STUDENT ID ____________________

UNIVERSITY OF OTAGO EXAMINATIONS 2014

COMPUTER SCIENCE

Paper COSC 343

ARTIFICIAL INTELLIGENCE

Semester 1

TIME ALLOWED: 3 HOURS

This examination paper comprises 7 pages including this title page.

Candidates should answer questions as follows:

Candidates must answer all **6** questions.
Each question is worth a total of 10 marks, and submarks are shown thus: (4)

The following material is provided OR candidates are permitted copies of:

Nil.

Use of calculators:

No restriction on the model of calculator to be used, but no device with communication capability shall be accepted as a calculator. Calculators are subject to inspection by the examiners.

Other instructions:

Even when a question asks for a numerical answer the derivation of how that answer was obtained is required.

Please write your student ID at the top of this page.

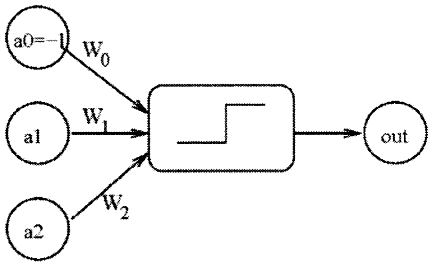
At the end of the exam, hand in this paper attached to the answer book(s).

1. Imagine the Department of Computer Science has a small database recording when the students' computer lab has been busy this semester in various simple circumstances. The database is expressed here as a truth table:

Sun shining	Assignment due	Lab busy
Yes	Yes	Yes
Yes	No	No
No	Yes	Yes
No	No	Yes

You are asked to train a single **binary** perceptron to predict whether the lab will be busy or not based on whether the sun is shining and whether there is an assignment due.

a) With the help of the following perceptron sketch, specify how the inputs are to be assigned and how the output (out) is to be calculated.



i) Why is the value of input a0 set to a fixed value (e.g. −1)?
ii) What will be the values of input a1?
iii) What will be the values of input a2?
iv) How will you calculate the value of perceptron output (out)? (4)

b) Is this task (as specified by the above truth table) linearly separable? Use a geometric illustration to show that it is (or that it is not). (1)

c) Why should we care whether this task is linearly separable? (1)

d) The parameters of the model, W_0, W_1 and W_2 are initially small random values. Write a pseudocode for iteratively adjusting these parameters so that in the end the actual output of the perceptron (out) is the same as the target output specified by the truth table. (3)

e) This prediction task can be easily handled by an AI tool called a decision tree. Draw the minimal decision tree that is consistent with the examples in the above truth table. (1)

TURN OVER

2. A survey is carried out to determine the causes of happiness. Respondents are asked three questions: (1) Are you rich? (2) Are you a student? (3) Are you happy? Based on the results of the survey, the following joint probability distribution is estimated:

	student		*¬ student*	
	rich	*¬ rich*	*rich*	*¬ rich*
happy	0.10	0.15	0.10	0.05
¬ happy	0.06	0.09	0.30	0.15

a) Explain how this distribution could have been calculated from the results of the survey among *N* respondents. (1)

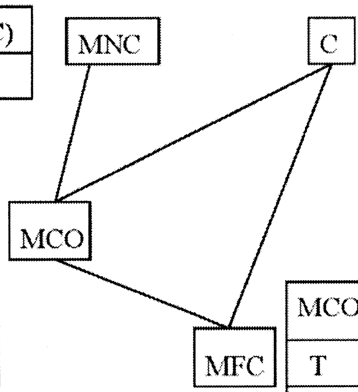
b) Are you more likely to be happy given you are a student or given you are rich, according to this distribution? Calculate appropriate posterior probabilities to justify your answer. (2)

c) Show that variables *Happy* and *Rich* are conditionally independent given *Student* in the above distribution. (3)

d) Draw a Bayesian network expressing the information in the above distribution. For each node, write down a corresponding conditional probability table that quantifies the effect of parent(s) on that node. (4)

TURN OVER

3. Consider the following Bayesian network, featuring the Boolean variables C (cold weather), MNC (Mary has a new coat), MCO (Mary has a coat on), and MFC (Mary feels cold).



P(MNC)
0.1

P(C)
0.4

MNC	C	p(MCO|MNC,C)
T	T	0.9
T	F	0.4
F	T	0.6
F	F	0.1

MCO	C	p(MFC|MCO,C)
T	T	0.4
T	F	0.0
F	T	0.9
F	F	0.2

a) Fill in the missing columns in the conditional probability tables, i.e. columns (from left to right) for P(¬ *MNC*), P(¬C), P(¬*MCO* | *MNC* & C) and P(¬*MFC* | *MCO* & *C*). (2)

b) Calculate the probability that Mary has a new coat, and it is not cold, and Mary has a coat on, and Mary is not feeling cold. In other words calculate the joint probability *P*(*MNC* & ¬C & *MCO* & ¬*MFC*). Show your working. (2)

c) How would you estimate the latter probability, i.e. *P*(*MNC* & ¬C & *MCO* & ¬*MFC*), using the method of approximate inference by rejection sampling based on evidence? Write down pseudocode for this approximate inference. (4)

d) There are some salient things to be read from the conditional probability tables in the Bayesian network above. Thus, what kind of person is Mary judging from this network with given probabilities? (2)

TURN OVER

4. This question concerns clustering and classification.

   a) Define the k-means clustering. (1)

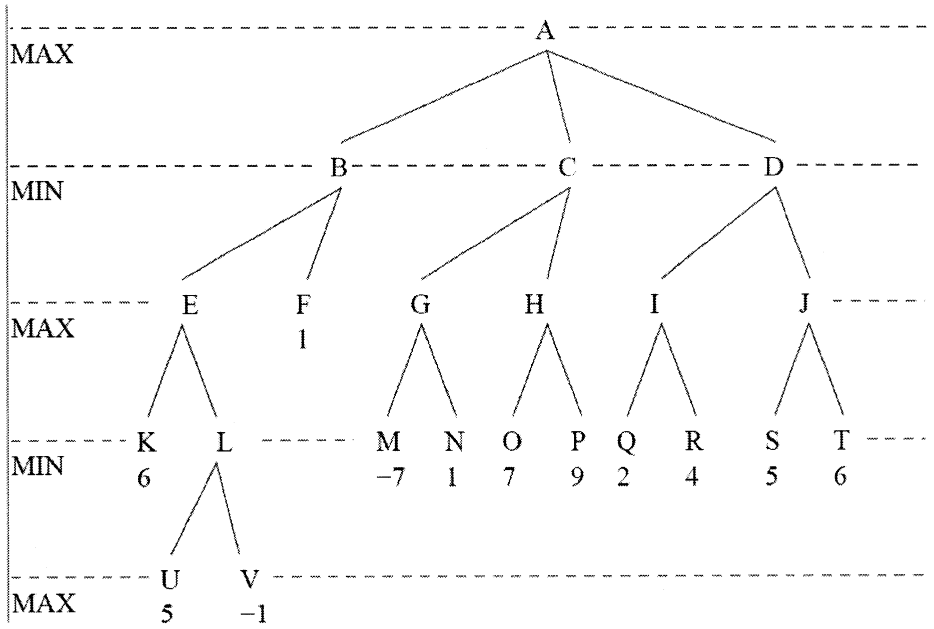
   b) Describe Manhattan, Euclidean and Hamming distance between two vectors a = (a1, a2, …, aN) and b = (b1, b2, …, bN). (3)

   c) Write pseudocode for the k-means clustering algorithm that will include initialization, iterative procedure and stopping criterion. Be as concrete as possible. (3)

   d) Is there a guarantee that the k-means produced by this algorithm will always converge to the global optimum? Why yes – or why not? (1)

   e) Suppose you have finished the clustering and a new data point arrives. Which of the existing clusters would you assign this new data point to? Describe the criteria or method you would use. (2)

5. Consider the following search tree for a two-player game. The letters are labels for nodes. The numbers on the leaf nodes are the results returned by a utility function on these nodes. High values are good for MAX while the low values are good for MIN. MAX starts to play:



a) Define the MINIMAX value of a node n in the search tree. (1)

b) Describe the MINIMAX algorithm. (2)

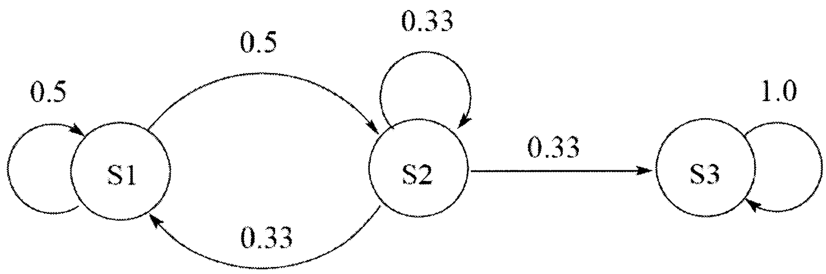
c) What is alpha-beta pruning? (1)

d) If the tree is searched using the MINIMAX algorithm improved by alpha-beta pruning, which nodes can be pruned (i.e. not expanded at all)? (3)

e) Using the MINIMAX algorithm, which node should MAX pick if MIN has picked the node C? (1)

f) Explain why in the worst case, applying alpha-beta pruning requires an exhaustive search of the state space. (2)

6. The figure illustrates a simple hidden Markov model with three states, S1, S2, and S3. At each time step, either the output A or B is observed with certain probability.



out	P
A	0.6
B	0.4

out	P
A	0.1
B	0.9

out	P
A	0.3
B	0.7

a) State a Markov assumption. Is the Markov assumption fulfilled by the model above? (2)

b) What is the order of the process illustrated in the figure? Why? (1)

c) Explain what is meant by a stationary temporal process. Is the process in the figure above stationary? Why yes or why not? (2)

d) What is the most probable sequence of states that would generate the sequence of observations ABBA, given we start in state 1? Explain why. (3)

e) Is it possible that the sequence of states S1, S2, S2, and S3 would have generated the output sequence of identical tokens AAAA? What is the probability of generating this output sequence? (2)

END