Two hours

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Machine Learning

Date: Friday 29th January 2010

Time: 14.00 – 16.00

Answer ALL 10 short questions in Section A
Answer ONE question from Section B
Answer ONE question from Section C

Use separate Answerbooks for EACH section

The use of electronic calculators is permitted provided they are not programmable and do not store text
1. Multiple Choice Questions – not published.
Section B

Answer ONE of the two questions

2. a) State the equation for the Perceptron decision rule, being sure to fully explain the following concepts: activation, threshold, decision boundary. (5 marks)

b) State the update rule for Perceptron learning, and use a small example to explain how and why it works. (5 marks)

c) Read the following passage, and then answer the question that follows it.

James works for a bank. He is given a historical database of 10,002 customers, and told to apply machine learning algorithms, enabling the bank to predict credit-worthiness of future customers. In the database, 502 of the customers have not paid back their loans, and so are considered unworthy of credit. Among these, James finds 2 customers who are bankrupt multi-millionaires, with overdrafts several thousand times the normal amount, so discards them as he thinks they are likely to adversely affect the learning algorithms.

The remaining 10,000 customers are characterised by their age, sex, living conditions, and bank transactions over the last 5 years. When choosing features to give to his algorithms, James includes only the last 6 months of transactions, and summarises the rest of the 5 years by calculating each customer’s average overdraft amount.

James uses the 10,000 customers as input to the Perceptron learning algorithm, using the learning rate 0.25. Because of the Perceptron convergence theorem, he calculates that it should correctly classify at least 0.25 x 10,000 = 2,500 of the customers, and in fact it vastly exceeds his expectations, correctly classifying 9,500 of them. James happily concludes that the data must be linearly separable, but would like better performance.

Due to his conclusion on the Perceptron, he decides to apply a linear classifier, so picks the K-NN rule. He sets k=10,000, and finds that he can again only correctly classify 9,500 customers, but is not sure why. He concludes that a Perceptron with learning rate 0.25 is optimal, and recommends this to the bank.

List 5 things that are wrong with James’s methodology and understanding of Machine Learning, giving reasons for each, and explain to him why his K-NN rule behaved in the way it did. (10 marks)
3. a) Give a definition of “overfitting”, including:

- how it can be detected in general,
- in decision tree classifiers: how it manifests, and how to control it. (5 marks)

b) State the equation to calculate the entropy of a feature. Apply this to an example binary feature x with \( p(x=1) = 0.7 \), and \( p(x=0) = 0.3 \). (5 marks)

c) A company is analysing financial trades over the last year to determine what caused the current financial crisis. It has compiled a dataset of 500 banks and financial institutions worldwide, and asked the Chief Financial Officer of each bank 10 yes/no questions on his management style, and 10 questions about the amounts of their current investments. Note that the answers to the second set of questions were continuous variables. In addition they have records of whether each bank is currently judged as being “in crisis” or “not in crisis”.

Describe how the ID3 algorithm works, and how it should be applied to this data. (10 marks)
Section C

Answer ONE of the two questions

4. A bank wishes to find a way to make initial binary decisions (acceptance or denial) automatically for credit card applications. A client was given a database that contains a number of previous applications where decisions were made by bank experts based on four critical features \((x_1, x_2, x_3, x_4)\). The client decided to train a Perceptron to produce the linear classifier with all the examples in the database for this task. With different initial conditions, however, the Perceptron training resulted in the following three linear classifiers of different weights but all have 100% accuracy on the training data set:

\[
H_1(x) = x_1 + x_2 + x_3 + x_4 + 1, \\
H_2(x) = 1.41 + 1.5x_2 + 1.2x_3 + 0.9, \\
H_3(x) = 1.2x_1 + 1.3x_2 + x_3 + 1.1x_4 + 1.5.
\]

As a result, you are going to help him solve the following problems.

a) Determine which classifier is likely to be the best one to make correct decisions for the future applications in general. (10 marks)

b) Determine whether or not there is another linear classifier that outperforms the three existing ones with the same training data set. (4 marks)

c) After another validation data set is collected, the best linear classifier achieved previously does not yield satisfactory performance, i.e., a high error rate on the validation data set. You are asked to use both previous training database and the validation data set to establish another classifier in a simple way that is likely to improve the performance in decision making for future applications. (6 marks)

You need to give details to justify your answers to all the above questions.
5. A doctor presents you with a list of test results with four measurements on five patients as follows.

<table>
<thead>
<tr>
<th>ID</th>
<th>Test-1</th>
<th>Test-2</th>
<th>Test-3</th>
<th>Test-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Negative</td>
<td>Negative</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>2</td>
<td>Positive</td>
<td>Negative</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>3</td>
<td>Negative</td>
<td>Positive</td>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td>4</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>5</td>
<td>Positive</td>
<td>Positive</td>
<td>Positive</td>
<td>Negative</td>
</tr>
</tbody>
</table>

He wishes to group patients based on the above test results for further diagnosis. You are going to apply the agglomerative clustering algorithm to this problem.

a) Describe an appropriate representation of test results and then calculate the distance matrix with a proper distance metric that best reflects the nature of medical data. (10 marks)

b) Based on the distance matrix achieved in a), draw a dendrogram tree with the single link to reflect the entire grouping relation among patients. It is essential to describe how you achieve this dendrogram representation. (6 marks)

c) Based on the dendrogram representation achieved in b), make a recommendation on how many patient groups are the most likely to form in terms of test results and what is the membership of each group. Explain why. (4 marks)