for this MDP using Value iteration.

   (iii) If you were not told the rules of the game, what algorithm would you use to obtain an optimal policy?