UNIVERSITY OF LONDON
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2004

BEng Honours Degree in Computing Part II
MEng Honours Degrees in Computing Part II
MSc in Computing Science
BSc Honours Degree in Mathematics and Computer Science Part II
MSci Honours Degree in Mathematics and Computer Science Part II
for Internal Students of the Imperial College of Science, Technology and Medicine

This paper is also taken for the relevant examinations for the Associateship of the City and Guilds of London Institute
This paper is also taken for the relevant examinations for the Associateship of the Royal College of Science

PAPER C231=MC231

ARTIFICIAL INTELLIGENCE I

Friday 14 May 2004, 14:30
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

1a i) Write the following linear algebra problem as a constraint satisfaction problem (CSP):

$$2 \leq x \leq 5, \quad 1 \leq y \leq 8, \quad 1 \leq z \leq 5, \quad x + y \leq 5, \quad y + z \geq 7, \quad x + z \geq 5$$

ii) Make the CSP arc-consistent.

iii) Describe the method of forward-checking in CSP solving.

b i) What is the main difference between genetic algorithms and genetic programming?

ii) Represent the following functions as graphs:
$f(a,b) = 2*(a + (b \div \sqrt{a}))$ $\qquad g(a,b) = 2*(a*b)$

iii) Suppose that functions f and g were to be taken as parents to produce offspring functions via crossover and that the crossover fragment for f was $(b \div \sqrt{a})$, and the crossover fragment for g was $(a*b)$. Draw the graphs for the resulting offspring and write out the functions they represent.

iv) Explain the terms *fitness function* and *function set* in genetic programming.

c Suppose that the following data has been collected:

Example	Property 1	Property 2	Property 3	Category
1	good	2	yes	A
2	good	3	no	A
3	bad	3	yes	B
4	bad	5	yes	B
5	good	5	no	C

i) Use the formula below to calculate the entropy of the set of examples. Also calculate the information gain for property 1.

$$Entropy(S) = \sum_{i=1}^{n} -p_i \, log_2(p_i)$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

[p_i = proportion of examples in category i, S_v = set of examples taking value v]

ii) Describe how information gain is used in the ID3 algorithm.

iii) For what kind of learning problems is decision tree learning appropriate?

The three parts carry, respectively, 35%, 30% and 35%.

2a In the context of machine learning, explain the terms *overfitting* and *predictive accuracy.*

b Cluedo is a board game where players make guesses about who committed a murder, in which room the murder occurred and what instrument was used to commit the murder. Suppose that both player 1 and player 2 know that: (a) "if the murderer was in the bathroom, they had the dagger". Suppose further that player 1 knows that (b) "if Professor Plum was in the bathroom, they had the dagger", and that player 2 knows that (c) "only Professor Plum was in the bathroom".

i) Using only the predicates: *is_pp(X)* to represent "X is Professor Plum", *in_bathroom(X)* to represent "X was in the bathroom", *is_murderer(X)* to represent "X is the murderer" and *had_dagger(X)* to represent "X had the dagger", write sentences (a) (b) and (c) in first order logic using implication signs and no constants.

The identification and absorption rules of inference are as follows:

Identification:

$$\frac{p \leftarrow A, B \qquad p \leftarrow A, q}{q \leftarrow B \qquad p \leftarrow A, q}$$

Absorption:

$$\frac{q \leftarrow A \qquad p \leftarrow A, B}{q \leftarrow A \qquad p \leftarrow q, B}$$

ii) If player 1 used the identification rule to induce a new statement using sentences (a) and (b) and substituting *is_pp(X)* for q, what would they induce? Show how you worked this out and explain in English how the induced statement explains sentence (b).

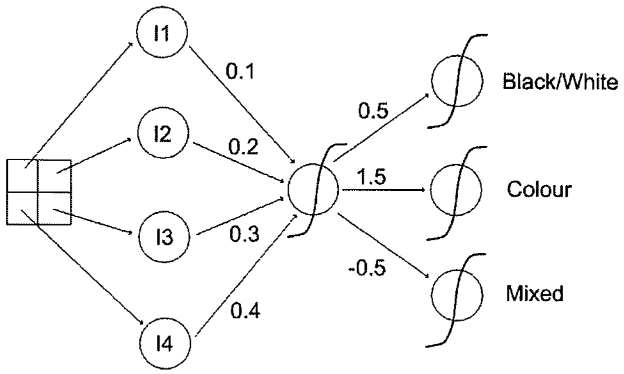
iii) If player 2 used the absorption rule with the same substitutions for A, B, p and q as in part (ii), what would they induce? [Show how you worked this out].

iv) Write sentences (a) and (c), and the induced sentence from part (v) in conjunctive normal form and use resolution to explain sentence (a).

v) How are rules such as identification and absorption used in machine learning?

The two parts carry, respectively, 35% and 65% of the marks.

3 A multi-layer Artificial Neural Network (ANN) is being trained to classify 2x2 pixel graphics. If a graphic has 3 or more pixels which are red(R), green(G) or blue(B), it is classified as "colour"; if it has 3 or more pixels which are black(K) or white(W), it is classified as "black and white", and it is classified as "mixed" otherwise. The network has been randomly initialised as follows:



i) This ANN is using sigmoid units, which, given weighted sum S as input, calculates the following for the output: $1/(1+e^{-S})$. Describe two other possibilities for the functions in the units.

The output from an input node is 1 if it corresponds to a pixel which is red, 2 if it corresponds to a green pixel, 3 for blue, 4 for black and 5 for white. The following are two graphics to be classified by the ANN:

Graphic 1:

R	R
G	W

Graphic 2:

K	W
B	B

ii) What is the input to and output from the hidden node for graphic 1?

iii) Calculate the output from output nodes 1, 2 and 3 for both graphics.

iv) Which of these graphics is correctly classified by the ANN? Why? For the graphic which is incorrectly classified, use the following formula to calculate the error term for the output units:

$$\delta_{O_k} = o_k(E)(1 - o_k(E))(t_k(E) - o_k(E))$$

[$o_k(E)$ = observed output from node k for example E, $t_k(E)$ = the target output].

v) Give an overview of the back-propagation ANN learning algorithm.

The five parts carry, respectively, 10%, 15%, 30%, 20%, and 25% of the marks.

4a i) How can we test whether one heuristic search is better than another?

ii) Briefly describe the minimax principle.

b i) Using the associativity of $\wedge$ and $\vee$ and the re-write rules below, show that the sentence below in propositional logic is false. Clearly mark which rewrite rules you use, wherever you use them.

Sentence: $(P \vee (P \rightarrow R)) \wedge (P \wedge \neg(\neg P \rightarrow \neg Q))$

Rules:
(i) $(X \vee \text{True}) \Leftrightarrow \text{True} \Leftrightarrow (X \vee \neg X)$
(ii) $(X \wedge \text{False}) \Leftrightarrow \text{False} \Leftrightarrow (X \wedge \neg X)$
(iii) $\neg(\neg X) \Leftrightarrow X$ (iv) $(\neg X \vee Y) \Leftrightarrow (X \rightarrow Y)$
(v) $\neg(X \wedge Y) \Leftrightarrow (\neg X \vee \neg Y)$ (vi) $\neg(X \vee Y) \Leftrightarrow (\neg X \wedge \neg Y)$

ii) What does it mean to say that a propositional sentence is false? Name another method for showing that a propositional sentence is false.

iii) What are rules (v) and (vi) known as?

c i) Translate the sentences below into universally quantified implications in first order logic, using only the predicates supplied and no constants.

Predicates:
host_olympics(X), european(X), big(X), on_river(X), is_london(X), is_paris(X)

Sentences:
(a) "The olympics will be hosted in a European city"
(b) "The olympics are held in big cities which are on rivers"
(c) "London and Paris are the only big cities in Europe"
(d) "All big cities are on rivers"
(e) "Paris is not going to host the olympics"

ii) Write sentences (a) to (e) in conjunctive normal form. You may need to use some of the rewrite rules given in part b above.

iii) Given that somewhere is going to host the olympics, use the resolution method to prove that London will host the olympics.

The three parts carry, respectively, 20%, 30% and 50% of the marks.