UNIVERSITY OF MANITOBA
Midterm

Winter 2003

COMPUTER SCIENCE

Machine Learning

Date: 8 March 2004
Time: 15:30 - 16:30
Room: Armes Building 115, University of Manitoba
(Time allowed: 50 Minutes)

NOTE: Attempt all questions.
This is a closed book examination.
Use of calculators is permitted.
Show your work to receive full marks.

SURNAME: ___________________________
FORENAME(S): _________________________
STUDENT ID: _________________________

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>100</td>
</tr>
</tbody>
</table>

CONTINUED
Section A: Candidate Elimination

1. Given below is a generalization hierarchy for the clothing domain.

Starting with \( S\text{-set}=0, G\text{-set}=\langle ?, ? \rangle \), the system is trained with a series of training instances. After training, the S-set and G-set are now as follows:

\[
\begin{align*}
S\text{-set}: & \langle \text{large sized, blue} \rangle \\
G\text{-set}: & \langle ?, \text{colored} \rangle
\end{align*}
\]
How many different concepts are still included in the version space (i.e., are consistent with the series of training instances)? List all concepts that are still included in the version space. If it is impossible to determine the number of concepts remaining in the version space or impossible to determine the list of concepts remaining in the version space, then say so in your answer and explain why it is impossible.

[5 marks]

4 concepts
VS = <LS, blue>, <LS, colored>, <*>blue>, <*>colored>

2. Show the trace of the execution of the Candidate Elimination algorithm on the clothing domain given the following training sequence. For each instance, show the resulting S and G sets. I already did the first entry for you.

[10 marks]

<S, other> + S-Set = <S, other>
G-Set = <?, ?>
<S, other> + S-Set = <Small sized, other>
G-set = <?, ?>
<XL, green> - S-Set = <Small sized, other>
G-set = <Small sized, ?><?, other>
<L, red> - S-Set = <Small sized, other>
G-set = <Small sized, ?><?, other>
<S, other> + S-Set = <Small Sized, other>
G-set = <Small Sized, ?><?, other>
<S, green> - S-Set = <Small Sized, other>
G-set = <?, other>

3. Given the generalization hierarchy of the clothing domain shown above, what is the maximum size (i.e., number of sets) of the G set that can be generated with a single instance. Show one instance and its classification (positive or negative) so that it generates a maximum G set.

[10 marks]
Section B: Decision Trees

The information gain $\text{Gain}(S,A)$ of an attribute $A$ for a sample set $S$ is defined as

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropy}(S_v)$$

A graph of the entropy function is shown in the figure below. You can use this graph when answering the following questions.

4. Given below is a set of instances from a hard disk manufacturing domain. The goal is to determine whether a HD is defective within the first 6 months.
5. Given below is a set of rules for a medical diagnosis domain. The attribute \textit{blood} refers to the blood pressure of the patient, the attribute \textit{stress} refers to the reported stress level. The rules in the rule set are ordered, which means that as soon as the preconditions of one rule are satisfied, the corresponding diagnosis will be output by the system.

For example, the instance \textit{stress=low and blood=high} will be classified as \textit{Yes}, because of rule 2 which is evaluated before rule 3.

Is it possible to convert these rules into a decision tree? If so, show the reduced decision tree (i.e., \textit{the decision tree with a minimal number of nodes}) for this problem domain, otherwise explain why these rules can not be converted into a decision tree.

Rule 1: \textit{if blood=normal and stress=low then No}
Rule 2: \textit{if blood=high then Yes}
Rule 3: \textit{if stress=low then No}
Rule 4: \textit{if stress=high and blood=low then Yes}
Rule 5: \textit{if stress=high then Yes}
Section C: Neural Nets

6. Given below is an artificial neural network (ANN) with three input nodes (X1,X2,X3), two hidden nodes, and one output node. The network uses simple threshold nodes (i.e., the node will output 1.0 if the sum of the weighted inputs is greater than the threshold, 0 otherwise).

You are trying to learn the boolean target function $f_1$.

Show a set of weights and thresholds for all nodes that implement the function $f_1$. If it is impossible to represent the function $f_1$ with the given neural network, then state this in your answer and explain why this is impossible.

[15 marks]
7. The perceptron theorem shows that any function that can be computed by a multi-layer network of linear threshold units can be computed by a network with only two layers.

The output $y$ of a linear threshold unit $n$ is given by the following formula:

$$y_n = \theta_n + \sum_i w_{in}x_i$$

where $\theta_n$ is the weight (threshold) of the node itself, $w_{in}$ is the weight of the link from node $i$ to node $n$ and $x_i$ is the input on the link from $i$ to $n$. The summation is over all inputs to node $n$.

$f_3(X0, X1, X2)$ is the function computed by the following network $N$. 

---

CONTINUED
Show the link and node weights for the two layer network $N'$ below, so that network $N'$ computes the same function $f_3$. If it is impossible to find such a set of weights, then say so in your answer and explain why.

Weights of links are given beside the link
Weight of nodes are given inside the node

[10 marks]

CONTINUED
8. A humanoid robot has a gyroscope and a force sensor to detect whether it is about to fall over. The associated probabilities are shown below.

<table>
<thead>
<tr>
<th>Gyroscope</th>
<th>Force sensor</th>
<th>Falling</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>N</td>
<td>N</td>
<td>0.2</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>0.1</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>0.0</td>
</tr>
<tr>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>0.1</td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>0.3</td>
</tr>
<tr>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>0.2</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>0.0</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Calculate the accuracy of the gyroscope if the robot is falling over. [5 marks]

\[
\text{The accuracy of the gyroscope if the robot is falling over is } \frac{0.2 + 0.1}{0.1 + 0.1 + 0.2 + 0.1} = 0.6
\]

9. What is the probability of the robot falling over if the gyroscope and the force sensor indicate a fall? [5 marks]

\[
\text{The probability of the robot falling over if the gyroscope and the force sensor indicate a fall is } \frac{0.1 + 0.0}{0.1 + 0.0} = 1.0
\]

10. Determine if the readings of the gyroscope are statistically independent of the force sensor or not. Show your work for full marks. [5 marks]

CONTINUED
The readings of the gyroscope and the readings of the force sensor are statistically dependent.
For statistical dependence of \( G \) and \( F \) is given iff:
\[
P(G=Yes) = P(G=Yes|F=Yes) \text{ (and } P(G=Yes|F=No))
\]
\[
P(G=Yes) = 0.3+0.2+0.0+0.1 = 0.6 \quad P(G=Yes|F=No) = 0.3+0.2 = 0.5
\]
Additional work pages
Additional work pages