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:

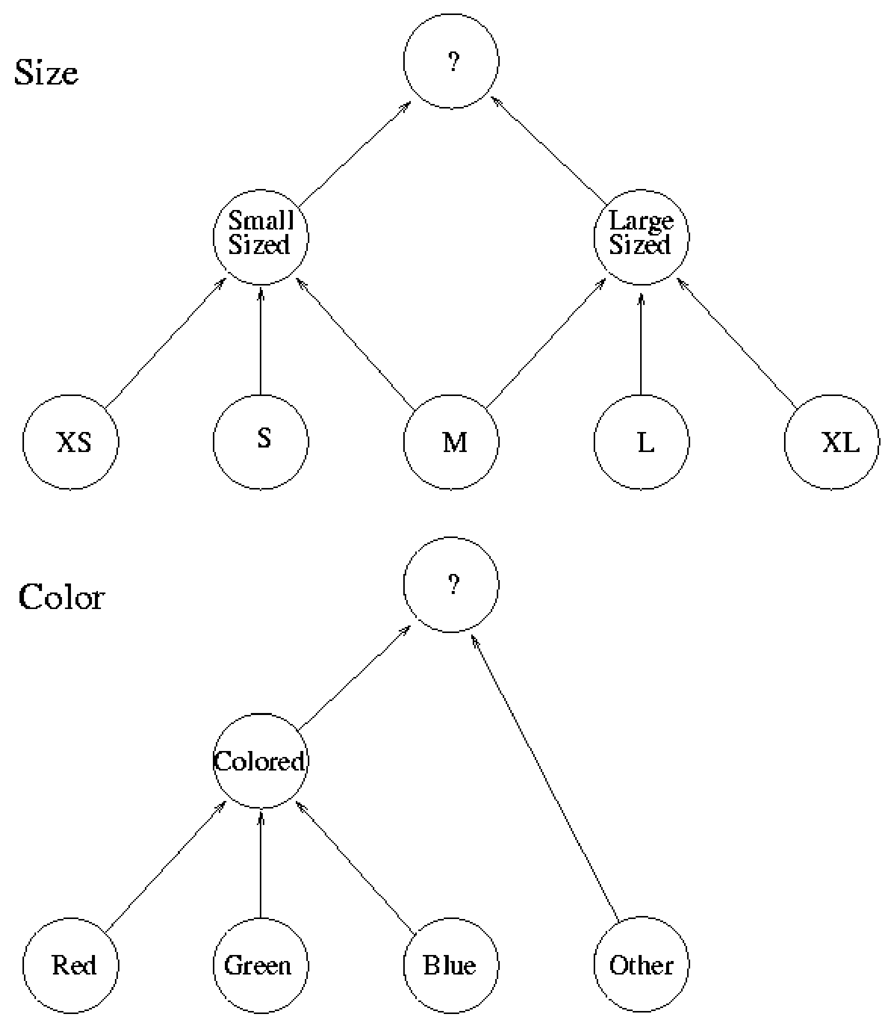
A	B	C	D	Total
25	25	25	25	100

Surname: ______________________________

Forename(s): ______________________________

Section A: Candidate Elimination

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



Starting with `S-set=0`, `G-set=<?,?>`, the system is trained with a series of training instances. After training, the S-set and G-set are now as follows:

```
S-set: <large sized, blue>
G-set: <?, colored>
```

CONTINUED

Surname: ______________________________

Forename(s): ______________________________

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 = <?,?>
<M, 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]

Surname: ____

Forename(s): ____

After a single instance, the G set contains at most _____7_____ sets.
<Medium, Red> -
G-Set: <XS,?>,<S,?>,<L,?>,<XL,?>,<?,Green>, <?, Blue>, <?,Other>

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.

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.

Surname: ____________________

Forename(s): ____________________

Size	Series	Plant	Target
40GB	X	Singapore	No
60GB	Y	US	Yes
80GB	X	Singapore	No
40GB	X	US	No
40GB	Y	Singapore	Yes
60GB	Y	US	Yes
80GB	X	Singapore	No
80GB	Y	US	Yes
70GB	Y	Singapore	No

Calculate the information gain of the attribute Series (Gain(S, Series)).

[10 marks]

```
Ent[4,5]=0.99
Series=X
  4: Ent[0,4] = 0
Series=Y
  5: Ent[4,1] = 0.72
Information gain = 0.99 - 4/9*0 - 5/9*0.72 = 0.56
```

5. 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 *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 stress=low and blood=high 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=normal and stress=low *then* No
Rule 2: *if* blood=high *then* Yes
Rule 3: *if* stress=low *then* No
Rule 4: *if* stress=high and blood=low *then* Yes
Rule 5: *if* stress=high *then* Yes

Surname: ______________________________

Forename(s): ______________________________

[15 marks]

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).

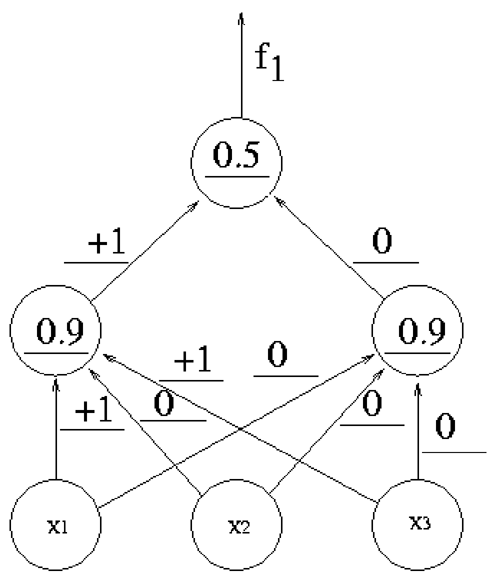
   You are trying to learn the boolean target function f_1.

   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]

Surname: ______________________________

Forename(s): ______________________________



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

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:

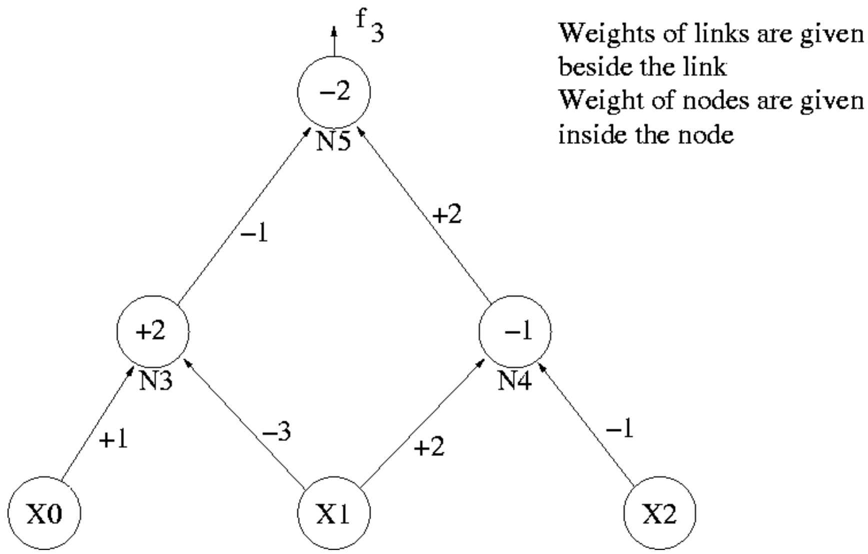
$$y_n = \theta_n + \sum_i w_{in} x_i$$

   where θ_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.

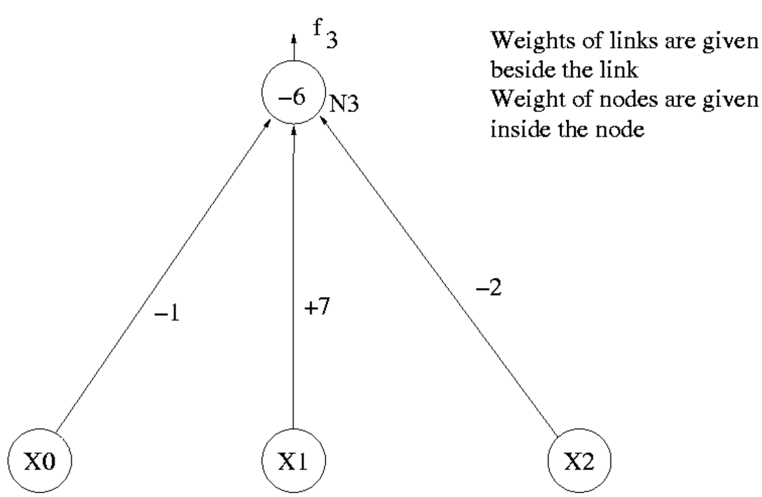
Surname: ______________________________

Forename(s): ______________________________



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.

[10 marks]



Surname: ____________________

Forename(s): ____________________

Section D: Bayesian Learning

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.

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

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

The accuracy of the gyroscope if the robot is falling over is 0.6

$$\frac{0.2 + 0.1}{0.1 + 0.1 + 0.2 + 0.1}$$

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

The probability of the robot falling over if the gyroscope and the force sensor indicate a fall is 1.0

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

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]

Surname: ______________________________

Forename(s): ______________________________

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) (and P(G=Yes|F=No))
P(G=Yes) = 0.3+0.2+0.0+0.1 = 0.6 P(G=Yes|F=No) = 0.3+0.2 = 0.5

CONTINUED

Surname: ____________________

Forename(s): ____________________

Additional work pages

Surname: ______________________________

Forename(s): ______________________________

Additional work pages