INFO8006 Introduction to Artificial Intelligence

Exam of August 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*.*

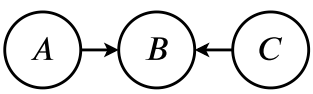
1. Which of the following is true? As an environment, the game of Chess is ...
   (a) partially observable.
   (b) stochastic.
   (c) known.
   (d) single agent.

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 should your agent used to be optimal?
   (a) $h_1(s) = (i_s - x_s)^3 + (j_s - y_s)^3$
   (b) $h_2(s) = 0$
   (c) $h_3(s) = t$
   (d) $h_4(s) = \max(h_1(s), h_2(s))$

3. In adversarial search,
   (a) the deeper in the tree the evaluation function is buried, the more the quality of the evaluation matters.
   (b) if not looked deep enough, good moves may appear as bad moves, because their consequences are hidden beyond the search horizon.
   (c) the horizon effect arises when the evaluation is perfect.
   (d) the search horizon is irrelevant if and only if the evaluation function is admissible but not consistent.

4. The Bayes' rule states that ...
   (a) $P(a|b) = P(b|a)P(b) \,/\, P(a)$.
   (b) $P(a|b) = P(b) \,/\, P(b|a)P(a)$.
   (c) $P(b|a) = P(a|b)P(a) \,/\, P(b)$.
   (d) $P(a|b) = P(b|a)P(a) \,/\, P(b)$.

5. Consider the Bayesian network below. Which of the following is always true?



   (a) A and C are causes of B.
   (b) A or C are an effect of B.
   (c) A and C are independent.
   (d) A and C are dependent given B.

6. The task of smoothing consists in computing
   (a) $\arg\max_{\mathbf{x}_{1:t}} P(\mathbf{x}_{1:t}|\mathbf{e}_{1:t})$.
   (b) $\mathbf{P}(\mathbf{X}_{t+k}|\mathbf{e}_{1:t})$, for $k > 0$.
   (c) $\mathbf{P}(\mathbf{X}_t|\mathbf{e}_{1:t})$.
   (d) $\mathbf{P}(\mathbf{X}_k|\mathbf{e}_{1:t})$, for $k < t$.

7. Linear regression usually considers a parameterized linear Gaussian model $p(y|\mathbf{x}) = \mathcal{N}(y|\mathbf{w}^T\mathbf{x} + b, \sigma^2)$, where $\mathbf{w}$ and b are parameters to determine. Which of the following is true?
   (a) This model can approximate any mathematical function arbitrarily well.
   (b) Maximum likelihood estimation for $\mathbf{w}$ and b reduces to minimizing $\sum_{j=1}^{N}(y_j - (\mathbf{w}^T\mathbf{x}_j + b))^2$.
   (c) Minimizing $\sum_{j=1}^{N}|y_j - (\mathbf{w}^T\mathbf{x}_j + b)|$ always leads to the same solution as minimizing $\sum_{j=1}^{N}(y_j - (\mathbf{w}^T\mathbf{x}_j + b))^2$.
   (d) Maximum a posteriori estimation for $\mathbf{w}$ and b always leads to the same solution as maximum likelihood estimation.

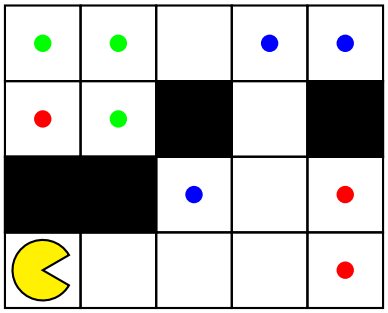
8. In a Markov decision process, the reward function...
   (a) has no effect on the optimal agent's behavior.
   (b) cannot be negative.
   (c) can shape the optimal agent's behavior from risk-taking to conservative.
   (d) produces sequences of values that sum to zero.

9. The update equation for TD Q-Learning is ...
   (a) $Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma \max_{a'} Q(s',a') - Q(s,a))$
   (b) $Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma \min_{a'} Q(s',a') - Q(s,a))$
   (c) $Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma Q(s,a) - \max_{a'} Q(s',a'))$
   (d) $Q(s,a) \leftarrow Q(s,a) + \alpha(r + \gamma Q(s,a) - \min_{a'} Q(s',a'))$

10. The search algorithm in AlphaGo is based on...
    (a) the Minimax algorithm.
    (b) Monte Carlo tree search.
    (c) A*.
    (d) a Markov Chain Monte Carlo method.

Question 2 [4 points]

Let us assume a grid-world in which there are three kinds of food pellets, each with a different color (blue, green and red). Mrs Pacman is only interested in tasting two different kinds of food: the game ends when she has eaten (at least) two pellets of distinct colors (red and blue, green and blue or red and green). Mrs Pacman has four actions: moving north, south, east, or west; although she cannot move into any of the B walls. There are L pellets of each kind and the dimensions of the grid-world are $W \times H$.



(a) Formally define an *efficient* search problem for Mrs Pacman, for which an optimal solution would minimize the number of moves while achieving the goal. Assume she starts in location (i, j).

(b) Provide a tight upper bound on the size of the state space.

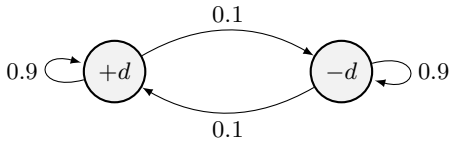
(c) Provide a tight upper bound on the branching factor of the search problem.

(d) For each of the following heuristics, indicate (yes/no) and motivate briefly (1-2 sentences) whether or not it is admissible.

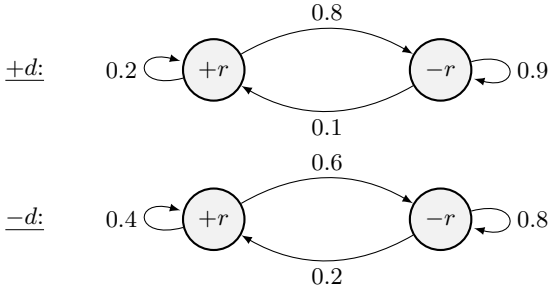
- h_1: the number of remaining red pellets.
- h_2: the Manhattan distance to the closest remaining pellet.
- h_3: the minimum Manhattan distance between any two remaining pellets.
- h_4: the maximum Manhattan distance between any two remaining pellets of different colors.

Question 3 [4 points]

In Belgium, whether it rains or not from each day to the next forms a Markov chain (note: this is a terrible model for real weather). However, due to climate changes, sometimes Belgium is in a drought and sometimes it is not. Whether Belgium is in a drought from each day to the next itself forms another Markov chain, and the state of the Markov chain affects the transition probabilities in the rain-or-shine Markov chain. The state diagram for droughts is:



The state diagrams for rain given that Belgium is ($+d$) and is not ($-d$) in a drought are respectively as follows:



(a) Draw a dynamic Bayesian network (DBN) which encodes this dynamic. Use variables $D_{t-1}, D_t, D_{t+1}, R_{t-1}, R_t$ and R_{t+1}. Assume that on a given day, it is determined whether or not there is a drought before it is determined whether or not it rains that day.

(b) Draw the conditional probability table for D_t in the DBN above. Fill in the actual numerical probability values.

(c) Draw the conditional probability table for R_t in the DBN above. Fill in the actual numerical probability values.

(d) Suppose we are observing the weather on a day-to-day basis, but we cannot directly observe whether Belgium is in a drought or not. We want to predict whether or not it will rain on day $t+1$ given observations of whether or not it rained on days 1 through t.

i. First, we need to determine whether Belgium will be in a drought on day $t+1$. Derive a formula for $P(D_{t+1}|r_{1:t})$ in terms of the given probabilities (the transition probabilities on the above state diagrams) and $P(D_t|r_{1:t})$ (that is, you can assume we have already computed the probability there is a drought today given the weather over time).

ii. Now derive a formula for $P(R_{t+1}|r_{1:t})$ in terms $P(D_{t+1}|r_{1:t})$ and the given probabilities.

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 additive Gaussian noise $\mathcal{N}(0, 1)$.

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 best 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,

$$h(i,j) = \text{ReLU}(iw_{1,i} + jw_{1,j} + b_1) + \text{ReLU}(iw_{2,i} + jw_{2,j} + b_2)$$
$$\text{ReLU}(x) = \max(x, 0).$$

(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]

ULiège's soccer team is playing against UCLouvain's team next week. With many losses this season, ULiège needs to improve their attack strategy to win the game and increase their popularity. Luckily, the team captain follows INFO8006 and knows how to model the attack as a Markov Decision Process. The captain considers four states `close`, `away`, `fail`, and `goal` and two actions `pass` and `shoot`. Although the transition probabilities are unsure, the possible transitions (s, a, s') are known. To each transition is associated an increase/decrease of the team's popularity.

s	a	s'	$R(s, a, s')$
close	pass	close	$+1$
close	pass	away	-1
close	pass	fail	-2
close	shoot	close	$+3$
close	shoot	fail	-5
close	shoot	goal	$+10$
away	pass	close	$+2$
away	pass	away	0
away	pass	fail	-3
away	shoot	close	$+3$
away	shoot	fail	-10
away	shoot	goal	$+20$

The current strategy of the team is to always shoot. During the last match, they had several attack opportunities, resulting in the following outcomes.

s	a	s'
close	shoot	goal
close	shoot	close
close	shoot	goal
close	shoot	fail
away	shoot	fail
away	shoot	close
away	shoot	fail
away	shoot	fail

Assuming a discount factor $\gamma = 0.75$ and a learning rate $\alpha = 0.25$,

(a) Build an estimator $\hat{P}(s'|s, a)$ of the transition model and, from it, determine the expected utility V^π of the team's current policy π.

(b) Perform direct utility estimation of the expected utility V^π. Do you observe a difference with the previous estimation? Why?

The captain found the tapes of the previous season where they had much more success. Together with the team, the captain selects the following instructive actions.

s	a	s'
close	pass	fail
close	pass	close
close	pass	away
close	pass	fail
away	pass	away
away	pass	close
away	pass	close
away	pass	fail

(c) Apply the Q-learning algorithm to obtain state-action-value $Q(s, a)$ estimates. Estimates are initialized to 0.

(d) Determine the optimal policy according to the state-action-value estimates.