IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2014-2015

BEng Honours Degree in Computing Part III
MEng Honours Degree in Electronic and Information Engineering Part IV
BEng Honours Degree in Mathematics and Computer Science Part III
MEng Honours Degree in Mathematics and Computer Science Part III
MEng Honours Degrees in Computing Part III
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

MACHINE LEARNING

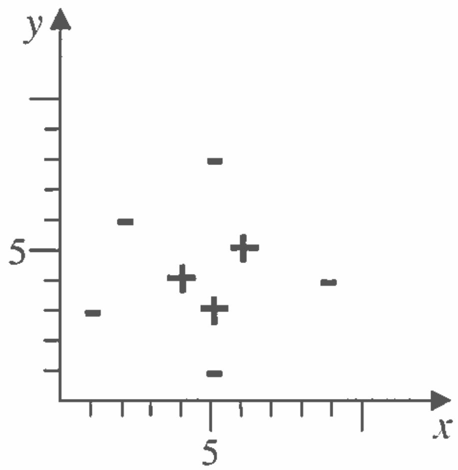
Monday 15 December 2014, 10:00
Duration: 90 minutes

Answer TWO questions

Paper contains 3 questions
Calculators required

1) Consider the instance space consisting of integer points in the x, y plane and the set of hypotheses H consisting of rectangles. More specifically, hypotheses are of the form $a \le x \le b, c \le y \le d$, where a, b, c, and d can be any integers.

Consider further the version space with respect to the set of positive (+) and negative (-) training examples shown in the diagram.



a) Apply the CANDIDATE-ELIMINATION learning algorithm. Write out the intermediate and the final results. Draw the final result on the diagram.

b) Now assume that you are a teacher, attempting to teach the target concept $(3 \le x \le 5, 2 \le y \le 9)$. What is the smallest number of training examples you can provide so that the CANDIDATE-ELIMINATION learning algorithm will perfectly learn the target concept?

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

$$o_d = w_0 + w_1x_1^2 + w_1x_1^3 + \ldots + w_nx_n^2 + w_nx_n^3.$$

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.

d) Given the target function representation defined in 1c, prove that Least Mean Squares (LMS) training rule performs a gradient descent to minimize the cost/error function E defined in 1c.

The four parts carry, respectively, 20%, 20%, 25%, 35% of the marks.

2)

a) Design three single-layer perceptrons, each having a single neuron, which implement the Boolean functions AND, OR and NAND, respectively. Assume that there are two binary inputs to each neuron, x1 and x2, and no bias. Which activation function should be used in each neuron? Evaluate the three perceptrons for all four possible combinations of x1(having a value 0 or 1) and x2 (having a value 0 or 1).

b) Combine the above neurons in two layers such that the resulting network implements the XOR function. Provide a geometric interpretation of the network and explain which region of the 2D input space is classified as 1 and which region is classified as 0.

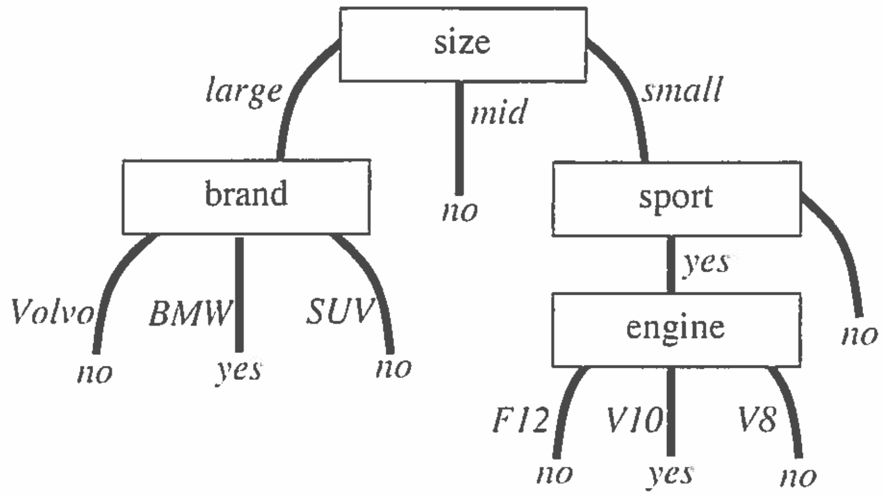
c) Regularization is one way to avoid overfitting while training a neural network. Explain how and why it works. You should also explain how the error function of the neural network should be modified.

d) Assume that you test classifier A on a dataset of 100 examples with the classification rate of 80%. Suppose that more examples (=10000) become available and another classifier B is tested on this new dataset achieving a classification rate of 77%. Compute the 99% confidence intervals for both classifiers and discuss what these intervals mean and which classifier you would trust more for an important application. Provide the equation for calculating confidence interval. The constant z can be computed from the following table:

$N\%$:	50%	68%	80%	90%	95%	98%	99%
z_N:	0.67	1.00	1.28	1.64	1.96	2.33	2.58

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

3 Suppose that we want to solve the problem of finding out what a good car is by using genetic algorithms. Suppose further that the solution to the problem can be represented by a decision tree as follows:



a) What is the appropriate chromosome design for the given problem? Which Genetic Algorithm parameters need to be defined? What would be the suitable values of those parameters for the given problem? Provide a short explanation for each.

b) 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.

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

	size	brand	sport	engine	Good car
1	large	SUV	no	V8	-
2	large	BMW	no	V8	+
3	mid	Ford	yes	V10	-
4	small	BMW	yes	V10	+
5	small	Ferrari	yes	F12	-

Apply a distance-weighted k-Nearest Neighbour algorithm, k=3, to classify the instance <small, BMW, sport, V8 >, assuming that the above-listed examples are already known. Write out the algorithm, the distance function, the weight function, and the intermediate results.

d) Which types of knowledge can we distinguish in Case-Based Reasoning (CBR)? Provide a short explanation of each of the types.

The four parts carry, respectively, 25%, 25%, 35%, 15% of the marks.