UNIVERSITY OF MANITOBA
Midterm

Winter 2004

COMPUTER SCIENCE

Machine Learning

Date: **9 March 2005**
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 non-programmable calculators is *permitted*.
Use of any other electronic equipment is strictly forbidden.
Show your work to receive full marks.

SURNAME:

FORENAME(S):

STUDENT ID:

A	B	C	D	Total
25	25	25	25	100

Surname: _____

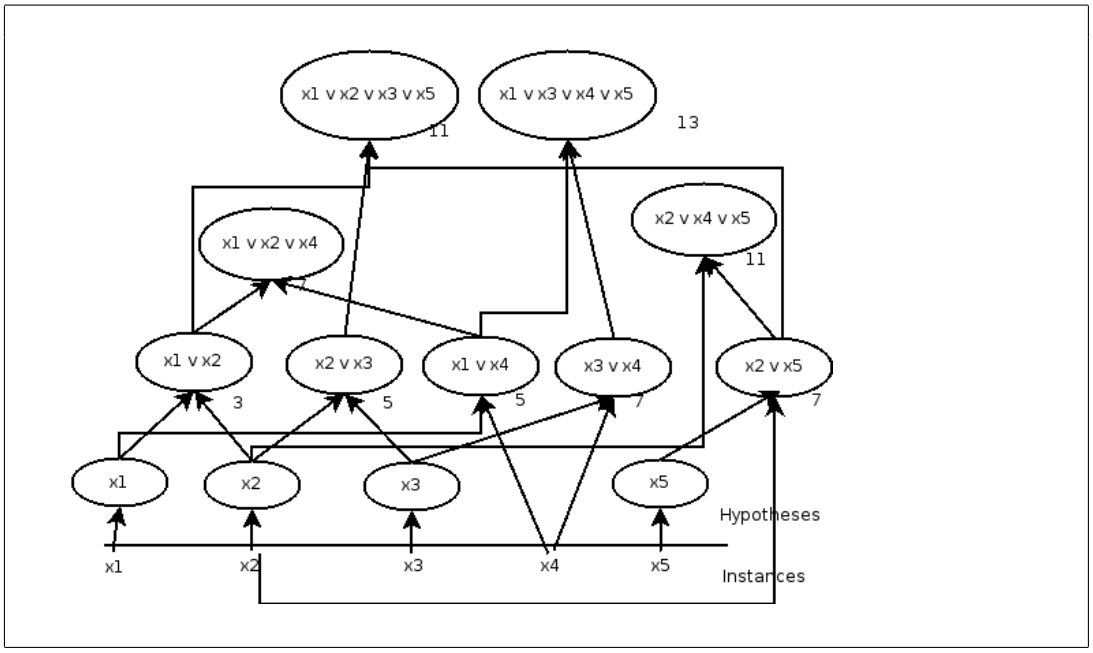
Forename(s): _____

Section A: Candidate Elimination

1. Given is a domain with five instances: $x_1, x_2, ...x_5$. The hypothesis space H consists of all concepts h such that the sum of the indices of all positive instances x_i of hypothesis h add up to a prime number (i.e., 1,2,3,5,7,11,13). For example, the hypothesis ($x_5 \cup x_4 \cup x_3 \cup x_1$) must be present in your generalization hierarchy because 5+4+3+1=13, which is a prime number. Note that there may be more than one hypothesis for every prime number. For example, the hypotheses $x_2 \cup x_3$, x_5, and $x_1 \cup x_4$ must **all** be included in your hypothesis space, since they all add up to 5.

   In the answer box below, show the generalization hierarchy for this hypothesis space. Show the positive instances for all hypothesis as well as the generalization links that form your hypothesis space. Some of the entries are already shown for you.

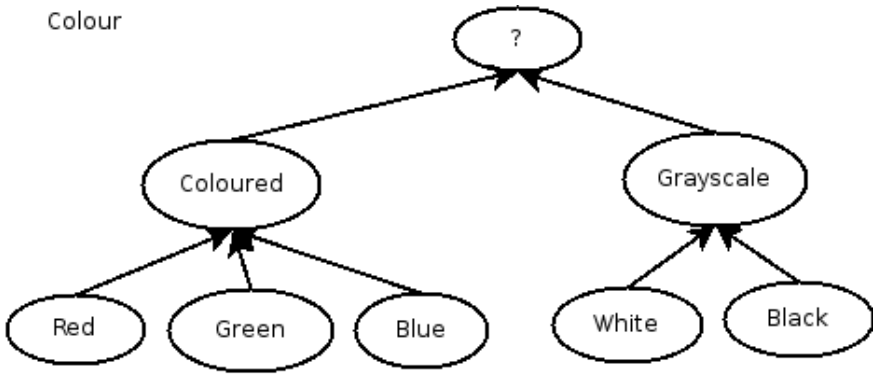
   [15 marks]

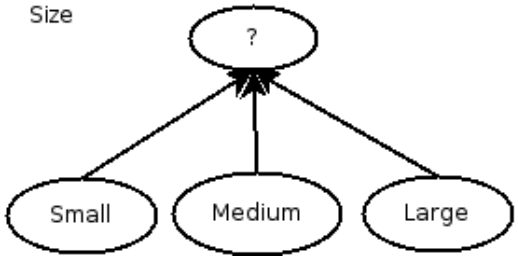


2. Show the trace of the execution of the Candidate Elimination algorithm on the clothing domain given the following training sequence. For each instance, show the resulting S and G sets. I already did the first entry for you.

Surname: _____

Forename(s): _____





[10 marks]

```
<black,small>   +  S-Set = <black,small>
                   G-Set = <?,?>
<white,medium>  -  S-Set = <black,small>
                   G-Set = <black,?>,<?,small>
<red,medium>    -  S-Set = <black,small>
                   G-Set = <black,?>,<?,small>
<blue,small>    -  S-Set = <black,small>
                   G-Set = <black,?>,<grey,small>
<black,medium>  -  S-Set = <black,small>
                   G-Set = <grey,small>
<black,large>   -  S-Set = <black,small>
                   G-Set = <grey,small>
```

Surname: _____

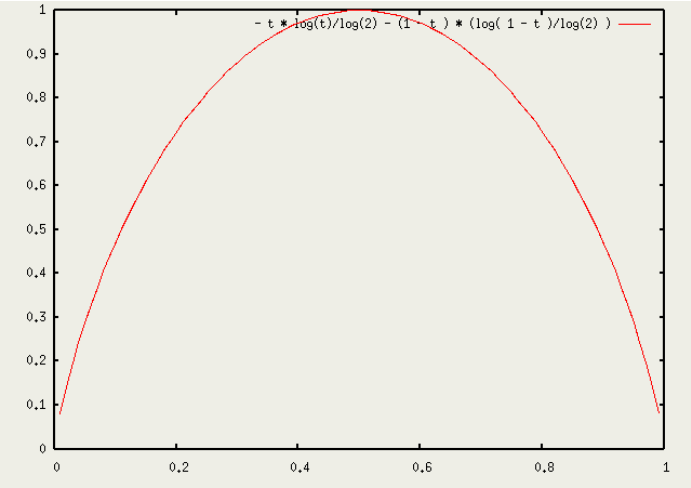
Forename(s): _____

Section B: Decision Trees

The information gain 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.



3. Given below is a set of instances from a hardware diagnosis domain. The target function is to specify whether a particular memory chip fails within six month. During transmission of the data, some of the size attributes were corrupted. The unknown entries are shown with a size attribute of "?."

Instance	Size	Company	Type	Target
x1	128M	A	DDR	Yes
x2	128M	B	SRAM	Yes
x3	?	A	SRAM	Yes
x4	128M	B	DDR	No
x5	256M	A	DDR	No
x6	128M	B	SRAM	No
x7	128M	A	SRAM	No
x8	?	B	DDR	No
x9	128M	B	DDR	No
x10	128M	A	DDR	Yes

Surname: _____

Forename(s): _____

A colleague tells you that the information gain of the size attribute in a previous run of the ID3 algorithm with the uncorrupted data was greater than 0.15 (i.e., Gain$(S, \text{Company}) > 0.15$). You also know that the companies A and B only make memory chips with 128M and 256M capacity.

Is this information sufficient to uniquely determine the missing attribute value Size for instances x3 and x8?

If yes, then show the missing attribute values for x3 and x8. If there is not enough information to uniquely determine the missing attribute values, then state this in your answer and explain why. If no possible assignment of attribute values result in an information gain that is greater than 0.15 for then state this in your answer and explain why.

[15 marks]

```
-->ent(6,4) - 7/10 * ent(3,4) - 3/10 * ent( 1, 2)
    0.0058021
-->ent(6,4) - 8/10 * ent(4,4) - 2/10 * ent( 2, 0)
    0.1709506
-->ent(6,4) - 8/10 * ent(3,5) - 2/10 * ent( 1, 1)
    0.0074034
-->ent(6,4) - 9/10 * ent(4,5) - 1/10 * ent( 1, 0)
    0.0789821
```

x3 must be 128M, x8 must be 256M.

4. Given below is a set of rules for a medical diagnosis domain. The attribute *blood* refers to the blood pressure of the patient, the attribute *size* refers to the height of the patient.

   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 blood = low and size = tall will be classified as Yes, because of rule 2 which is evaluated before rule 3.

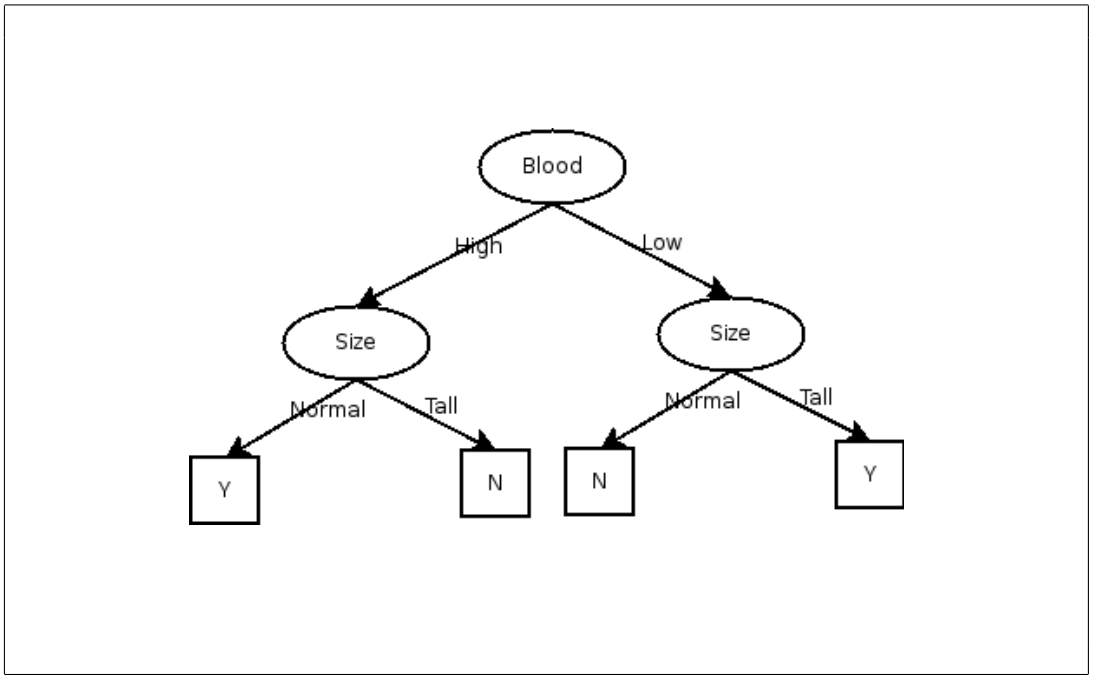
   Is it possible to convert these rules into a decision tree? If so, show the reduced decision tree (i.e., *the decision tree with a minimal number of nodes*) for this problem domain, otherwise explain why these rules can not be converted into a decision tree.

   Rule 1: *if* blood=high and size=normal *then* Yes
   Rule 2: *if* blood=low and size=tall *then* Yes
   Rule 3: *if* size=tall *then* No
   Rule 4: *if* blood=low and size=normal *then* No
   Rule 5: *if* blood=high *then* Yes

   [10 marks]

Surname: _____

Forename(s): _____



Section C: Neural Nets

5. 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).

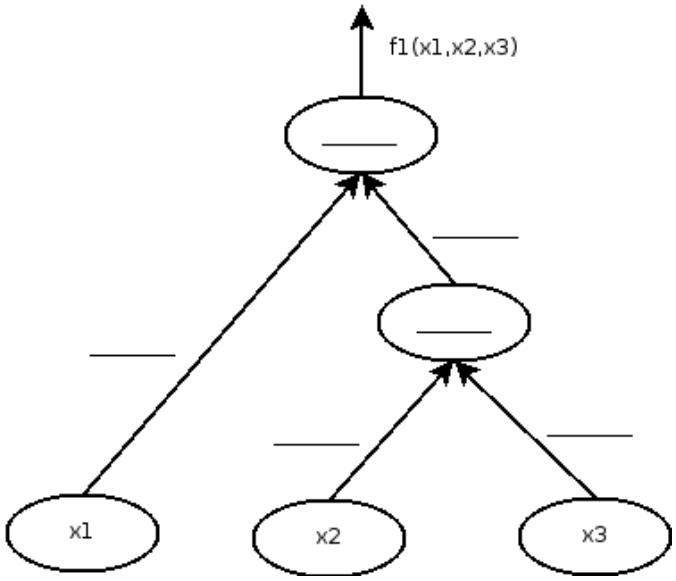
   You are trying to learn the boolean target function f_1.

   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.

   [15 marks]

Surname: _____

Forename(s): _____



x_1	x_2	x_3	f_1
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	1

It is impossible to implement f_1 using this network since it would require to compute the XOR of x_2 and x_3. Another way of proving it is that the given weights only allow you cut in the x_2, x_3 plane and the x_1 plane.

6. Is it possible to represent the **continous** function f_3 using an ANN?

$$f_3(x_1, x_2) = \begin{cases} 1 & \text{if } x_1 >= 2 * x_2 \\ 0 & \text{else} \end{cases}$$

Show the network with the **minimum** number of nodes that computes f_3.

[10 marks]

Surname: _____

Forename(s): _____

The network must have at least 1 node(s).
Weight for x_2 is twice the weight for x_1, threshold at 0+.

Section D: Bayesian Learning

7. 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.

Gyroscope	Force sensor	Falling	Probability
N	N	N	0.1
N	N	Y	0.1
N	Y	N	0.3
N	Y	Y	0.1
Y	N	N	0.1
Y	N	Y	0.2
Y	Y	N	0.0
Y	Y	Y	0.1

Calculate the probability of the robot falling over, given that no sensor feedback is available (i.e., the prior probability):

[7 marks]

Surname: _____

Forename(s): _____

The probability of the robot falling over is 0.5

$$0.1 + 0.1 + 0.2 + 0.1$$

8. What is the accuracy of the gyroscope sensor given that the force sensor indicates a fall.

[8 marks]

The accuracy of the gyroscope sensor given that the force sensor indicates a fall is 0.8

$$\frac{0.3 + 0.1}{0.3 + 0.1 + 0.0 + 0.1}$$

9. Currently, the **correctness** of the gyroscope is statistically dependent on the correctness of the force sensor. Make the minimum number of modifications necessary to the table of probabilities below, so that the correctness of the gyroscope and the correctness of the force sensor are **statistically independent**.

   If it is impossible to change the table so that the correctness of the gyroscope and the correctness of force sensor are statistically independent, then say so in your answer and explain why.

[10 marks]

Surname: _____

Forename(s): _____

Gyroscope	Force sensor	Falling	Probability	New probability
N	N	N	0.1	
N	N	Y	0.1	
N	Y	N	0.3	
N	Y	Y	0.1	
Y	N	N	0.1	
Y	N	Y	0.2	0.0
Y	Y	N	0.0	0.2
Y	Y	Y	0.1	

P(Gyro=correct — Force = Correct) = P(Gyro=correct — Force = Not Correct) = P =(p0+p7)/(p0+p3+p4+p7) = (p2 + p5)/(p1+p2+p5+p6)

Surname: _____

Forename(s): _____

Additional work pages

Surname: _____

Forename(s): _____

Additional work pages