COMP 61011

Two hours

EXAM PAPER MUST NOT BE REMOVED FROM
THE EXAM ROOM AND MUST BE RETURNED

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Foundations of Machine Learning

Date: Friday 20th January 2017

Time: 09:45 - 11:45

Please answer ALL Questions provided

Section A and Section B should be answered in separate answerbooks.

Answer ALL Questions in Section C
Write your answers directly on the exam paper. Only answers written in the boxes on the exam paper will be marked.

This is a CLOSED book examination

The use of electronic calculators is permitted provided they are not programmable and do not store text

[PTO]

Section A
Answer *all* questions

1.

i) State the algorithm for Random Forests.

(4 marks)

ii) State two advantages of Random Forests over a simple decision tree.

(2 marks)

2.

i) Derive Bayes Theorem from the definition of conditional probability and symmetrical considerations.

(3 marks)

ii) At a college there are 10 students from France (5 girls & 5 boys), 15 from the UK (5 girls & 10 boys), and 20 from Canada (5 girls & 15 boys). If we choose a student at random what is the probability that this student is French?

(1 mark)

iii) Using the same college data, if we notice that the student chosen at random is a boy, how does this change the probability that the student is French? Show your working.

(2 marks)

iv) Explain what you understand by the statement 'the probability of it raining tomorrow in Manchester is 0.5.'

(1 mark)

v) I give you a coin and tell you that it is biased (i.e. the probability of tossing it and getting a head is not the same as tossing it and getting a tail), but I tell you no more information. Why might it be still rational to assume that the probability of heads and tails is 0.5?

(1 mark)

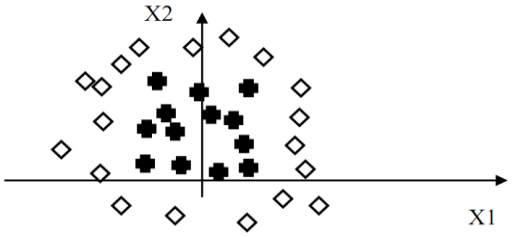
vi) There is evidence that TV weather shows systematically predict worse weather than actually occurs. Given what you know about decision theory why might this be rational for them to do so?

(1 mark)

Section B
Answer *all* questions

1. For a linear SVM, what is the reason behind to minimise $\| w \|$, where w is the parameter vector of a linear decision boundary and $\| w \|$ is the length of w? (1 Mark)

2. Suppose that you apply SVM to classify the dataset as follows:



Now answer the following questions:

i) Please write down the kernel function you want to use. (1 Mark)

ii) Assume that you have applied SVM with your chosen kernel and its parameter. The resulting classifier is underfitted. Are you going to increase or decrease the parameter of your chosen kernel function in order to achieve better fit? (1 Mark)

3. You are applying an SVM with the following error function to classify a dataset

$$E = \sum_{i=1}^{N} \max\{0, 1 - y_i f(x_i)\} + \frac{1}{2} \sum_{j=1}^{d} w_j^2$$

where (x_i, y_i) $(i = 1,2,...,N)$ are the training data, $f(x)$ is an SVM, and w_j $(j = 1,2,...,d)$ are its parameters. After a few tries, you realise that there are some "outlier" datapoints. Now answer the following questions:

i) How do you modify the above error function to handle these "outlier" datapoints? (1 Mark)

ii) What is the meaning of each new item in your modified error function above? (1 Mark)

4. In comparison with Perceptron classifiers, list three advantages of SVM classifiers. (3 Mark)

5. Assume that a decision tree is built using ID3 based on information gain to classify the following dataset

X1	X2	Label Y
1	1	1
1	1	1
0	1	0
0	0	1
1	0	0
1	0	0

Now answer the following questions:

i) What is the entropy of variable Y before the split? (1 Mark)

ii) Which attribute should be selected for the first split and why? You need to calculate the related information gains and then justify your answer. (Note. You need to give the formulas for your calculation rather than just give answers.) (6 Marks)

[PTO]

Section C

This Section contains Multiple Choice Questions and is therefore restricted