Artificial Intelligence
COMS4033A/7044A
Class Test

9 April 2025 09h00 – 11h00, MSL006

Name: Memo

Student Number: _____

For marking purposes only

Question 1	
Question 2	
Question 3	
Question 4	
Question 5	
Total	

Instructions

- Answer all questions in pen. **Do not write in pencil.**
- This test consists of 10 pages. Ensure that you are not missing any pages.
- This is a **closed-book** test: you may not consult any written material or notes.
- You are allocated 2 hours to complete this test.
- There are 5 questions and 50 marks available.
- Ensure your cellphone is switched off.

Question 1 Agents and Search [10 Marks]

1. Explain what it means for an agent to be categorised as "rational". [1]

   A rational agent ACT in its environment according to what it has PERCEIVED

   in order to MAXIMISE its expected PERFORMANCE MEASURE

   1 for all, 0.5 for some

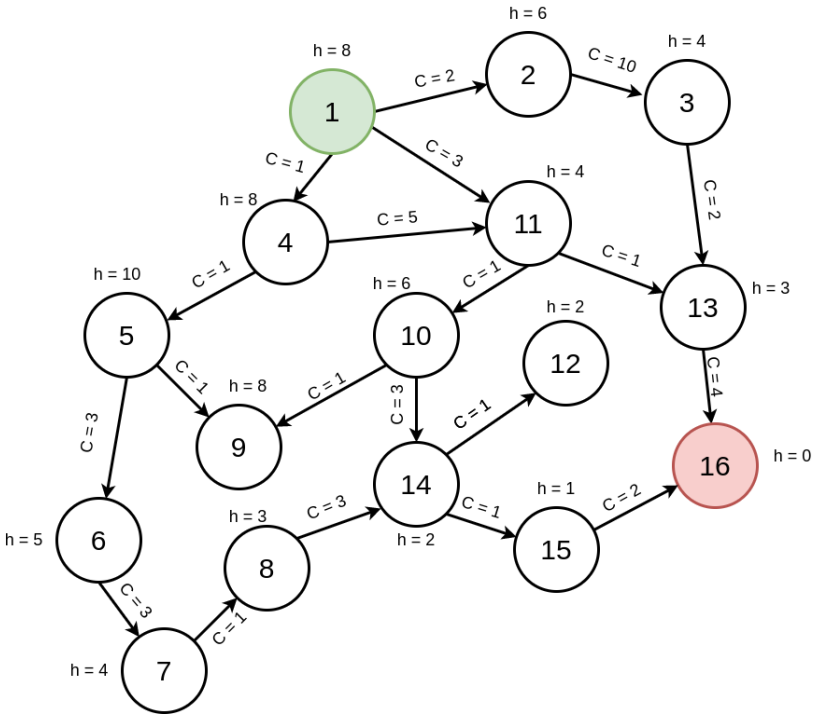
2. Consider a robot operating on Mars with the purpose of collecting rock samples. We can categorise this task along the following lines:

   0.5 each

   (a) partially observable or fully observable — PO
   (b) deterministic or stochastic — stochastic
   (c) episodic or sequential — sequential
   (d) static or dynamic — static/dynamic (arguments for both I guess)
   (e) discrete or continuous — continuous
   (f) single agent or multiagent — signle

   For each option above, circle the correct one. [3]

1. Let C be the cost of traversing an edge between nodes, and let h be the heuristic estimate of a given node to the goal. Consider the graph below, where nodes are labelled with integers and directed edges are indicated by lines with arrows. Each edge is labelled with a cost C, and each node with a heuristic value h.



Assume that we wish to search for the goal node 16 starting at node 1 using the uniform-cost search algorithm (UCS). Further, assume that if there are ties in the priority queue, they are resolved by removing the node with the smallest label first (e.g. if two nodes 2 and 1 both have priority 10, then 1 will be removed before 2).

Given this, detail each step of the search algorithm. At each iteration, describe what node is popped from the queue and then what nodes are added to the queue subsequently (and their corresponding priority). [4]

Start at 1. Add 2(2), 11(3), 4(1)

Pop 4. Add 5(2)
Pop 2. Add 3(12)

Pop 5. Add 9(3),6(5)
Pop 9. Pop 11. Add 10(4), 13(4)

Pop 10. Add 14(7)
Pop 13. Add 16(8)

Pop 6. Add 7(8)

Pop 14. Add 12(8), 15(8)
Pop 7. Add 8(9)

Pop 12. Pop 15. Pop 16. Stop.

2. When employing alpha-beta pruning, many chess engines will try to expand the children of a node using hand-coded rules, such as by expanding moves that capture an opponent's piece first. What advantage does this give? [1]

Trying better moves first can lead to more pruning (anything that mentions how order can affect pruning)

3. Describe under what conditions uniform-cost search is identical to breadth-first search. [1]

when all actions have the same cost (give mark if they pick a particular cost)

Question 2 Game Theory [10 Marks]

1. Your lecturer is lazy and has decided to use ChatGPT to generate questions. ChatGPT outputs the following payoff matrix with a description of the various Nash equilibria:

	A	B	C	D
E	$(3,2)$	$(0,0)$	$(2,3)$	$(1,1)$
F	$(0,0)$	$(2,3)$	$(1,1)$	$(3,2)$
G	$(1,1)$	$(3,2)$	$(0,0)$	$(2,3)$
H	$(2,3)$	$(1,1)$	$(3,2)$	$(4,4)$

```
In this payoff matrix:

Player 1 has four strategies: E, F, G, and H.
Player 2 has four strategies: A, B, C, and D.
There are multiple Nash equilibria in this game:

(E, A): In this equilibrium, Player 1 chooses strategy E,
  and Player 2 chooses strategy A.

(F, B): In this equilibrium, Player 1 chooses strategy F,
  and Player 2 chooses strategy B.

(G, C): In this equilibrium, Player 1 chooses strategy G,
  and Player 2 chooses strategy C.

(H, D): In this equilibrium, Player 1 chooses strategy H,
  and Player 2 chooses strategy D.
```

For each claimed equilibrium, state whether ChatGPT is correct and motivate your answer [2]

EA is not - P2 can do better by picking C (0.5)

FB is not - P1 can rather pick G

GC is not - P1 or P2 can pick any others

HD is - neither P1/P2 can do better by switching

0.5 each

2. Consider the following payoff matrix representing a game between two players A (the "row" player) and B (the "column" player).

	B_1	B_2	B_3	B_4
A_1	$(1,7)$	$(2,5)$	$(7,2)$	$(0,1)$
A_2	$(5,2)$	$(3,3)$	$(5,2)$	$(0,1)$
A_3	$(7,0)$	$(2,5)$	$(0,4)$	$(0,1)$
A_4	$(0,0)$	$(0,2)$	$(0,9)$	$(9,-1)$

3. Execute the *Iterated Elimination of Dominated Strategies* algorithm on the payoff matrix to find the outcome, beginning with player B. At each step, list the strategy that is eliminated and why it is eliminated, as well as the final outcome. [3]

B4 dominated by B3/B2. Delete

A4 dominated by A1/A2/A3. Delete

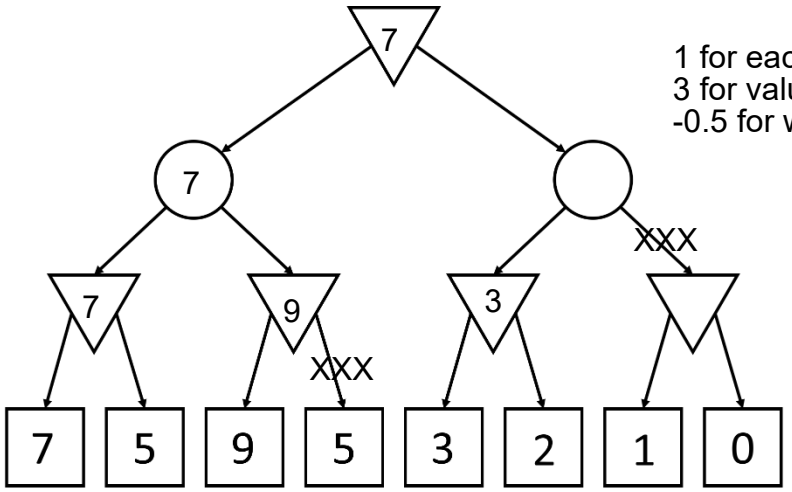
B3 dominated by B2. Delete 0.5 for each step

A1 dominated by A2/A3. Delete must mention outcome to get last 0.5

B1 dominated by B2. Delete

A3 dominated by A2. Delete. Leaves A2B2 = (3,3)

4. Consider the game tree below. Triangles represent the maximising player, circles the minimising player and rectangles terminal nodes. Execute alpha-beta pruning on the tree. Annotate the tree by writing down the value backed up to each node in the tree (including the root), and indicate where pruning takes place by crossing out the relevant edges. [5]



1 for each correct prune
3 for values
-0.5 for wrong prunes (max -2)

Question 3 Logic [10 Marks]

1. Describe how first-order logic is used to represent knowledge. [2]

   Using objects, predicates and functions

   [2 for all three, 1 for two, 0.5 for one]

2. Briefly describe what is meant by the *grounding problem*. [1]

   The grounding problem refers to the challenge of connecting

   abstract symbols or words used by an AI system to their real-world meanings

3. What does it mean for a knowledge base K to *entail* sentence β? [1]

   beta is true in EVERY model in which ALL sentences in K are true

4. With reference to physical sensors, argue why propositional logic is insufficient in general for modelling real-world problems. [1]

   They are faulty/noisy and so cannot know for certain if some condition is true

   (something about noise/faults and certainty)

5. Describe when one would make use of a Skolem constant. [1]

   When grounding the existential operator

6. Encode the statement "Everyone who Mary loves loves someone who is happy" in first-order logic. [2]

   for all x, love(Mary, x) -> exists y (love(x, y) AND happy(y))

   1 for LHS of implication, 1 for right

7. Alice wishes to encode the English sentence "all students have phones" in first-order logic. She writes this as $\forall x Student(x) \wedge HasPhone(x)$. Explain the fault in her reasoning and then write down the correct logical sentence. [2]

   Means that all objects are students who have phones 1

   For all x, student(x) -> hasphone(x)

Question 4 Uncertainty [7 Marks]

1. Describe the difference in how frequentists and subjectivists interpret the concept of probability. [2]

   Frequentist: relative frequency of repeated events (something about counts/repetitions)

   Bayesian: degree of belief of event occuring (something about belief)

   1 mk each

2. Suppose $P(A), P(\neg A), P(B|A)$ and $P(B|\neg A)$ are known. Find an expression for $P(A|B)$ in terms of these four probabilities. [2]

   P(A|B) = numerator/denom

   numerator = P(B|A)P(A) [1]

   denom = P(B|A)P(A) + P(B|not A)P(not A) [1]

3. Consider the joint probability distribution for three Boolean random variables below:

A	**B**	**C**	Prob
T	T	T	0.108
T	T	F	0.072
T	F	T	0.012
T	F	F	0.008
F	T	T	0.016
F	T	F	0.144
F	F	T	0.064
F	F	F	0.576

   (a) Compute $P(B = True)$. Show your working. [1]

   0.34

   (b) Compute $P(B = True|A = False)$. Show your working. [2]

   P(B,-A) = P(B AND -A) / P(-A)

   P(B AND -A) = 0.16 (1 mk)

   P(-A) = 0.8 0.5 mk
   Answer = 0.2 0.5 mk

Question 5 Bayesian Networks [13 Marks]

1. In the context of Hidden Markov Models, what is the difference between determining the most likely path and filtering? Answer this question by writing down the probabilities that each approach attempts to calculate (and define your notation), and then suggest one real-world application for each of them. [3]

2. We wish to model the following three real-world events: whether it is raining (`rain`), whether the sprinkler system is on (`sprinkler`), and whether the grass is wet (`wet`). We occasionally water the grass with the sprinklers when it is too dry. However, unless we forget to disable the system, we never run the sprinkler system when it is raining, since this would waste water.

   (a) Draw the corresponding Bayesian network for the above three variables. [2]

   rain to wet and sprinkler
   sprinkler to wet

   2 if correct, 1 if partial

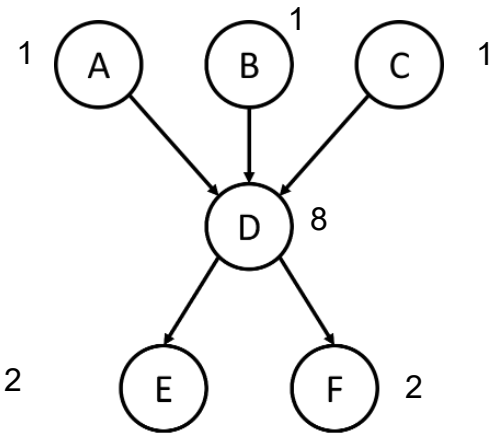
   (b) Write down the factorisation of joint probability distribution given the Bayesian network's structure in part a). [1]

   P(R)P(S|R)P(W|S,R)

   (or answer according to a)

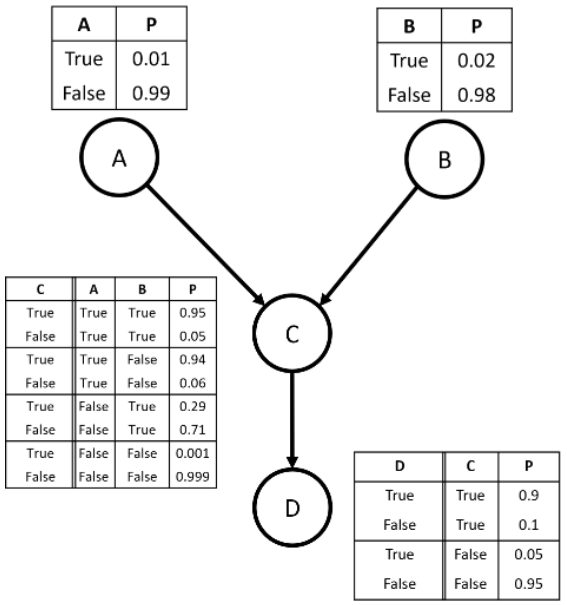
3. A Hidden semi-Markov model (HSMM) is similar to a hidden Markov model, but the transition model is sem-Markov. This means that the next time step depends on more than just the previous timestep. Draw the graphical representation of an HSMM where the current state depends on the previous two states, but the observation model is still Markov. [2]

4. Consider the Bayesian network below, where each node is a Boolean random variable. How many entries are required to complete the CPTs for this network? [1]



15 entries

5. Consider the Bayesian network and associated CPTs below:



A	P
True	0.01
False	0.99

B	P
True	0.02
False	0.98

C	A	B	P
True	True	True	0.95
False	True	True	0.05
True	True	False	0.94
False	True	False	0.06
True	False	True	0.29
False	False	True	0.71
True	False	False	0.001
False	False	False	0.999

D	C	P
True	True	0.9
False	True	0.1
True	False	0.05
False	False	0.95

6. We wish to estimate $\Pr(c|\neg a, \neg b, \neg d)$ using rejection sampling. Below are some samples that have been produced by sampling from the network. First, cross out the samples that would be rejected by rejection sampling. Then, write down the estimate for $\Pr(c|\neg a, \neg b, \neg d)$ using those samples. [3]

Sample 1: $\neg a,\ \neg b,\ c,\ d$

Sample 2: $a,\ \neg b,\ \neg c,\ d$

Sample 3: $\neg a,\ \neg b,\ c,\ \neg d$

Sample 4: $a,\ \neg b,\ \neg c,\ \neg d$

Sample 5: $\neg a,\ b,\ c,\ d$

Sample 6: $\neg a,\ \neg b,\ \neg c,\ \neg d$

cross out 1, 2, 4, 5

0.5 each, -0.5 for each wrong

Remaining = 0.5 (1 mk)

---

7. Consider an email spam classifier where S denotes that the email is spam and $W_1, W_2, \ldots, W_n$ denote the observed words in the email. Using Bayes' rule, the probability that an email is spam given the observed words is given by:

$$\Pr(S \mid W_1, W_2, \ldots, W_n) = \frac{\Pr(W_1, W_2, \ldots, W_n \mid S)\Pr(S)}{\Pr(W_1, W_2, \ldots, W_n)}.$$

Explain why, in the context of spam classification, we do not need to explicitly compute the denominator $\Pr(W_1, W_2, \ldots, W_n)$ when deciding if an email is spam. [1]

We care only about the decision rule / most likely, so we don't need the exact prob / denom (or spam and not spam would have the same denom, so their numerator is proportional) (or can compute it because only two options that must sum to 1)