

INFO8006 Introduction to Artificial Intelligence

Exam of January 2023

Instructions

- The exam lasts for 4 hours.
- You are allowed to use a calculator during the exam, but documents of any kind are forbidden.
- The last two pages can be used for scratch work or for extra space. If you want work done there or on backs of pages to be graded, mention where to look **in big letters with a box around them**, on the page with the question.
- Write your last name, first name, and ULiège ID on the first page. Write only your ULiège ID on all the other pages.
- Before handing in your exam, **sort all the pages according to the page numbers** (even if you used additional pages to answer a question).

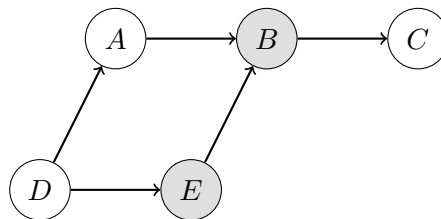
Good luck!

LAST NAME, FIRST NAME (in uppercase):

ULiège ID (s201234):

Question 1 [4 points] Multiple choice questions. Choose one of the four choices by filling in its circle. 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

- A goal-based planning agent ...
 - selects its actions using the Policy Evaluation algorithm and an admissible heuristic.
 - selects its actions on the basis of the current percept only, ignoring the rest of the percept history.
 - cannot decide which action to take next if its transition model is wrong.
 - can be implemented using the A^* algorithm in a discrete, deterministic and known environment.
- 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?
 - $h_1(s) = |i_s - x_s| + |j_s - y_s|$
 - $h_2(s) = \sqrt{(i_s - x_s)^2 + (j_s - y_s)^2}$
 - $h_3(s) = \mathcal{N}(i_s + j_s | \mu = x_s + y_s, \sigma^2 = 1)$
 - $h_4(s) = \min(h_1(s), h_2(s), h_3(s))$
- In H-Minimax, ...
 - if not looked deep enough, bad moves may appear as good moves because their consequences are hidden beyond the search horizon.
 - good moves will always appear as good moves, regardless of the search horizon.
 - bad moves will always be avoided by taking random actions with a probability ϵ .
 - the deeper in the tree the evaluation function is called, the more the quality of the evaluation function matters.
- Which of the following statement is always true?
 - $\mathbf{P}(X, Y) = \sum_a \mathbf{P}(X, a) \sum_b \mathbf{P}(Y, b)$
 - $\mathbf{P}(X) = \sum_a \sum_b \sum_c \mathbf{P}(X, a, b, c)$
 - $\mathbf{P}(X_1, X_2, \dots, X_n) = \mathbf{P}(X_1) \prod_{i=2}^n \mathbf{P}(X_i | X_{i-1})$
 - $\mathbf{P}(X|Y) = \frac{\mathbf{P}(Y|X)\mathbf{P}(Y)}{\mathbf{P}(X)}$
- Consider the Bayesian network shown below. We want to infer $\mathbf{P}(A|b, e)$ where A is the query variable, B and E are evidence variables, and C and D are hidden variables. Which of the following statements is true?



- $\mathbf{P}(A|b, e) \propto \sum_c \sum_d P(c|b) \mathbf{P}(b|A) P(b|e) \mathbf{P}(A|d) P(e|d) P(d)$
- $\mathbf{P}(A|b, e) \propto \sum_c \sum_d P(c|b) \mathbf{P}(b|A, e) \mathbf{P}(A|d) P(e|d)$
- $\mathbf{P}(A|b, e) \propto \sum_c \sum_d P(c|b) \mathbf{P}(b|A, e) \mathbf{P}(A|d) P(e|d) P(d)$
- $\mathbf{P}(A|b, e) \propto \sum_b \sum_e P(c|b) \mathbf{P}(A|b, e) \mathbf{P}(A|d) P(e|d) P(d)$

6. The Bayes filter requires the specification of ...
- a prior $\mathbf{P}(\mathbf{X}_0)$ and a transition model $\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t)$.
 - a transition model $\mathbf{P}(\mathbf{X}_{t+1}|\mathbf{x}_t)$ and an observation model $\mathbf{P}(\mathbf{E}_t|\mathbf{x}_t)$.
 - a prior $\mathbf{P}(\mathbf{X}_0)$, a transition model $P(\mathbf{X}_{t+1}|\mathbf{x}_t)$, and an observation model $P(\mathbf{E}_t|\mathbf{x}_t)$.
 - 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. In deep learning, a layer in a multi-layer perceptron is defined as ...
- $\mathbf{h} = \sigma(\mathbf{W}^T \mathbf{x} + \mathbf{b})$, where σ is the standard deviation function.
 - $\mathbf{h} = \sigma(\mathbf{W}^T \mathbf{x} + \mathbf{b})$, where σ is an activation function, such as the sigmoid function.
 - $\mathbf{h} = \sigma(\mathbf{W}^T + \mathbf{x} - \mathbf{b})$, where $\mathbf{W} \in \mathbb{R}^{d \times q}$ is matrix of weights.
 - $\mathbf{h} = \sigma(\mathbf{W}^T \mathbf{x} + \mathbf{b})$, where $\mathbf{b} \in \mathbb{R}^d$ is the most likely vector of hidden states given \mathbf{x} .
8. Arnaud is trying to perform gradient descent on a function $f(x)$ using the update

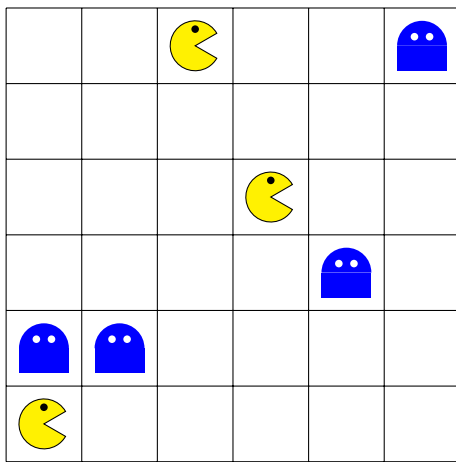
$$x_{t+1} := x_t - \frac{\partial f}{\partial x}(x_t).$$

Is gradient descent guaranteed to converge to the global minimum of f ?

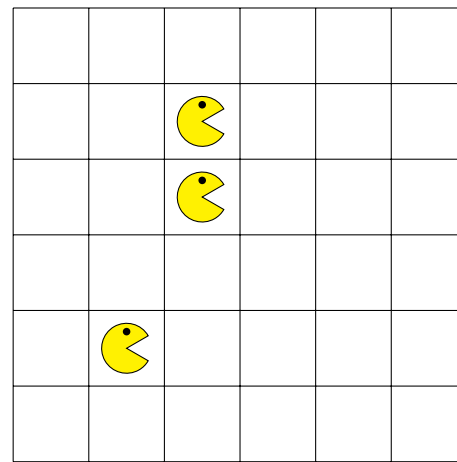
- Yes, since he's updating using the gradient of x .
 - Yes, but not for the reason above.
 - No, since he is updating x in the wrong direction.
 - No, but not for the reason above.
9. Which of the following is true? In Markov Decision Processes, ...
- the closer the discount factor γ to 0, the higher the utility of future rewards.
 - the closer the discount factor γ to 0, the longer Value Iteration may take to converge.
 - the closer the discount factor γ to 1, the greedier the optimal agent.
 - the closer the discount factor γ to 1, the longer Value Iteration may take to converge.
10. Assume that we run ϵ -greedy Q-learning until convergence. What is the optimal policy π^* we obtain for an arbitrary state s ?
- $\pi^*(s) = \arg \max_s V(s)$
 - $\pi^*(s) = \begin{cases} \arg \max_a Q(s, a) & \text{with probability } 1 - \epsilon \\ \text{random action} & \text{with probability } \epsilon \end{cases}$
 - $\pi^*(s) = \arg \max_a Q(s, a)$
 - $\pi^*(s) = \arg \max_s Q(s, a)$

Question 2 [4 points] Pacman and his friends have decided to combine forces and go on the offensive, and are now chasing ghosts instead! In a grid of size $M \times N$, Pacman and $P - 1$ of his friends are moving around to collectively eliminate all of the ghosts in the grid by stepping on the same square as each of them. Moving into the same square as a ghost will eliminate it from the grid. At every turn, Pacman and his friends may choose one of the four $\{north, south, east, west\}$ actions but may not collide with each other. In other words, any action that would result in two or more Pacmen occupying the same square will result in no movement for either of the Pacmen. Additionally, Pacman and his friends are *indistinguishable* from each other. There are a total of G ghosts, which are indistinguishable from each other and cannot move.

Treating this as a search problem, we consider each configuration of the grid to be a state, and the goal state to be the configuration where all ghosts have been eliminated from the board. Below is an example starting state, as well as an example of goal state.



Possible starting state.



Possible goal state.

Suppose that Pacman has no friends ($P = 1$).

- What is the size of the minimal state space representation? Explain your answer.
- Explain whether each of the following heuristics is admissible, consistent, neither, or both.
 - $h_1(s)$ = the sum of the Manhattan distances from Pacman to every ghost.
 - $h_2(s)$ = the number of ghosts times the maximum Manhattan distance between Pacman and any of the ghosts.
 - $h_3(s)$ = the number of remaining ghosts.

Suppose that Pacman has exactly one less friend than the number of ghosts ($P = G$). Keep in mind that Pacman and his friends cannot be distinguished from one another.

- What is the size of the minimal state space representation? Explain your answer.
- Explain whether each of the following heuristics is admissible, consistent, neither, or both.
 - $h_4(s)$ = the largest of the Manhattan distances between each Pacman and its closest ghost.
 - $h_5(s)$ = the smallest of the Manhattan distances between each Pacman and its closest ghost.
 - $h_6(s)$ = the number of remaining ghosts.
 - $h_7(s)$ = the number of remaining ghosts / P .

See Exercise 6 of the 1st exercise session (Searching).

Question 3 [4 points] You are waiting for an oral exam for which the professor is known to be moody. Some students have easy questions while others have hard ones. You learned from your seniors that if a student has hard questions, there is a 40% chance that the next student will also have hard ones. If a student has easy questions, there is 70% chance that the next student will also have easy ones. You observe the students that come out of the professor's office. They either look happy, neutral or sad. You assume that, if their questions were hard, there are 20% and 60% chance that they look respectively happy and sad. If their questions were easy, there are 50% and 30% chance that they look respectively happy and sad.

- (a) Define the components of a hidden Markov model (HMM), where the state $X_t \in \{\text{hard}, \text{easy}\}$ is the difficulty of the questions for the t -th student and the evidence $E_t \in \{\text{happy}, \text{neutral}, \text{sad}\}$ is the emotion you observe.
- (b) At some point, the professor decides that one in two students will come out of their office by another exit, which prevents you from observing them. Sketch a Bayesian network describing this new process.
- (c) Express, in terms of the HMM components, the distribution $\mathbf{P}(X_t|e_{1:t-2})$ that the t -th student will have hard questions given your observations $e_{1:t-2} = \{e_i : i \bmod 2 = 1 \mid i = 1, 2, \dots, t\}$ of half of the previous students. Separate the cases where t is even and odd.
- (d) Given the observations $e_{1:3:2} = (\text{sad}, \text{happy})$, compute $\mathbf{P}(X_3|e_{1:3:2})$.
- (e) The last student passing the exam arrives late and has not observed any of the emotions of the previous students. Since many students have passed before, there is a chance however that the professor has reached the stationary distribution of his mood. Compute this distribution.

See Exercise 6 of the 4th exercise session (Reasoning over time).

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 (x, y) 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 $\mathcal{N}(0, 1)$.

i	Starting room	Measured best time
1	(4, 5)	4
2	(5, 3)	6
4	(4, 1)	6
3	(3, 3)	8
5	(1, 2)	9

From the leaderboard, you wish to learn a heuristic approximating the best time to get to the exit, starting from room (x, y) . You decide to use a small neural network as approximator, described by the following parametric function,

$$h(x, y; \phi) = \text{ReLU}(xw_1 + yw_2 + w_3) + \text{ReLU}(xw_4 + yw_5 + w_6)$$

$$\text{ReLU}(x) = \max(x, 0),$$

where $\phi = (w_1, w_2, w_3, w_4, w_5, w_6)$ is the set of parameters/weights of the neural network.

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

Set	w_1	w_2	w_3	w_4	w_5	w_6
ϕ_A	-1	2	3	0	-1	4
ϕ_B	-1	0	4	-2	1	5
ϕ_C	-2	1	4	1	-1	6

ϕ_C


- (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? Motivate your answer.

ϕ_A

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

See Exercise 7 of the 5th exercise session (Learning) for detailed answers.

Question 5 [4 points] Consider the grid-world given below and an agent who is trying to learn the optimal policy. The agent starts from the bottom-left corner and can take the actions **north** (N), **south** (S), **west** (W) and **east** (E). Rewards are only awarded for reaching the terminal (shaded) states. You observe the following trials, whose trajectories are sequences of tuples $(s_t^i, r_t^i, a_t^i, s_{t+1}^i)$. We assume a discount factor $\gamma = 1$.

3	-6	-1	+5
2			
1		-8	+3
	1	2	3

t	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
0	$(1, 1), 0, N, (1, 2)$	$(1, 1), 0, N, (1, 2)$	$(1, 1), 0, N, (1, 2)$	$(1, 1), 0, N, (1, 2)$	$(1, 1), 0, N, (1, 2)$
1	$(1, 2), 0, E, (2, 2)$	$(1, 2), 0, E, (2, 2)$	$(1, 2), 0, E, (2, 2)$	$(1, 2), 0, E, (2, 2)$	$(1, 2), 0, E, (2, 2)$
2	$(2, 2), 0, N, (2, 3)$	$(2, 2), 0, E, (3, 2)$	$(2, 2), 0, S, (2, 1)$	$(2, 2), 0, E, (3, 2)$	$(2, 2), 0, E, (3, 2)$
3	$(2, 3), -1, \emptyset, \emptyset$	$(3, 2), 0, N, (3, 3)$	$(2, 1), -8, \emptyset, \emptyset$	$(3, 2), 0, W, (2, 2)$	$(3, 2), 0, S, (3, 1)$
4		$(3, 3), +5, \emptyset, \emptyset$		$(2, 2), 0, N, (2, 3)$	$(3, 1), +3, \emptyset, \emptyset$
5				$(2, 3), -1, \emptyset, \emptyset$	

- Describe the problem setting of reinforcement learning formally.
- Perform direct utility estimation of the expected utilities $V^\pi(s)$, given the four first trials.
- Update the estimated expected utilities with respect to the fifth trial using temporal-difference learning. Assume a learning rate $\alpha = 0.5$.
- Consider the scenario where the agent continues to interact with the environment while following its current policy and updating it with new trials. Explain if this will ultimately lead to the optimal policy. If so, provide reasoning. If not, identify necessary changes.

See Exercise 1 of the 7th exercise session (Reinforcement Learning).