FINAL

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

EXAMINATIONS 2018

BEng Honours Degree in Computing Part III
BEng Honours Degree in Electronic and Information Engineering Part III
MEng Honours Degree in Electronic and Information Engineering 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 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

PAPER C395

MACHINE LEARNING

Thursday 22 March 2018, 10:00
Duration: 90 minutes

Answer ALL TWO questions

Paper contains 2 questions
Calculators not required

1.

a) Derive 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, the quadratic loss as the error function and both L1 and L2 regularisation are applied.

b) What is the learning rate? What is its effect on training when it is too low or too high? What is a good updating policy for the learning rate during training?

c) Describe what early stopping is and why it is used. Draw a figure which explains how it works. What is the main assumption made by early stopping and what happens when it is violated?

d) Assume that you test classifier A on a dataset of 50 examples and the classification rate is 88%. Then you collect more examples (=5000) and you test another classifier B on this new dataset which achieves a classification rate of 84%. 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. 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.

2.

a) You are given a dataset with 150 examples. How would you implement cross-validation for test set performance estimation and parameter optimization? Describe in detail all the necessary steps. How would you train a single model in order to classify new unseen examples?

b) Compute the classification rate for the given confusion matrix. Compute the unweighted average recall and the F1 measure for each class. Which performance measure is suitable in this case? Explain why and what the issue is with the given test set.

	Class 1 - Predicted	Class 2 - Predicted	Class 3 - Predicted
Class 1 - Actual	20	0	0
Class 2 - Actual	40	700	20
Class 3 - Actual	40	40	0

c) 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	-

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

d) 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.

e) What is the difference between the CANDIDATE-ELIMINATION and ID3 algorithms? What is the relationship between the learned decision tree in (3d) and the version space that is learned from the same examples in (3c)?

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