University of St Andrews

DECEMBER 2018 EXAMINATION DIET

SCHOOL OF COMPUTER SCIENCE

MODULE CODE:	**CS5010**
MODULE TITLE:	Artificial Intelligence Principles
EXAM DURATION:	2 hours
EXAM INSTRUCTIONS	(a) Answer **three** questions. (b) Each question carries 20 marks. (c) Answer questions in the script book.
PERMITTED MATERIALS	Non-programmable calculator

YOU MUST HAND IN THIS EXAM PAPER AT THE END OF THE EXAM.

DO NOT TURN OVER THIS EXAM PAPER UNTIL YOU ARE INSTRUCTED TO DO SO.

1. Search

(a) Explain what is meant by informed and uninformed symbolic search, Iterative Deepening, Uniform Cost, Greedy Best First, and A* symbolic search. [5 marks]

(b) Figure 1 below shows a global map of semi-linear obstacles a robot has to navigate separately from start position S to goal position G. In this case, the robot uses the map and symbolic search to plan a shortest path to G. Table 1 shows some inter-nodal straight-line distances, e.g. S to 1 has a straight-line distance of 10.

(i) Suppose the state space consists of all positions (x, y) in the plane. How many states are there? How many paths are there to the goal? [1 mark]

(ii) Explain briefly why the shortest path from one node to any other in the map must consist of straight-line segments joining some of the nodes. Define a better state space for symbolic search than in (i) for finding the shortest path cost using this feature. Explain why it is an improvement. You should take shortest to mean shortest total distance travelled along the straight-line segments. [2 marks]

(iii) Give, with explanation, a prediction for likely node sequences for the uniform cost, greedy, A* techniques using the shortest path cost in (ii) from start position S and based on the partial entries in Table 1. Nodes S,2,4,G may be taken to be collinear. The A* cost may be taken to differ amongst its frontier nodes at each stage. The planning should be done without checking for repeated states. [6 marks]

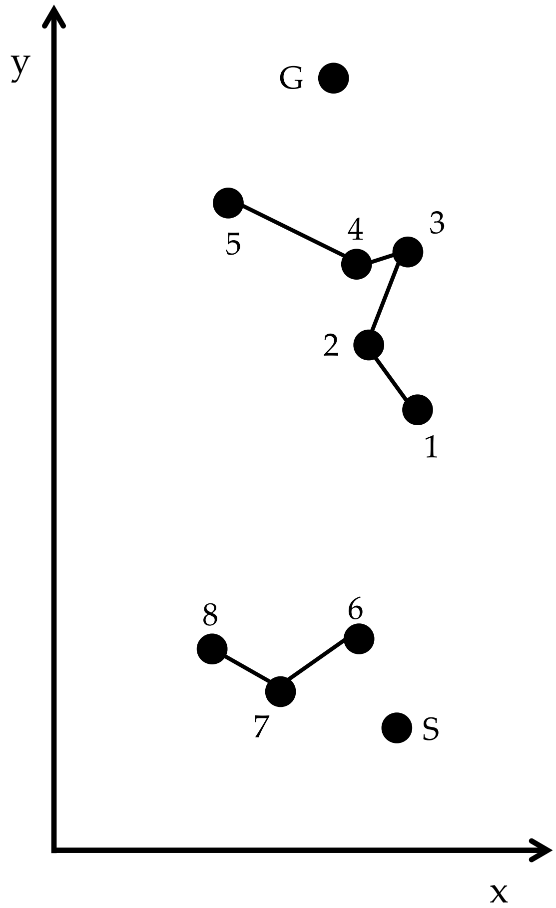
(iv) Suppose instead that checking for repeated states is incorporated, is there a benefit that results for any one of the techniques and if so, what is it? [2 marks]

(c) Suppose that, instead of symbolic search and a global map, a robot uses local sensors to follow a BUG-0 algorithm so that it hits obstacles and then leaves an obstacle edge when progress is possible in terms of straight-line distance-to-goal. Does this provide a guarantee of completeness? If so, why? If not, why not and what would provide such a guarantee? Use Figure 1 to illustrate your answer.

If the BUG-0 algorithm is replaced by a static potential field approach, what start position would cause the approach to fail to reach G and why? [4 marks]

[Total marks 20]

Figure 1.



node	1	2	3	4	5	6	7	8	G
S	**10**	**12**				**1**	**3**		**19**
7						**2**		**1**	
4		**1**	**4**		**1**				
G	**8**	**7**	**5**	**6**	**7**	**19**	**20**		

Table 1.

2. Logic

(a) What are Russell and Norvig's circuit-based and inference-based agent types? Compare them along 2 separate dimensions for relative advantages and disadvantages. [3 marks]

(b) What is Russell and Norvig's TT-entails procedure? [1 mark]

(c) What are two improvements of the DPLL algorithm over TT-entails?

[2 marks]

(d) Suppose an agent in a version of Wumpus World knows one of three Squares (A, B, or C) contains a Wumpus and aims to infer which square this is. The inference is based on these squares having certain characteristics and one extra fact. These are

If a denotes "square A contains a Wumpus", then:

$a \wedge \neg g \wedge h$;

If $\neg b$ denotes "square B does not contain a Wumpus", then:

$\neg b \wedge g \wedge m$;

If $\neg c$ denotes "square C does not contain a Wumpus", then

$\neg c \wedge \neg g \wedge \neg m \wedge (a \vee b)$

Extra fact: "Square B contains a Wumpus or B is not smelly or C has animal droppings", i.e.

$\neg b \vee g \vee h$

where g is "B is not smelly", h is "C has animal droppings", and m is "B has a breeze".

Table 2 below represents the application of the DPLL algorithm to the above knowledge at the stage of conversion into disjunctive clauses for Conjunctive Normal Form (CNF), with (a,b,c) representing a OR b OR c for example. You may assume that rows 3 to 6 represent the CNF for Squares B and C.

(i) What knowledge do the top two clause rows represent? [1 mark]

(ii) Which rows of clauses represent the fact that there is only 1 Wumpus square and why? [2 marks]

(iii) Set g to be false and apply unit propagation to produce a modified column. [2 marks]

(iv) Apply unit propagation in further stages with each such stage producing a further modified column so that the sequence provides a result. [4 marks]

(v) What can you conclude now about whether there is or is not a unique solution? If there is definitely a unique solution, what is it? If there is no such certainty, say what else would need to be done to be sure. [2 marks]

(vi) Why is Wumpus World limited to about 100 x 100 squares in practice for inference-based propositional solvers? How does this limitation generalize to a limitation of propositional logic? [3 marks]

[Total marks 20]

Clause
(¬a, ¬g)
(¬a, h)
(b, g)
(b, m)
(c, ¬g)
(c, ¬m)
(¬b, g, h)
(a, b, c)
(¬a, ¬b)
(¬a, ¬c)
(¬b, ¬c)

Table 2.

3. Learning

(a) Describe, in order, the two main steps taken by back-propagation to compute the weight transition Δw_{ij} for an individual I/O pattern and a weight w_{ij} from a unit i to an output unit u_j in a feedforward neural network.

Additionally, show how there are symbolic expressions for the second step that enable Δw_{ij} to be computed for an individual I/O pattern using just values for input (in_i) to u_j, output (out_j) from u_j, a target ($targ_j$) for u_j, and a learning rate ε. The symbolic terms for these expressions should be drawn from the following set:

$\frac{\delta E}{\delta w_{ij}}, \frac{\delta out_j}{\delta ex_j}, \frac{\delta ex_j}{\delta w_{ij}}, in_i, out_j, targ_j, \Delta w_{ij}$

where E is error, and ex is excitation.

You may use the relations of

$$\frac{\delta E}{\delta out_j} = (out_j - targ_j); \frac{\delta out_j}{\delta ex_j} = out_j \times (1 - out_j); \frac{\delta ex_j}{\delta w_{ij}} = in_i$$

[3 marks]

(b) What is the exploding and vanishing gradient problem for feedforward neural networks with hidden layers? Describe a method for overcoming the problem and give a known drawback for the method.

[3 marks]

(c) Why is the Ravine problem common for neural training using least mean square error? Illustrate your answer with suitable diagrams.

[5 marks]

(d) Training and Validation: Figure 2 below shows a neural Input Space with training points marked as T and validation points marked as V. White points indicate a class with a target output of 0.2 and black points indicate a class with a target output of 0.8.

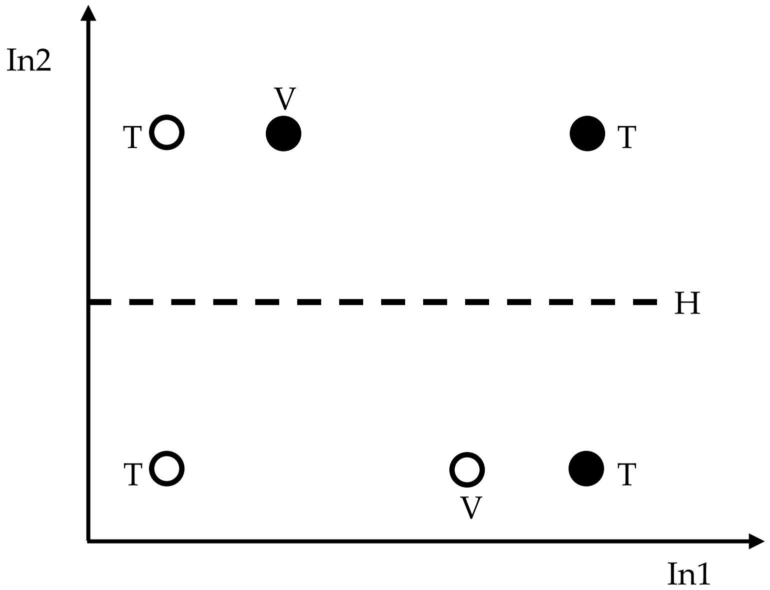
(i) Explain with reasons where the hyperplane H moves towards during training for a 2-1 net. [3 marks]

(ii) Explain how early stopping can be used to produce minimum classification error over the two sets of points as a whole. You should assume back-propagation is used without momentum. [3 marks]

(e) Using a training set containing all the known data points without any validation is likely to have less error for these data points compared to when some points are put in a validation set. So, in general, why not train without validation? Discuss whether you agree or disagree with this rhetorical assertion and motivate your answer. [3 marks]

[Total marks 20]

Figure 2.



4. Uncertainty

(a) Give a broad reason as to why Bayes' Rule is important for AI. Derive Bayes' Rule from the product rule for conditional probability. What benefit does normalisation enable for computation with Bayes' Rule? [2 marks]

(b) Suppose the chance of a rare disease (R) is $P(R) = 1/2000 = 0.0005$, and the chances of certain symptoms (S) occurring given the disease is or is not present are $P(S \mid R) = 0.4$, and $P(S \mid \neg R) = 0.05$ respectively.

Use these values to calculate unnormalized values for $P(R \mid S)$ and $P(\neg R \mid S)$. Normalize these values, explaining your working. [6 marks]

(c) Suppose you are a witness to a terrorist hit-and-run incident involving a hired VansRus van. All VansRus vans are either red or orange. You tell the police that the van was red. Extensive testing shows that, under such stressful conditions, discrimination between red and orange is 80% reliable.

(i) Compute the most likely colour for the van using two random variables: R meaning the van was red, and LR meaning the van looked red, and also using Laplace's Principle of Indifference. Explain the steps in the computation. [3 marks]

(ii) How is your computation of this most likely colour affected numerically if you come to know that 9 out of 10 of the vans are orange? [3 marks]

(d) In your house, there is an alarm that senses when a gauge measuring the temperature of your boiler exceeds a given threshold. Consider the Boolean variables A (alarm sounds), FA (alarm is faulty), and FG (gauge is faulty) and the multi-valued nodes G (gauge reading) and T (actual boiler temperature).

(i) Draw a Bayesian network for this domain, given that the gauge is only likely to fail when the boiler temperature gets too high. Justify your design briefly. [3 marks]

(ii) Suppose there are just two possible actual and measured temperatures, normal and high; the probability that the gauge gives the correct temperature is x when it is working, but y when it is faulty. Tabulate the conditional probabilities associated with G. [3 marks]

[Total marks 20]

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