IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2009

BEng Honours Degree in Computing Part III
MEng Honours Degree in Information Systems Engineering Part IV
MEng Honours Degrees in Computing Part IV
MSc in Advanced Computing
MSc in Computing Science (Specialist)
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

PAPER C395=I4.58

MACHINE LEARNING

Thursday 30 April 2009, 14:30
Duration: 120 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

Section A *(Use a separate answer book for this Section)*

1 Consider the following set of positive (+) and negative (-) training examples:

	sky	air	humid	wind	water	forecast	Enjoy Sport
1	sunny	warm	normal	strong	warm	same	+
2	sunny	warm	high	strong	warm	same	+
3	rainy	cold	high	strong	warm	change	-
4	sunny	warm	high	strong	cool	change	+
5	sunny	warm	normal	weak	warm	same	-

1a Apply the CANDIDATE-ELIMINATION learning algorithm. Write out the intermediate and the final results.

1b What are the differences (if any) between FIND-S and CANDIDATE-ELIMINATION algorithms?

1c Show the decision tree that would be learned by ID3 assuming that it is given the above-listed training examples. Write out the intermediate calculations.

1d What is the difference between the CANDIDATE-ELIMINATION and ID3 algorithms? What is the relationship between the learned decision tree in (1c) and the version space that is learned from the same examples in (1a)?

The five parts carry, respectively, 30%, 10%, 40%, 20% of the marks.

2

2a Give three examples of activation functions with the corresponding mathematical formulas. Which activation function is most commonly used in Multilayer Feed Forward Neural Networks and why?

2b Explain the principle of the gradient descent algorithm. Accompany your explanation with an illustrative diagram. Explain the use of all the terms and constants that you introduce and comment on the range of values that they can take.

2c Derive the gradient descent training rule for a simple network with a single unit with output o given by the formula:

$$o = w_0 + w_1 x_1 + w_1 x_1^2 + \ldots + w_n x_n + w_n x_n^2.$$

Define explicitly the cost/error function E, while assuming that a set of training examples D is provided, where each training example $d \in D$ is associated with the target output t_d.

The three parts carry, respectively, 15%, 35%, 50% of the marks.

3 Suppose that we want to solve the problem of finding out when we shall enjoy doing sports. Suppose further that the solution to the problem can be learned based on the following training examples:

	sky	air	wind	humid	wind	water	forecast	Enjoy Sport
1	sunny	warm	strong	normal	strong	warm	same	+
2	sunny	warm	strong	high	strong	warm	same	+
3	rainy	cold	strong	high	strong	warm	change	-
4	sunny	warm	strong	high	strong	cool	change	+
5	sunny	warm	weak	normal	weak	warm	same	-

3a What is a suitable chromosome design for the given problem? Provide a short explanation of the solution.

3b What does "fitness function" refer to in Genetic Algorithms applications and how it can be defined? What is the suitable fitness function for the given problem? Provide a short explanation of the solution?

3c Which other Genetic Algorithm parameters need to be defined? What is the suitable definition of those parameters for the given problem? Provide a short explanation for each parameter.

3d What is the result of applying a single round of the prototypical Genetic Algorithm? Explain your answer in a clear and compact manner by providing the pseudo code of the algorithm.

The four parts carry, respectively, 20%, 20%, 30%, 30% of the marks.

END Section A *(Use a separate answer book for question 4)*

Section B (Use a separate answer book for this Section)

4a Let

a_1 be $p(b, 4)$,

a_2 be $p(b, X)$,

a_3 be $p(Y, 4)$ and

a_4 be $p(Z, Z)$.

i) State the definition of $\succeq$ (subsumption) with respect to atoms.

ii) For the atoms $a_1 - a_4$ above, state all true relations of the form $a_i \succeq a_j$.

iii) For each pair of atoms $\langle a_i, a_j \rangle$ for which $i < j$ state the least general generalisation of a_i and a_j.

b Let

C be $member(U, cons(V, W))) \leftarrow member(U, W)$,

D be $member(U1, cons(V1, cons(W1, Z1)))) \leftarrow member(U1, Z1)$

Is it the case that $C \succeq D$? Explain your answer based on the definition of $\succeq$.

c Consider the following two statements.

C: Henri lives in Paris
D: Henri lives in France

In each case below explain your answer.

i) Represent statement **C** as a definite clause.

ii) Represent statement **D** as a definite clause.

iii) Define a background knowledge clause **B** which allows the clauses for **C** and **D** to be related according to their generality.

iv) What is the generality relation between the clauses for **C** and **D** given **B**?

d Let

B_1 be $animal(X) \leftarrow cat(X)$

B_2 be $cat(kitty) \leftarrow$

$B = B_1 \wedge B_2$ be the background knowledge

$E = nice(kitty) \leftarrow$ be an example.

i) Letting H stand for the hypothesis, state the condition in Inductive Logic Programming which H, B and E must satisfy.

ii) For B and E above, what is $\perp$ (the most specific hypothesis)? Explain.

The four parts carry, respectively, 40%, 20%, 20%, and 20% of the marks.