<table>
<tr><td>  </td><td>Student’s full name: .....<br>Students’ code: …….....<br>Student’s class: .....</td></tr>
<tr><td>ARTIFICIAL INTELLIGENCE<br><br>Date: April 21, 2022<br>Duration: 100 minutes</td><td colspan="2">FINAL EXAMINATION<br><br><u>Invigilator 1</u><br><br><u>Invigilator 2</u></td></tr>
</table>

- *Writing test consists of ---- **PAGES (Answer Sheet included).***
- *Students are required to use provided **blue-color** pens to write responses.*
- *Learning materials, dictionaries and any kinds of electronic devices are **NOT** allowed during the test time.*

**Student Full Name: .....**

**HCMUS Student ID: …….....**

**Student's class: .....**

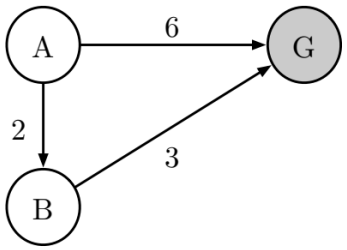
**FINAL EXAMINATION**

**ANSWER SHEET**

| **Result** | **Examiners** |
|---|---|
| | |

## PART I: Problem solving with Search Strategies (2pts)

**Question 1 (1pt)** *Heuristic admissibility and consistency*. Consider the search problem shown on the left. It has three states and three directed edges. **A** is the start node and **G** is the goal node. To the right, four different heuristic functions are defined, numbered I through IV.




| | h($A$) | h($B$) | h($G$) |
|---|---|---|---|
| I | 4 | 1 | 0 |
| II | 5 | 4 | 0 |
| III | 4 | 3 | 0 |
| IV | 5 | 2 | 0 |

For each heuristic function, check whether it is admissible and whether it is consistent. Given an explanation if it is inadmissible and/or inconsistent.

| Heuristic | Admissible? | Consistent? | Explanation |
|---|---|---|---|
| I | | | |
| II | | | |
| III | | | |
| IV | | | |

**Question 2 (1pt)** *A* algorithm*. Is graph search A* using an admissible heuristic guaranteed to be optimal? Justify your answer.

## PART II: Knowledge representation (4pts)

**Question 3 (1pt)** *Backward chaining in propositional logic*. Consider the knowledge base shown aside. Apply **Backward chaining** to answer the query **"KB entails D?"**

For every step, state the sentences required for inference.
Note that you can choose any sentence of your choice, not necessary to exhaustively scan the knowledge base.

(1) $A \wedge B \rightarrow D$
(2) $Q \wedge R \rightarrow A$
(3) $\neg P \vee Q$
(4) $\neg Q \vee \neg B \vee R$
(5) $B$
(6) $P \leftrightarrow B$

**Question 4 (1pt)** *Unification*. Find the MGU for each of the following pairs of clauses. If there exists such a MGU, write the corresponding substitution θ. Otherwise, write No MGU.

a) F( x, G( F(A), u ) ) and F( G(u, v), x ) where A is a constant

b) P( A, x, F( G(y) ) ) and P( z, F(z), F(A) ) where A is a constant

**Question 5 (2pts)** *Refutation resolution in first-order logic*. Consider the following passage.

*"Every student takes Databases or Artificial Intelligence. Every student who takes Artificial Intelligence knows Python. John is a student who did not take Databases".*

a) Build a FOL knowledge base from the text above, using only the given predicates

- STUDENT(x): x is a student
- TAKES(x, y): x takes y
- KNOWS(x, y): x knows y

b) Prove via refutation resolution that "*John knows Python.*"

*Rewrite the given clauses in an appropriate form*

*Write down new clauses generated during the proof. For each new clause, state the source clauses required and the corresponding substitutions.*

## PART III: Machine learning (4pts)

**Question 6 (2pts)** *Decision tree*. Consider the following set of training examples. There are two features: A and B. The binary target variable (also known as the class label) is Y ∈ {1, 0}. You came up with a decision tree classifier that is shown in Figure 1.

| A | B | Y |
|---|---|---|
| 3 | 3 | 0 |
| 3 | 6 | 0 |
| 5 | 5 | 0 |
| 7 | 2 | 0 |
| 9 | 7 | 0 |
| 9 | 3 | 0 |
| 2 | 7 | 1 |
| 7 | 6 | 1 |
| 8 | 5 | 1 |
| 10 | 5 | 1 |
| 10 | 7 | 1 |
| 11 | 2 | 1 |

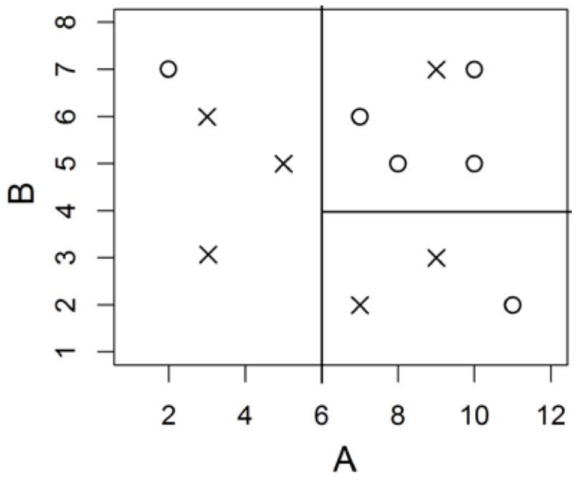
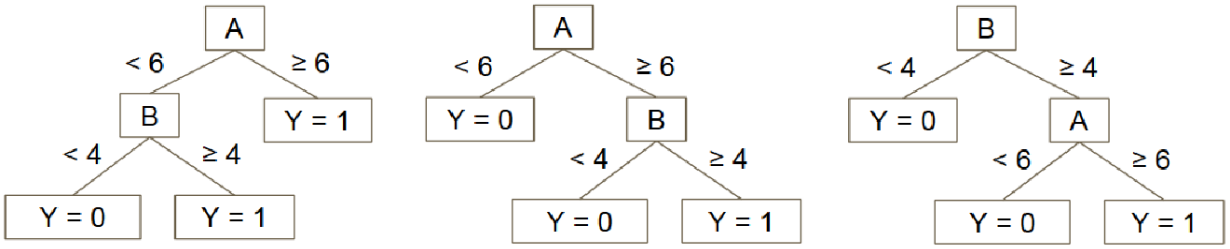



**Figure 1.**

a) In Figure 1, do circles represent Y = 0 or Y = 1? **(Write 0 or 1)** .....

b) Consider the three decision trees shown below (labeled 1, 2, and 3).Which tree represents the same decision tree as that which is shown in the above figure? Note that when a leaf node is not pure, take the majority class as its label. **(Write 1, 2, or 3)** .....




**(1)** **(2)** **(3)**

c) Classify the new examples below using the decision tree shown in Figure 1. When a decision region is not pure, take the majority class as its label. Indicate your answer as a circle or an X.

i. What class is [A = 2, B = 7] ? **(Write circle O or X)** .....

ii. What class is [A = 10, B = 2]? **(Write circle O or X)** .....

*You may refer to the formulas shown in the last page if necessary.*

**Question 7 (2pts)** *Artificial neural network*. The table aside shows the input and output of a boolean function f(x, y).

| x | y | f(x,y) |
|---|---|---|
| 1 | 1 | 0 |
| 0 | 0 | 0 |
| 1 | 0 | 1 |
| 0 | 1 | 0 |

a) Is the function linearly separable? .....

b) Let us represent the function with a perceptron. Assume that learning rate is 0.1, threshold is 0.5 and intial weights are $\mathbf{w_x = 1}$ and $\mathbf{w_y = -1}$. Show how the weights change in the first epoch. *You may refer to the formulas shown in the last page if necessary.*

| Iteration | x | y | f(x,y) | Initial weights | | Actual output | Error e | Final weights | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | $w_x$ | $w_y$ | | | $w_x$ | $w_y$ |
| 1 | 1 | 1 | 0 | 1 | -1 | | | | |
| 2 | 0 | 0 | 0 | | | | | | |
| 3 | 1 | 0 | 1 | | | | | | |
| 4 | 0 | 1 | 0 | | | | | | |

c) Could the function be completely represented with a perceptron? If yes, show the weights (without any calculation). If not, explain why not in 1-2 sentences.

-The end-

The Entropy measure evaluates the uncertainty of a random variable $V$ with values $v_k$.

$$H(V) = \sum_k P(v_k)\, log_2 \frac{1}{P(v_k)} = -\sum_k P(v_k)\, log_2\, P(v_k)$$

- $v_k$ is a class in $V$ (e.g., yes/no in binary classification)
- $P(v_k)$ is the proportion of the number of elements in class $v_k$ to the number of elements in $V$

Functions for perceptron

- The actual output is calculated as follows: $y(p) = \text{step}[(\sum_{i=1}^{n} x_i(p) \times w_i(p)) - \theta]$

  where $n$ is the number of inputs, $x_i$ is the $i$th input, $\theta$ is the threshold, and step is the activation function.
- The step activation function: $step(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & if\ x < 0 \end{cases}$
- Update the weights: $w_i(p+1) = w_i(p) + \Delta w_i(p)$, where $\Delta w_i(p)$ is the weight correction at iteration $p$
- The delta rule determines how to adjust the weights by $\Delta w_i(p) = \alpha \times x_i(p) \times e(p)$

  where $\alpha$ is the learning rate ($0 < \alpha < 1$) and $e(p) = Y_d(p) - Y(p)$