Two hours

QUESTION 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: Monday 19th January 2015
Time: 14:00 - 16:00

Please answer ALL Questions

Section A should be answered in an Answer book.

Section B is multiple choice and should be answered 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
Section A

Answer all questions.

1. The accuracy and error rate, on both the training and testing datasets, is the most commonly used way of evaluating a model. Other than these, state two (2) additional factors that could be used to assess the performance of a model and the efficiency of its learning algorithm. (2 marks)

2. What is the main difference between the decision boundary induced by a Perceptron, and the decision boundary induced by a linear SVM? (2 marks)

3. State the formula for the entropy. Using whatever log base you wish (but you must state which) calculate the entropy of the feature \( x = \{10, 10, 10, 10, 2, 2, 2, 0, 0\} \). (2 marks)

4. For each of the FOUR models below, state ONE parameter that controls the fit of the model. For each of these, state whether increasing or decreasing the parameter causes the model to become more finely tuned to the training data.

   Logistic regression,
   Support Vector Machine with a Polynomial kernel,
   K-nearest neighbour,
   Bayesian Network. (8 marks)
5. State the assumption made by the Naive Bayes model, in the form of an equation (1 mark) and its meaning in words (1 mark). Now draw the Bayesian Network corresponding to this assumption (1 mark).

(3 marks)

6. Given the data in the table below, identify all the probabilities that a Naive Bayes classifier would need to calculate during training. Use these probabilities to calculate what probability the model will predict for $p(y=1|x)$, given input $x = \{1, 0, 0\}$.

(4 marks)

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

7. There is a horrible disease in the Manchester student population, “Examophobia”, but 96% of all students are immune to it. The reliability of a particular skin test for Examophobia is as follows: if the student has the disease, the test correctly identifies it 98% of the time. If the student does not have it, the test comes back negative 99% of the time. Imagine we take the entire student population of the University of Manchester, 50,000 students (52% female, 48% male) and a random person is chosen. It’s you. Your test comes back “positive” for Examophobia.

What is the probability that you have Examophobia? (4 marks)
Section B contains Multiple Choice Questions and is therefore restricted