IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2011

BEng Honours Degree in Computing Part III
MSc in Computing for Industry
MEng Honours Degree in Information Systems Engineering Part IV
MSci Honours Degree in Mathematics and Computer Science Part IV
BSc Honours Degree in Mathematics and Computer Science Part III
MSci Honours Degree in Mathematics and Computer Science Part III
MSc in Advanced Computing
MSc in Computing Science
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
This paper is also taken for the relevant examinations for the Associateship of the Royal College of Science

PAPER C395

MACHINE LEARNING

Wednesday 4 May 2011, 14:30
Duration: 90 minutes

Answer THREE questions

Paper contains 4 questions
Calculators required

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

1a Given the target function output representation

$$o_d = w_0 + w_1x_1 + w_1x_1^2 + \ldots + w_nx_n + w_nx_n^2$$

what is the Least Mean Squares (LMS) training rule used for and how is it defined?

1b Explain the principle of the gradient descent algorithm. Accompany your explanation with a diagram. Explain the use of all the terms and constants that you introduce and comment on the range of values that they can take.

1c Derive the gradient descent training rule assuming that the target function representation is:

$$o_d = w_0 + w_1x_1 + w_1x_1^2 + \ldots + w_nx_n + w_nx_n^2.$$

Define explicitly the cost/error function E, 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.

1d Prove that the LMS training rule performs a gradient descent to minimize the cost/error function E defined in 1c.

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

2a Given a set of n training examples $D = \{(x_1, y_1), ..., (x_n, y_n)\}$, let us assume the model $y_i = f(x_i) + e_i$, according to which are those examples generated, where e_i follows a distribution $N(0,\sigma)$ and is independently drawn from it. By further assuming that any two examples (x_i, y_i) and (x_j, y_j), $i \neq j$, are mutually independent, prove that the Maximum Likelihood (ML) estimation for the target function approximation f^* is given by:

$$f^* = \text{argmin}_f \sum_{i=1}^{n} \left(f(x_i) - y_i\right)^2 .$$

2b Lets assume now that the target function f is defined as $f(x) = ax + b$. Prove that the ML estimations for a and b are given by:

$$b = \bar{y} - a\,\bar{x}, \quad a = \frac{S_{xy}}{S_{xx}}$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $S_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ and

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y}) .$$

The two parts carry each 50% of the marks.

END Section A *(Use a separate answer book for question 3)*

3 a Let

B_1 be $happy(X) \leftarrow at(imperial, X)$

B_2 be $at(imperial, jane) \leftarrow$

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

$E = student(jane) \leftarrow$ 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.

b Let

C be $append(c(U, V), W, c(U, X)) \leftarrow append(V, W, X)$,

D be
$append(c(U1, c(V1, W1)), X1, c(U1, c(V1, Y1))) \leftarrow append(W1, X1, Y1)$

i) Is it the case that $C \models D$? Explain your answer.

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

iii) What is the lgg of C and D? Explain your answer in terms of compatible literals.

c 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.

d Consider the following two statements.

C: Jean lives in Paris
D: Jean 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**?

The four parts carry, respectively, 15%, 35%, 30%, and 20% of the marks.