University of St Andrews

DECEMBER 2020 72 HOUR ASSESSMENT

SCHOOL OF COMPUTER SCIENCE

MODULE CODE:	**CS5010**
MODULE TITLE:	AI Principles
TIME TO HAND IN:	72 hours
EXAM INSTRUCTIONS:	(a) Answer **ALL THREE** questions. (b) Each question carries 20 marks.

This assessment consists of exam-style questions and you should answer as you would in an exam. As such, citations of sources are not expected, but your answers should be from your own memory and understanding and significant stretches of text should not be taken verbatim from sources. Any illustrations or diagrams you include should be original (hand or computer drawn). You may word-process your answers, or hand-write and scan them. In either case, please return your answers as a single PDF. If you handwrite, please make sure the pages are legible, the right way up and in the right order. Your submission should be your own unaided work. While you are encouraged to work with your peers to understand the content of the course while revising, once you have seen the questions you should avoid any further discussion until you have submitted your results. You must submit your completed assessment on MMS within 72 hours of it being sent to you. Assuming you have revised the module contents beforehand, answering the questions should take no more than three hours.

1. Agents and Learning

Let E be the restaurant domain data given in the table below, in which input attributes influence decisions on whether or not to wait for a table.

EXAMPLE	PRICE	ESTIMATED WAIT	HUNGRY?	BAR?	Wait
D1	Expensive	30-60	No	Yes	No
D2	Expensive	30-60	No	No	No
D3	Medium	30-60	No	Yes	Yes
D4	Cheap	10-30	No	Yes	Yes
D5	Cheap	0-10	Yes	Yes	Yes
D6	Cheap	0-10	Yes	No	No
D7	Medium	0-10	Yes	No	Yes
D8	Expensive	10-30	No	Yes	No
D9	Expensive	0-10	Yes	Yes	Yes
D10	Cheap	10-30	Yes	Yes	Yes
D11	Expensive	10-30	Yes	No	Yes
D12	Medium	10-30	No	No	Yes
D13	Medium	30-60	Yes	Yes	Yes
D14	Cheap	10-30	No	No	No

(a) Explain how a decision tree based on E can be used to predict a wait decision for a new restaurant domain example. (If you sketch an example decision tree as part of your answer, there is no need to ensure that it conforms to the algorithms for decision tree construction given in the lectures.) [4 marks]

(b) Explain the use of learning curves to measure performance of decision trees based on domain examples such as E, giving three possible performance outcomes. [4 marks]

(c) How many bits are needed to classify a new example? Give your answer to three decimal places [2 marks]

(d) Using your answer to part (c), calculate to three decimal places the information gain for the **PRICE**, **ESTIMATED WAIT**, **HUNGRY?** and **BAR?** attributes. [4 marks]

(e) Based on your answer to part (d), which of these four attributes would be the correct choice for the root node of a decision tree based on E? Justify your answer. [2 marks]

(f) Without performing any calculations, explain why the **EXAMPLE** attribute would not be a suitable choice for the root node of a decision tree. [2 marks]

(g) Suppose that a perfectly rational agent A is based on a full decision tree for E, and therefore maximises expected performance. When A is given a new restaurant domain example, A decides to wait when in fact the correct decision was to not wait. Explain this apparent contradiction. [2 marks]

[Total marks 20]

2. Logic

(a) Write the sentence "*Every woman who is the mother of a child loves that child*" as a sentence in first order logic, and convert that sentence into conjunctive normal form. [4 marks]

(b) Consider the knowledge base:

(i) *All migraine sufferers take painkillers.*

(ii) *Anyone who has a virus will not be at work.*

(iii) *Professors do not take painkillers.*

(iv) *Tom either has a virus or suffers from migraine.*

Use resolution to prove that if Tom is a professor, then Tom will not be at work. [10 marks]

(c) From the following axioms, give a forward-chaining proof of the sentence $7 \leq 3 + 9$. Show only the steps that lead to success, and use only the axioms given here and not anything else you know about arithmetic. Make explicit any renaming of variables needed in your proof.

(i) $0 \leq 3$

(ii) $7 \leq 9$

(iii) $\forall x \quad x \leq x$

(iv) $\forall x \quad x \leq x + 0$

(v) $\forall x \quad x + 0 \leq x$

(vi) $\forall x, y \quad x + y \leq y + x$

(vii) $\forall w, x, y, z \quad w \leq y \land x \leq z \implies w + x \leq y + z$

(viii) $\forall x, y, z \quad x \leq y \land y \leq z \implies x \leq z$

[6 marks]

[Total marks 20]

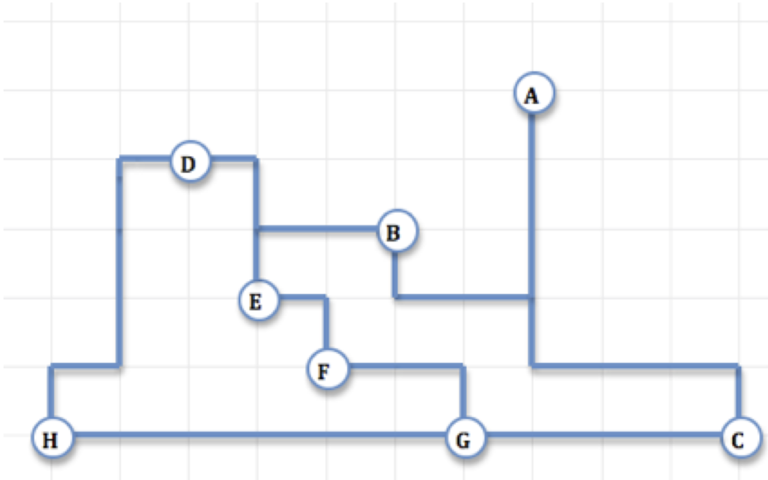

Figure 1: The map required for Q3 (a).

3. Search and Uncertainty

(a) Search

The diagram in Figure 1 shows bus routes between locations in a city. The heavy blue lines represent routes the buses can take, with the stops represented by letters. The lighter grid lines represent city blocks. We can use Manhattan distance in estimates and heuristics, since all bus routes are on grid lines. For example, from location A our initial estimate of distance to H is 12 blocks (5 down + 7 across).

We are to calculate the shortest possible bus route between A and H. Routes are given by the sequence of locations visited, without repeats. From any given location we can move to any other bus stop with no bus stops in-between. E.g. a legal (but obviously bad) route would be ABCGH. On the other hand, AEFGH is not legal, since to get from A to E we have to go through either B or C.

As a heuristic estimate in this question, we will use the sum of the distance travelled from the source to the current node added to the Manhattan distance between the current location and the goal (location H).

(i) Give in detail the search that would be explored by Branch-and-Bound using depth-first search to find the best route from A to H and to prove it is the best route, using the Manhattan distance as the lower bound heuristic for future moves. Assume that when nodes are added during depth-first search, they are added in order of the bounding heuristic. [6 marks]

(ii) Give in detail the search that would be performed by A* to find the best route from A to H, using the total distance heuristic of distance travelled plus Manhattan estimate to goal. [4 marks]

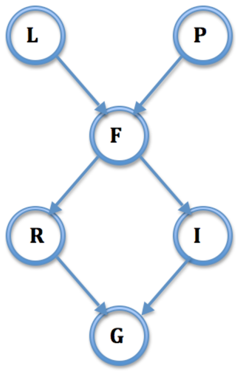

Figure 2: The Bayesian network required for Q3 (b).

(b) Uncertainty.

Consider the following simple description of faults in the brakes of a bus. Two problems may cause a brake repair to be necessary. These are a leak in the brake fluid cable, and worn brake pads. If either of these happens, one of two things may alert the driver to the problem: either noticeably reduced braking power in the bus, or an indicator light of a brake fault on the dashboard. If the driver observes either of these problems, they should take the bus to a garage for repair.

Figure 2 shows the Bayesian network for modelling repairs in a bus. Nodes represent the following: L for a leak in the brake cable; P for a worn brake pad; F for a fault in the brakes; R for reduced braking power; I for an indicator light of a fault on the dashboard; and G for the bus is taken to a garage for a brake repair.

(i) Give an informal description of why the arrows in the diagram are necessary, and why the arrows which are not present may be omitted correctly. [4 marks]

(ii) Suppose we are given the following conditional probability values for the Bayesian Network: $P(L) = 0.1$, $P(P) = 0.2$, $P(F|L,P) = 0.999$, $P(F|notL,P) = 0.99$, $P(F|L,notP) = 0.99$, $P(F|notL,notP) = 0.1$.
From these values and the diagram, calculate the value of P(F) to three decimal places [3 marks]

(iii) Briefly suggest the advantages of using Bayesian reasoning for fault diagnosis, compared to other topics taught in this module, specifically Machine Learning, Logic, and Search. [3 marks]

[Total marks 20]

***** END OF PAPER *****