IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2010

BEng Honours Degree in Computing Part III
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

Tuesday 4 May 2010, 10:00
Duration: 120 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

$$o_d = w_0 + w_1x_1 + \ldots + w_nx_n,$$

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. Illustrate 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:

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

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, 25%, 20%, 20%, 35% of the marks.

2a Sketch a diagram that describes the architecture of a standard Multilayer Neural Network with one hidden layer that aims at solving a binary (i.e. a two class) classification problem. Which key limitation of single-layer perceptrons is overcome by Multilayer Neural Networks?

2b Assume that the Neural Network in part (a) uses the sigmoid activation function $\varphi(z) = 1/(1+e^{-z})$. Write down the equation that gives the output of the nodes in the hidden layer and the equation that gives the output of the nodes in the output layer. Show that the derivate $\varphi'(z)$ of the activation function $\varphi(z)$ is given by $\varphi'(z) = \varphi(z)(1-\varphi(z))$.

2c Consider the alternative cost function for training a Multilayer Neural Network with $k = [1,K]$ outputs given below:

$$E(w) = (1/2)\sum_d\sum_k(t_{dk} - y_{dk})^2 + \lambda\sum_{j,k}w_{jk}^2.$$

In the equation above t_{dk} is the target (desired) output of a node k in the output layer and y_{dk} is the obtained output of the network when presented with the example $d = [1,D]$. The summation $\lambda\sum_{j,k}w_{jk}^2$ in the second term is over the pairs of nodes j, k that are connected. D is the number of training examples.

Describe how this cost function differs from the classical cost function used for training Multilayer Neural Networks. Explain why this cost function is more appropriate for training a Neural Network with a complex structure when only few training examples are available.

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

3 Suppose that we want to predict whether a household will respond or not to a survey a company will send to it. Suppose further that correct prediction can be learned based on the following training examples:

District	House Type	Income	Previous Customer	Outcome
Suburban	Detached	High	No	No Respond
Suburban	Detached	High	Yes	No Respond
Rural	Detached	High	No	Responded
Urban	Semi-detached	High	No	Responded
Urban	Semi-detached	Low	No	Responded
Urban	Semi-detached	Low	Yes	No Respond
Rural	Semi-detached	Low	Yes	Responded
Suburban	Terrace	High	No	No Respond
Suburban	Semi-detached	Low	No	Responded
Urban	Terrace	Low	No	Responded
Suburban	Terrace	Low	Yes	Responded
Rural	Terrace	High	Yes	Responded
Rural	Detached	Low	No	Responded
Urban	Terrace	High	Yes	No Respond

3a What is the result of applying the ID3 algorithm to this problem? Provide all calculations needed for the solution.

3b Design a genetic algorithm to learn conjunctive classification rules for the above-given problem. Describe precisely the bit string encoding of hypotheses and the full set of Genetic Algorithm parameters.

The two parts carry, respectively, 50% and 50% 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

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

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

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.

b Let

a_1 be $p(3, 4)$,

a_2 be $p(3, X)$,

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.

c Consider the following two statements.

C: John lives in Paris
D: John 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 three parts carry, respectively, 40%, 40%, and 20% of the marks.