INFO8006 Introduction to Artificial Intelligence

Exam of August 2021

Instructions

- *Duration: 4 hours.*
- *Answer each question on a separate sheet, labelled with the question number, your first name, last name and student id.*
- *Question 1 must be answered on the SMART form in attachment.*
- *Answer in English or French.*

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*. Question 1 must be answered on the SMART form in attachment.*

1. There exists task environments in which no simple reflex agent can behave rationally.
   (a) True.
   (b) True, but only if the task environments are static.
   (c) False.
   (d) False, provided the agent implements compressed condition-actions rules with a multi-layer perceptron.
2. In a search problem, ...
   (a) a solution is a next action to take from the initial state.
   (b) a solution is an action sequence that leads from the initial state to a goal state.
   (c) a solution is a policy $\pi : \mathcal{S} \to \mathcal{A}$ that describes an entire mapping from states to actions while ensuring convergence to a goal state.
   (d) a solution is a policy $\pi : \mathcal{S} \to \mathcal{A}$ that describes an entire mapping from states to actions while maximizing the expected utility.
3. Which of the following is wrong? In Pacman, the evaluation function of an agent running the Minimax algorithm could be implemented as ...
   (a) a hardcoded constant value.
   (b) a look-up table.
   (c) a Bayesian network.
   (d) a neural network.
4. Let us assume 80% of the population is vaccinated and that the vaccine grants immunity with 99.9% probability. If 1% of the population is currently sick, then the proportion of vaccinated among the sick population is
   (a) 0.1%.
   (b) 8%.
   (c) 80%.
   (d) 92%.
5. The Normal distribution $\mathcal{N}(\mu, \sigma)$ is described by the density function
   (a) $p(x) = \frac{1}{\sigma}\exp\left(-(z + \exp(-z))\right)$, with $z = \frac{x-\mu}{\sigma}$.
   (b) $p(x) = \frac{1}{2\sigma}\exp\left(-\frac{|x-\mu|}{\sigma}\right)$.
   (c) $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
   (d) $p(x) = 1 \,/\, \pi\sigma\left[1 + \left(\frac{x-\mu}{\sigma}\right)^2\right]$.

6. The Bayes filter requires the analytical specification of ...
   (a) a prior $\mathbf{P}(\mathbf{X}_0)$ and a transition model $\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t)$.
   (b) a transition model $\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t)$ and an observation model $\mathbf{P}(\mathbf{E}_t|\mathbf{x}_t)$.
   (c) a prior $\mathbf{P}(\mathbf{X}_0)$, a transition model $P(\mathbf{X}_{t+1}|\mathbf{x}_t)$, and an observation model $P(\mathbf{E}_t|\mathbf{x}_t)$.
   (d) a prior $\mathbf{P}(\mathbf{X}_0)$, a prior $\mathbf{P}(\mathbf{E}_0)$, a transition model $P(\mathbf{X}_{t+1}|\mathbf{x}_t)$, and an observation model $P(\mathbf{E}_t|\mathbf{x}_t)$.
7. Which of the following is wrong? Convolutional neural networks ...
   (a) have a topology that encodes conditional independence assumptions between random variables.
   (b) are usually trained by solving a maximum likelihood estimation problem.
   (c) usually count thousands to millions of parameters.
   (d) are fit for processing spatially structured data, such as images or sequences.
8. Which of the following is true? In Markov Decision Processes, ...
   (a) the closer the discount factor γ to 0, the higher the utility of future rewards.
   (b) the closer the discount factor γ to 0, the longer Value Iteration may take to converge.
   (c) the closer the discount factor γ to 1, the greedier the optimal agent.
   (d) the closer the discount factor γ to 1, the longer Value Iteration may take to converge.
9. In reinforcement learning, ...
   (a) direct utility estimation is an efficient learning algorithm.
   (b) the temporal-difference update is equivalent to an exponential moving median.
   (c) exploration strategies can help train the agent faster but they are never strictly necessary.
   (d) the approximate Q-Learning algorithm replaces the tabular TD-update by a gradient step on the parameters of the function approximator of the Q-table.
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]

Let us consider the family of generalized tic-tac-toe games, defined as follows. Each particular game is specified by a set $\mathcal{S}$ of squares and a collection $\mathcal{W}$ of winning positions. Each winning position is a subset of $\mathcal{S}$. For example, in standard tic-tac-toe, $\mathcal{S}$ is a set of 9 squares, and $\mathcal{W}$ is a collection of 8 subsets of $\mathcal{S}$: the three rows, the three columns, and the two diagonals. In other respects, the game is identical to standard tic-tac-toe. Starting from an empty board, players alternate placing their marks on an empty square. A player who marks every square in a winning position wins the game. It is a tie if all squares are marked and neither player has won.

(a) Is the environment fully observable or partially observable? deterministic or stochastic? static or dynamic? Known or unknown? Explain your answers briefly.

(b) Let $N = |\mathcal{S}|$ the number of squares. Give an upper bound on the number of nodes in the complete game tree for generalized tic-tac-toe as a function of N.

(c) Propose and motivate a plausible evaluation function that can be used for any instance of generalized tic-tac-toe. The function may depend on $\mathcal{S}$ and $\mathcal{W}$.

(d) Assume it is possible to generate a new board and check whether it is a winning position in $100N$ machine instructions and assume a 2 gigahertz processor. Ignoring memory limitations and using your previous estimate for the size of a complete game tree, roughly how large Minimax can completely solve a game tree in i) a second of CPU time? ii) a minute?, iii) an hour?

Question 3 [4 points]

Two astronomers in different parts of the world make measurements M_1 and M_2 of the number of stars N in some small region of the sky, using their telescopes. In normal conditions (i.e. when the focus is right), there is a little possibility e of error by up to one star in each direction. Each telescope can also (with a much smaller probability f) be badly out of focus (events F_1 and F_2), in which case the scientist will undercount by three or more stars (or if N is less than 3, fail to detect any stars at all). Consider the three networks shown in Figure 1 below.

(a) Which of these Bayesian networks is a wrong representation of the preceding information?

(b) Which is the best network? Explain.

(c) Write out a conditional distribution for $P(M_1|N)$, for the case where $N \in \{1, 2, 3\}$ and $M_1 \in \{0, 1, 2, 3, 4\}$. Each entry in the conditional distribution should be expressed as a function of the parameters e and/or f.

(d) Suppose $M_1 = 1$ and $M_2 = 3$. What are the possible numbers of stars if you assume no prior constraint on the values of N?

(e) What is the most likely number of stars, given these observations? Explain how to compute this, or if it is not possible to compute, explain what additional information is needed and how it would affect the result. (Hint: to find the final answer, use bounds on the probability of the possible values found in the previous question.)

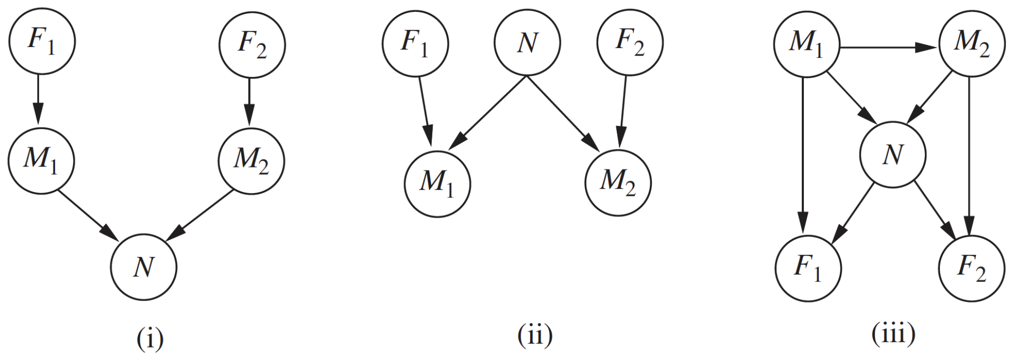

Figure 1: Three possible networks for the telescope problem.

Question 4 [4 points]

You observe a grandmaster playing Pacman and wish to learn from her games. To this end, you write down in a table all the state-action pairs (s, a) played by the grandmaster, together with their corresponding Q-values $Q(s, a)$. You describe each state-action pair with six features: the horizontal and vertical position of Pacman (x_P, y_P) and of the ghost (x_G, y_G); the distance d to the closest food pellet; and an action feature. Unfortunately, you did not sleep too well the night before and make random errors when computing and reporting the Q-values. Assuming Gaussian errors (of zero mean and unit variance), how would you learn a model of the Q-function using your data?

(a) Describe formally the learning problem you would have to solve in the case of a linear model of the Q-function (i.e., the data, the model, and its parameters). Write down the optimization problem to estimate the model parameters.

(b) Assuming a Gaussian prior (of zero mean and unit variance) on the model parameters, revise your optimization problem above. What is the name of the resulting estimator?

(c) Derive a closed-form formula for computing the solution of this optimization problem.

(d) For the data in the table below, compute the parameters of the linear model.

x_P	y_P	x_G	y_G	d	a	Q
2	4	4	2	4	1	-0.5
2	2	2	2	-2	10	1
0	0	4	4	3	3	1

Question 5 [4 points]

In the game Farmland, players alternate taking turns drawing a card from one of two piles, *Pig* and *Cow*. Pig cards have an equal probability of being 3 points or 1 point, and Cow cards are always worth 2 points. Players are trying to be the first to reach 5 points or more. We are designing an agent to play Farmland.

We will use a Markov Decision Process to come up with a policy to play the game. States will be represented as tuples (x, y) where x is our score, and y is the opponent's score. The value $V(x, y)$ is an estimate of the probability that we will win at the state (x, y) when it is our turn to play and both players are playing optimally. Unless otherwise stated, assume both players play optimally.

First, suppose we work out by hand V^*, the table of actual probabilities.

		Opponent				
		0	1	2	3	4
	0	0.75	0.5	0.5	0	0
	1	1	1	0.5	0	0
You	2	1	1	0.75	0.5	0.5
	3	1	1	1	1	1
	4	1	1	1	1	1

(a) At the beginning of the game, would you choose to go first or second? Motivate your answer using the table above.

(b) If our current state is (x, y), but it is the opponents turn to play, what is the probability that we will win if both players play optimally in terms of V^*?

(c) Describe with pseudo-code how you implement the optimal policy using the table above?

(d) As Farmland is a very simple game, you quickly grow tired of playing it. You decide to buy the Farmland extension, Bovine Bonanza, which adds loads of exciting cards to the *Cow* pile! Of course, this changes the transition model for our MDP, so the table V^* above is no longer correct. We need to come up with an update equation that will ultimately make V_i converge on the actual probabilities that we will win. You are given the transition model $P((x', y)|(x, y), a)$ and the reward function $R((x, y))$. Since we are only interested in trying to find the probability of winning and we do not care about the margin of victory, the reward $R((x, y))$ is 1 whenever (x, y) is a winning state and 0 otherwise. Write an update equation for V_{i+1} in terms of $P((x', y)|(x, y), a)$, R and V_i.