IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2015-2016

BEng Honours Degree in Computing Part III
MEng Honours Degree in Electronic and Information Engineering Part IV
MEng Honours Degree in Mathematics and Computer Science 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)
MRes in High Performance Embedded and Distributed Systems
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 14 December 2015, 10:00
Duration: 90 minutes

Answer TWO questions

Paper contains 3 questions
Calculators required

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

	house	trance	detroit	dutch	Pete likes
1	acid	goa	no	moom	+
2	electro	goa	yes	fidget	+
3	acid	beat	yes	fidget	-
4	deep	psycho	yes	moom	-
5	acid	goa	no	fidget	+
6	electro	beat	yes	fidget	+
7	electro	psycho	no	moom	-
8	deep	goa	no	fidget	-
9	acid	psycho	yes	fidget	+
10	acid	beat	no	moom	-
11	deep	beat	no	moom	-
12	electro	goa	no	fidget	-

a) Apply the ID3 algorithm. Write out the intermediate and the final results. Draw the final result as a diagram.

b) Apply the distance-weighted k-Nearest Neighbour algorithm, k=3, to classify the instance <electro, goa, no, moom>, assuming that the above-listed examples are already known. Write out the algorithm, the distance function, the weight function, and the intermediate results.

c) Suppose that we want to solve the problem of finding out what music Pete likes by using genetic algorithms. Suppose further that the solution to the problem can be represented by the result of the ID3 algorithm in 1(a). What is the appropriate chromosome design for the given problem? Which Genetic Algorithm parameters need to be defined? 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 and writing down all intermediate results.

The three parts carry, respectively, 30%, 35%, 35% of the marks.

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

	house	trance	detroit	dutch	Pete likes
1	acid	goa	no	moom	+
2	deep	psycho	yes	moom	-
3	acid	goa	no	fidget	+
4	electro	psycho	no	moom	-
5	deep	goa	no	fidget	-
6	acid	goa	yes	fidget	+
7	deep	beat	no	moom	-

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

b) What is the difference between the CANDIDATE-ELIMINATION and ID3 algorithms?

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

$$o_d = w_0 + w_1 x_1 + w_1 x_1^2 + w_1 x_1^3 + \ldots + w_n x_n + w_n x_n^2 + w_n x_n^3.$$

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 2c, prove that Least Mean Squares (LMS) training rule performs a gradient descent to minimize the cost/error function E defined in 2c.

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

3.

a) Compute the update rule for the weights in the output layer of a neural network using gradient descent. Assume that the sigmoid function is used as an activation function and the squared loss as the error function.

b) Based on the update rule from (a) show that the squared loss is not a good error function when the sigmoid is used as an activation function for the output neurons.
What is a good error function in this case?

c) Dropout is one way to avoid overfitting while training a neural network, especially if the network is deep. Explain how and why it works and how the network should be modified during test time. What are usually the optimal values of the retention probability for the input and hidden neurons?

d) Compute the recall, precision and F1 rates for each class for the given confusion matrix. Do you think some of the computed performance measures are misleading? If yes, explain what the problem is and propose a solution.

	Class 1 - Predicted	*Class 2 - Predicted*
Class 1 - Actual	*400*	*100*
Class 2 - Actual	*0*	*20*

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