香 港 中 文 大 學
The Chinese University of Hong Kong
二0一一至一二年度 上學期科目考試

Course Examination 1st Term, 2011 – 2012

科目編號及名稱 Course Code & Title	:	CSCI3230 Fundamentals of Artificial Intelligence
時間 Time allowed	:	2 小時 hours 00 分鐘 minutes
學號 Student I.D. No.	:	座號 Seat No. :

Answer Four out of Five questions
Total: 100 marks

1.

a) What is the crucial difference between the following two first-order sentences where R is a binary predicate? [3 marks]

i. $\forall x \exists y\, R(x, y)$.

ii. $\exists x \forall y\, R(x, y)$.

b) Express each of the following statements (i to vi) as First Order Logic sentences using the symbols below (together with $\forall, \exists, \neg, \wedge, \vee, \rightarrow$): [6 marks]

Symbol	Meaning
Student(X)	X is a student
Tough(X)	X is tough
Boring(X)	X is boring
HasProject(X)	X has a project
LightWorkLoad(X)	X has light workload
Likes(X,Y)	X likes Y
Course(X)	X is a course

i. Students like only courses with light workload.
ii. *CSCI3230* is a course with a project.
iii. Any course with a project is either tough or boring or both.
iv. Tough courses do not have light workload.
v. *Peter* does not like boring courses.
vi. If *Peter* likes *CSCI3230*, then *Peter* is not a student.

c) Convert your answers for b) into **Conjunctive Normal Form (CNF)**. [6 marks]

d) Prove statement vi from statements i to v using *any one* of the **Forward Chaining**, **Backward Chaining** and **Resolution** approaches. [5 marks]

e) What are the differences between propositional logic, first-order logic and higher-order logic? [5 marks]

2.

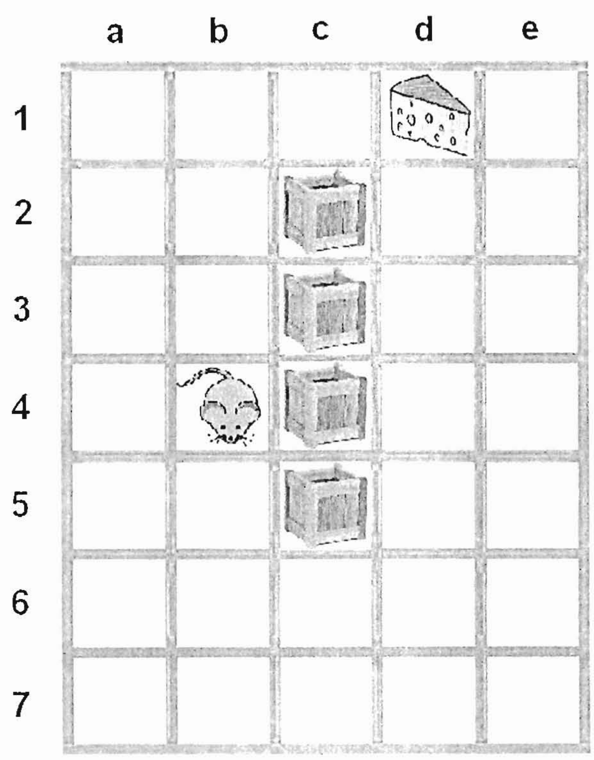
a) Name a multi-point stochastic search algorithm and describe it briefly. [3 marks]

b) The graph below shows a map of a house. A little rat is located at (b,4) and is searching for a piece of cheese which is located at (d,1). The rat cannot get through the four boxes located from (c,2) to (c,5). Since the rat is so hungry, it wants you to use the A* algorithm to help it find the shortest path from its location to the cheese. Treat each square as a node and the rat can only move one square at a time vertically or horizontally.

i. Design a heuristic function for your A* algorithm and prove that it is admissible. [4 marks]

ii. Perform the A* search to find the cheese (using your own heuristic function and draw the search tree(s)) [8 marks]

iii. If you use the best-first search without heuristics to find the cheese, how many nodes you need to expand to find the cheese? Show how you obtain your answer. [4 marks]



c) What is an imperfect game? Explain why we have to make imperfect decisions in game playing? [3 marks]

d) In the back-propagation updating rule $w'_{i,j,k} \leftarrow w_{i,j,k} - \alpha \dfrac{\partial E}{\partial w_{i,j,k}}$, there is a parameter $\alpha > 0$ named Learning Rate. What are the general effects of varying the value of α? [3 marks]

3.

a) Consider the following puzzle (Figure A): use each of the digits 1 to 9 once and only once to fill in the circles such that for each line, the sum of the digits in the two neighboring circles is equal to the number on the line, e.g. the digits in the first two circles should add up to a12. Formulate the puzzle as a search problem (state space, state transitions and goal state(s)).

[8 marks]

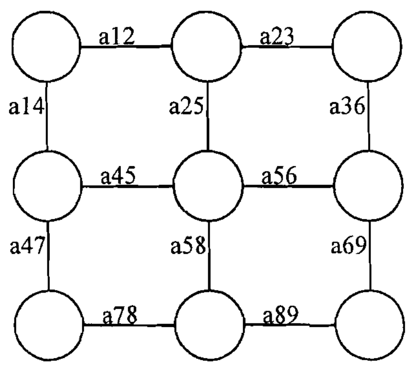


Figure A

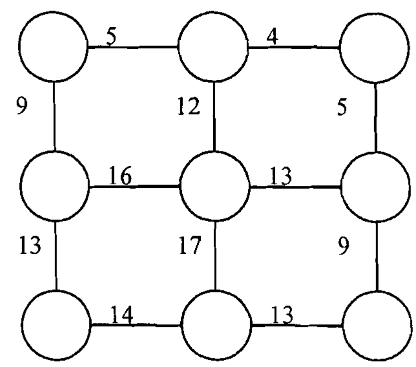


Figure B

b) Use *any heuristic* you like to solve the instance in Figure B above. Copy Figure B to your answer book and show the solution. [3 marks]

c) The following is the pseudocode of Minimax algorithm using recursive negation. Modify the pseudocode into **alpha-beta pruning**. [6 marks]

```
function minimax(node, depth)
    if node is a terminal node or depth == 0
        return the heuristic value of node
    else
        best = -∞
    for child in node           // evaluation is identical for both players
        best = max(best, -minimax(child, depth-1))
    return best
```

d) A programmer claims that his Chess playing program can never lose because it is using MiniMax with Alpha-Beta pruning, and MiniMax can search the game tree optimally. Do you agree? Explain briefly. [4 marks]

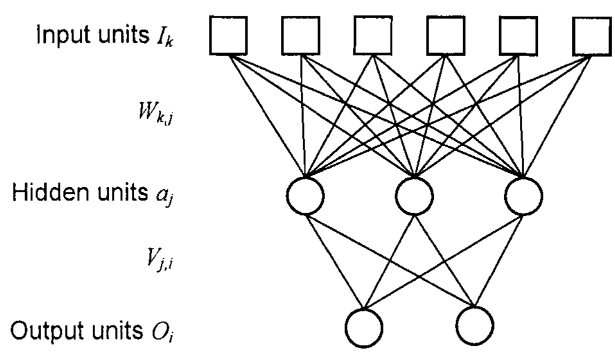
e) In real problems, the search spaces are usually very large. Suggest **TWO** methods to handle the huge spaces. [4 marks]

4.

a) What are the differences between Single-Layer Perceptron (SLP) and Multi-Layer Feed-Forward Network (MLN)? Give a concrete example of classification that can be classified by MLN but not SLP. [3 marks]

b) For the sigmoid function $f(z)=\frac{1}{1+e^{-bz}}$, find its derivative $f'(z)$. [3 marks]

c) Derive the **TWO** updating rules ($\frac{\partial E}{\partial V_{j,i}}$ and $\frac{\partial E}{\partial W_{k,j}}$) of weights and biases in back-propagation algorithm for Multi-Layer Feed-Forward Network (MLN). The activation function in the output units is the above sigmoid function, i.e., $O_i = f(u_i)$, where $u_i = \sum_j a_j \times V_{j,i} + B_{O_i}$. The activation function in the hidden units is the above sigmoid function, i.e., $a_j = f(s_j)$, where $s_j = \sum_k I_k \times W_{k,j} + B_{a_j}$. The error function is the sum of squared error, i.e., $E = \frac{1}{2}\sum_i (O_i - T_i)^2$ where T_i is the target value. The notations and the diagram of the MLN are shown below: [10 marks]



d) What is the shortcoming of the above back-propagation algorithm? Suggest a method to handle this shortcoming. [3 marks]

e) When training a neural network, a beginner observes that when he uses more hidden layers and increases the number of hidden neurons, the training accuracy will increase, but the training time also increases accordingly. What is your advice to him? Explain briefly. [4 marks]

f) Why $E = \frac{1}{2}\sum_i (O_i - T_i)^2$ is preferred to $E = \frac{1}{2}\sum_i | O_i - T_i |$ as the error function in back-propagation algorithm of MLN? Explain briefly. [2 marks]

5.

Teaching Assistant	Attributes						Evaluation Score
	Gender[1]	Exp[2]	Eng[3]	Man[4]	CSize[5]	Abs[6]	
X_1	Male	0	Fair	Fair	Medium	None	Pass
X_2	Male	>=2	Good	Good	Medium	None	Pass
X_3	Female	1	Good	Good	Medium	None	Pass
X_4	Female	0	Fair	Good	Medium	None	Pass
X_5	Male	>=2	Poor	Poor	Medium	None	Pass
X_6	Male	0	Poor	Fair	Medium	None	Not Pass
X_7	Female	0	Poor	Good	Medium	None	Not Pass
X_8	Female	1	Fair	Fair	Medium	None	Pass
X_9	Male	1	Poor	Poor	Medium	None	Not Pass
X_{10}	Male	1	Fair	Fair	Medium	None	Pass
X_{11}	Male	0	Fair	Poor	Medium	None	Not Pass

Table 1: Training Dataset

1. Gender: {Male, Female}
2. Teaching Experience (in terms of year): {0, 1, >=2}
3. Fluency in English: {Poor, Fair, Good}
4. Fluency in Mandarin: {Poor, Fair, Good}
5. Class size: {Small, Medium, Large}
6. Absence rate: {None, Some, All}

Teaching Assistant	Attributes						Evaluation Score
	Gender[1]	Exp[2]	Eng[3]	Man[4]	CSize[5]	Abs[6]	
X_{11}	Male	0	Good	Fair	Medium	None	Pass
X_{12}	Male	>=2	Fair	Poor	Small	None	Not Pass
X_{13}	Female	1	Good	Fair	Large	Some	Not Pass
X_{14}	Female	0	Good	Poor	Medium	None	Pass
X_{15}	Male	1	Poor	Good	Large	All	Pass
X_{16}	Male	0	Good	Fair	Medium	None	Pass

Table 2: Validation Dataset

a) We need to classify whether a teaching assistant will pass the course evaluation based on the attributes in Table 1. There are altogether 11 training samples (Table 1) and 5 validation samples (Table 2). Please use Information Gain (IG) to learn **a decision tree** based on the attributes in the training dataset and draw a diagram of the decision tree. **Show the detailed calculations on choosing the root attribute only.** You may find the following logarithm table useful. [10 marks]

x	1/2	1/3	2/3	1/4	3/4	3/7	4/7
$\log_2(x)$	-1	-1.5850	-0.5850	-2	-0.4150	-1.2224	-0.8074
x	1/5	2/5	3/5	4/5	4/11	7/11	
$\log_2(x)$	-2.3219	-1.3219	-0.7370	-0.3219	-1.4594	-0.6521	

b) Evaluate the decision tree using the validation dataset. Construct a **Confusion Matrix** to show the results, using "Pass" as positive and "Not Pass" as negative. [2 marks]

c) Based on your **Confusion Matrix**, calculate the performance metrics: $Accuracy = \dfrac{TP+TN}{TP+FP+FN+TN}$, $Precision = \dfrac{TP}{TP+FP}$, $Recall = \dfrac{TP}{TP+FN}$, $F\_measure = \dfrac{1}{\left(\frac{1}{Recall}+\frac{1}{Precision}\right)/2}$. Interpret each of them. [4 marks]

d) Briefly explain the **Ockham's Razor** principle. [4 marks]

e) What is cross-validation? What is overfitting? How can cross-validation help prevent overfitting? [5 marks]

全卷完

END