INFO8006 Introduction to Artificial Intelligence

Exam of August 2025

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 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.

1. A model-based reflex agent ...
   - ◯ computes a plan once and then executes it eyes closed.
   - ◯ stores all its percepts in a database.
   - ◯ maintains some internal state of how the world is now.
   - ◯ makes use of a model of how the world evolves to predict the consequences of its actions.
2. In H-Minimax with depth limit d and evaluation function h, if we increase d, then ...
   - ◯ the quality of play always strictly improves.
   - ◯ the evaluation function h becomes admissible.
   - ◯ the compute time per move decreases.
   - ◯ more nodes must be evaluated.
3. Consider a game tree where MAX is to move at the root node. If we multiply all terminal values by -1 and then add 5 to each, then the Minimax algorithm will ...
   - ◯ select the same move as before.
   - ◯ select the best move for MIN.
   - ◯ select some other move than before, but not necessarily the best move for MIN.
   - ◯ fail to terminate due to negative values.
4. If $P(A) = 0.3$, $P(B) = 0.4$, and A and B are independent, then ...
   - ◯ $P(A \cap B) = 0.12$.
   - ◯ $P(A \cap B) = 0.3$.
   - ◯ $P(A \cap B) = 0.7$.
   - ◯ $P(A \cap B) = 0.88$.
5. In a Bayesian network with variables A, B, C, and D, if we have edges $A \rightarrow B$, $A \rightarrow C$, $B \rightarrow D$ and $C \rightarrow D$, then ...
   - ◯ $P(A, B, C, D) = P(A)P(B)P(C)P(D)$.
   - ◯ $P(A, B, C, D) = P(D)P(C|D)P(B|D)P(A|B, C)$.
   - ◯ $P(A, B, C, D) = P(A)P(B|A)P(C|A)P(D|B, C)$.
   - ◯ $A \perp D \mid B$.
6. The Viterbi algorithm ...
   - ◯ computes $P(X_t|e_{1:t})$ recursively.
   - ◯ finds $\arg\max_{x_{1:t}} P(x_{1:t}|e_{1:t})$.
   - ◯ has exponential time complexity in the length of the sequence $e_{1:t}$.
   - ◯ requires the transition and sensor models to be Gaussian.
7. The ReLU activation function ...
   - ◯ outputs values in $[0, 1]$.
   - ◯ is primarily used in the output layer of a neural network.
   - ◯ has derivative 0 for negative inputs and 1 for positive inputs.
   - ◯ is a linear function.

8. Convolutional neural networks are particularly well suited for ...
    - ◯ processing tabular data such as spreadsheets.
    - ◯ processing images and spatial data.
    - ◯ processing sequences of words such as text.
    - ◯ processing graphs and networks.
9. Policy Iteration for MDPs ...
    - ◯ often requires fewer iterations than Value Iteration.
    - ◯ requires solving non-linear equations.
    - ◯ may cycle indefinitely.
    - ◯ improves the current policy by sampling the state-action values.
10. Temporal-difference learning differs from Monte Carlo evaluation in that ...
    - ◯ it updates the value function after each step.
    - ◯ it requires a model of the environment.
    - ◯ always converges faster.
    - ◯ does not require exploration mechanisms.

Question 2 [4 points] Consider a two-player turn-taking game called "Hidden Treasure Hunt" played on a 1×7 board. Each player has a token that starts at opposite ends: Player A starts at position 1 and Player B starts at position 7. There are two hidden treasures on the board whose locations are unknown to both players until a player lands on them.

The rules of the game are as follows:

- Players alternate turns, starting with Player A.
- On their turn, a player can move their token 1 or 2 positions towards the opponent's starting position (and cannot move backwards).
- If a player lands on a treasure, they collect it and gain +5 points. A treasure can only be collected once.
- If players land on the same position, the moving player pushes the other player back 1 position (if possible) and gains +2 points.
- If a player is pushed to a position with a treasure, they collect it and gain +5 points.
- The game ends when both treasures are collected or when a player reaches the opponent's starting position, with that player gaining +3 bonus points.

(a) Formalize the game as a search problem.

(b) Explain whether Minimax is sufficient, or not, to solve this game optimally. Motivate your answer.

(c) Assume the treasures are located at positions 4 and 6 and none of them have been collected yet. Player A is at position 3 and Player B is at position 5. Player A is to move next and the current score is 0 for both players. Draw the game tree for the next 3 moves (A, then B, then A again).

(d) If Player A is omniscient and knows the exact locations of the treasures, what is the best move for Player A in the situation described in the previous part? Explain your reasoning.

Question 3 [4 points] A medical diagnostic system uses a Bayesian network to assess patients with respiratory symptoms. The network models the relationships between the following variables:

- Smoking S: whether the patient is a smoker (yes or no).
- Pollution P: whether the patient is exposed to high pollution levels (yes or no).
- Lung disease L: whether the patient has a lung disease (yes or no).
- Coughing C: whether the patient has a persistent cough (yes or no).
- Fatigue F: whether the patient experiences fatigue (yes or no).

The network structure is $S \rightarrow L \leftarrow P$, $L \rightarrow C$, and $L \rightarrow F$. Its conditional probability distributions are defined as follows:

- $P(S = \text{yes}) = 0.3$.
- $P(P = \text{yes}) = 0.4$.
- $P(L = \text{yes}|S = \text{yes}, P = \text{yes}) = 0.8$, $P(L = \text{yes}|S = \text{yes}, P = \text{no}) = 0.5$, $P(L = \text{yes}|S = \text{no}, P = \text{yes}) = 0.3$, $P(L = \text{yes}|S = \text{no}, P = \text{no}) = 0.1$.
- $P(C = \text{yes}|L = \text{yes}) = 0.7$, $P(C = \text{yes}|L = \text{no}) = 0.2$.
- $P(F = \text{yes}|L = \text{yes}) = 0.6$, $P(F = \text{yes}|L = \text{no}) = 0.3$.

(a) Write the factorization of the joint probability distribution $P(S, P, L, C, F)$ according to the Bayesian network structure.

(b) Determine whether the following conditional independence statements are true, false, or cannot be determined.

(i) $S \perp P$.

(ii) $S \perp P|F$.

(iii) $C \perp F|L$.

(iv) $S \perp F|L$.

(c) A patient arrives. They live in a polluted area and have a persistent cough. Compute $P(L = \text{yes}|P = \text{yes}, C = \text{yes})$.

(d) Consider a routine screening X-ray X that can detect either lung disease or unrelated abnormalities U, where U is independent of the other variables.

(i) Draw the updated Bayesian network structure including the new variables X and U.

(ii) A patient's X-ray shows abnormalities (X = yes). If we then learn the patient has a persistent cough (C = yes), explain qualitatively (without computing exact probabilities) whether this would increase, decrease, or not change our belief in U = yes.

Question 4 [4 points] A master student in electrical engineering wants to verify her circuit knowledge before the exam. To do so, she builds an RLC circuit that she will analyze. Unfortunately, she is only able to probe the voltage across the resistor and she knows that the components are not perfect and that the response of the system might not follow exactly the theoretical equations mainly due to noise. Fortunately, she also has to pass the AI exam and remembers that she can estimate the state of a stochastic dynamical system from partial observation using a Kalman filter (under some assumptions).

From what she remembers, we can model the evolution of the current in the circuit as

$$\begin{aligned}
I_t &= I_{t-1} + \Delta t\, \dot{I}_{t-1} + \frac{1}{2}\Delta t^2\, \ddot{I}_{t-1}, \\
\dot{I}_t &= \dot{I}_{t-1} + \Delta t\, \ddot{I}_{t-1}, \\
\ddot{I}_t &= -\omega_0^2\, I_{t-1} - 2\alpha\, \dot{I}_{t-1},
\end{aligned}$$

where I_t, $\dot{I}_t$ and $\ddot{I}_t$ are the current and its two first order derivatives at timestep t, Δt is the time elapsed between $t-1$ and t, ω_0 is the angular resonance frequency of the circuit, and α is the attenuation coefficient. You estimate that the system starts at a current of $1.0 \pm 0.1\,\mathrm{A}$ with negligible derivatives $0.0 \pm 0.1\,\mathrm{A\,s^{-1}}$ and $0.0 \pm 0.1\,\mathrm{A\,s^{-2}}$. Finally, you assume that the voltmeter across the resistor follows an unbiased Gaussian distribution with some noise defined in the manual. She also remembers that the current inside a resistor is linked to the voltage via the relation

$$V_t = RI_t$$

where R is the resistance value.

(a) You wish to predict the state of the circuit given the the voltmeter measures. Define and instantiate the components of a Kalman filter for this problem.

(b) Express the distribution $p(x_t|e_{1:t})$ with respect to the components defined previously. Use the Gaussian identities in Appendix A if needed.

(c) Represent the transition and sensor models as a dynamic Bayesian network.

Question 5 [4 points] Let us consider a neural network f with one hidden layer taking as input a scalar $x \in \mathbb{R}$ and producing as output a positive scalar $\hat{y} = f(x; \theta) \in \mathbb{R}$. The neural network is defined as

$$f(x; \theta) = \text{ReLU}(w_4 \text{ReLU}(w_0 x + w_1) + w_5 \text{ReLU}(w_2 x + w_3)),$$

where $\theta = (w_0, w_1, w_2, w_3, w_4, w_5, w_6)$ is the set of parameters and $\text{ReLU}(x) = \max(x, 0)$ is the rectified linear unit function.

(a) Draw the computation graph representing the neural network and the flow of information from inputs to outputs. Your diagram should be a directed graph that follows the following conventions:

- circled nodes correspond to variables (input, output, parameters or intermediate variables),
- squared nodes correspond to primitive operations (addition, multiplication, ReLU) and produce an intermediate variable as output,
- directed edges correspond to the flow of information, from inputs to outputs.

(b) For $\theta = (1, -2, -1, -2, 1, 2, 0)$, draw the function $\hat{y} = f(x; \theta)$ for $x \in [-10, 10]$.

(c) Using the data point $(x, y) = (5, 5)$ and the value of θ given above, we want to fine-tune the parameter w_0 such that $f(x; \theta)$ produces a more accurate prediction of y.

(i) Evaluate the squared error loss at the data point $(x, y) = (5, 5)$ and the current value of θ.

(ii) Derive an expression for the derivative of the squared error loss with respect to w_0.

(iii) Update the parameter w_0 using one step of gradient descent with a learning rate $\gamma = 0.01$.

(iv) Verify that the value of the loss function has decreased after the update.

Appendix A. Gaussian identities (Särkkä, 2013).

(a) If $\mathbf{x}$ and $\mathbf{y}$ have the joint Gaussian distribution

$$p\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \mathcal{N}\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \middle| \begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}, \begin{pmatrix}\mathbf{A} & \mathbf{C}\\ \mathbf{C}^T & \mathbf{B}\end{pmatrix}\right),$$

then the marginal and conditional distributions of $\mathbf{x}$ and $\mathbf{y}$ are given by

$$\begin{aligned} p(\mathbf{x}) &= \mathcal{N}(\mathbf{x}|\mathbf{a}, \mathbf{A}) \\ p(\mathbf{y}) &= \mathcal{N}(\mathbf{y}|\mathbf{b}, \mathbf{B}) \\ p(\mathbf{x}|\mathbf{y}) &= \mathcal{N}(\mathbf{x}|\mathbf{a} + \mathbf{C}\mathbf{B}^{-1}(\mathbf{y} - \mathbf{b}), \mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^T) \\ p(\mathbf{y}|\mathbf{x}) &= \mathcal{N}(\mathbf{y}|\mathbf{b} + \mathbf{C}^T\mathbf{A}^{-1}(\mathbf{x} - \mathbf{a}), \mathbf{B} - \mathbf{C}^T\mathbf{A}^{-1}\mathbf{C}). \end{aligned}$$

(b) If the random variables $\mathbf{x}$ and $\mathbf{y}$ have Gaussian probability distributions

$$\begin{aligned} p(\mathbf{x}) &= \mathcal{N}(\mathbf{x}|\mathbf{m}, \mathbf{P}) \\ p(\mathbf{y}|\mathbf{x}) &= \mathcal{N}(\mathbf{y}|\mathbf{H}\mathbf{x} + \mathbf{u}, \mathbf{R}), \end{aligned}$$

then the joint distribution of $\mathbf{x}$ and $\mathbf{y}$ is Gaussian with

$$p\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} = \mathcal{N}\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \middle| \begin{pmatrix}\mathbf{m}\\ \mathbf{H}\mathbf{m} + \mathbf{u}\end{pmatrix}, \begin{pmatrix}\mathbf{P} & \mathbf{P}\mathbf{H}^T\\ \mathbf{H}\mathbf{P} & \mathbf{H}\mathbf{P}\mathbf{H}^T + \mathbf{R}\end{pmatrix}\right).$$

Extra page 1 / 2.

Extra page 2 / 2.