UNIVERSITY OF MANITOBA

Winter 2003

COMPUTER SCIENCE

Machine Learning

(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	20	25	95

Surname: ______________________________

Forename(s): ______________________________

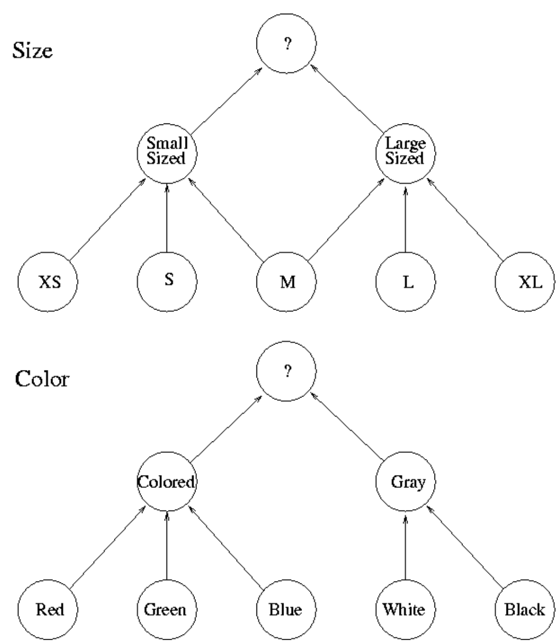
Section A: Candidate Elimination

1. You apply the candidate elimination algorithm to a machine learning problem. However, after the third example, you notice that the version space has collapsed (i.e., the S- and G-sets are empty.) State and explain two possible causes for this.

[10 marks]

A version space may collapse because:
1.) The correct hypothesis is not included in the version space. For example, trying to learn a disjunctive concept in a conjunctive version space.
2.) The training data may be noisy, that is a positive/negative instance is wrongly classified as negative/positive.

2. Given the following generalization hierarchy, trace the execution of the candidate elimination algorithm for the instances shown below. Show the S- and G-set after each new training instance.



Surname: ______________________________

Forename(s): ______________________________

[15 marks]

```
<S, White> - S-Set =   0
             G-Set = {<Large Sized,?>,<XS,?>,<?,black>,<?,color>}

<S, red>   - S-Set =   0
             G-Set = {<Large Size,?>,<XS,?>,<?,black>,<?,green>,<?,blue>}

<L, green> + S-Set = <L,green>
             G-Set = {<Large Sized,?>,<?,green>}
```

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 Fig. 1 below. You can use this graph when answering the following questions.

3. Given below is a set of instances from the domain in question 1 and its classification under the target function.

   Calculate the information gain of the attribute color ($\text{Gain}(S, \text{Color})$).

   [10 marks]

Surname: ______________________________

Forename(s): ______________________________

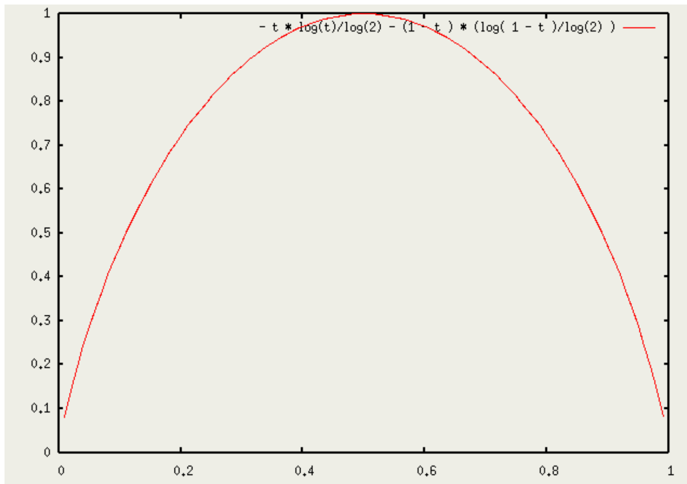


Figure 1: Graph of the Entropy Function

```
S = [4+, 4-]
E(S) = 1.0
S_white = [1+,0-]
E(S_white) = 0.0
S_black = [0,4-]
E(S_black) = 0.0
S_green = [+3,0]
E(S_green) = 0.0
Gain(S,Color) = 1 - ( 1/8 * 0 + 4/8 * 0 + 3/8 * 0 ) = 1
```

4. Select a subset S_2 of at least *three* instances from table 1 above such that the information gain of the attribute size ($\text{Gain}(S_2, \text{Size})$) is equal to 1.

[15 marks]

Surname: ______________________________

Forename(s): ______________________________

Size	Color	Target
S	black	No
L	green	Yes
XL	black	No
XS	black	No
XL	green	Yes
XL	white	Yes
M	black	No
M	green	Yes

Table 1: Training set for the Size, Color domain

It is impossible to get an information gain of 1 with three instances. Since it requires an initial entropy of 1, which can only be achieved with an even number of instances. So to get an information gain of 1, you must have 2+,2- instances that are split into perfectly classified sets using the attribute size. One such set is:
$S_2 =< S, black >, < XL, green >, < XL, white >, < M, black >$

Section C: Neural Nets

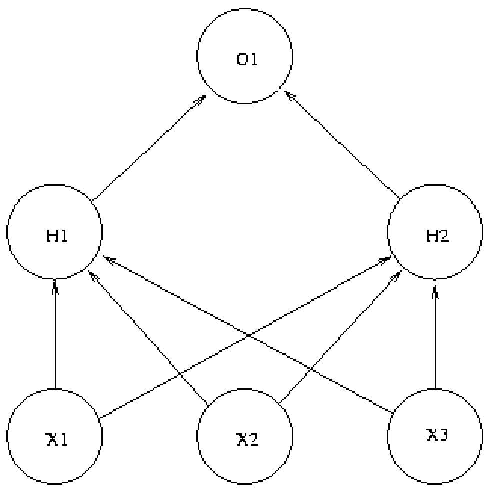
5. Given below is an artificial neural network (ANN) with three input nodes (X1,X2,X3), two hidden nodes (H1,H2), and one output node (O1). The network uses simple threshold nodes (i.e., the node will be fully activated if the sum of the weighted inputs is greater than the threshold). You are trying to learn a boolean target function.

   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 say this in your answer and explain why this is impossible.

   [10 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	0
1	1	0	0
1	1	1	1

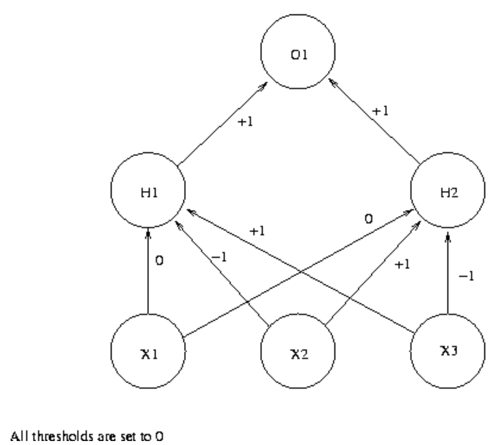
It is impossible to learn this function. The three input nodes form a cube. The two hidden nodes can split the cube into 8 regions, each one contains a positive and negative example. This means that an XOR has to be implemented by the single output node.

6. Show a set of weights and thresholds for all nodes that implement the function f_2. If it is impossible to represent the function f_2 with the given neural network, then say this in your answer and explain why this is impossible.

[10 marks]

Surname: ______________________________

Forename(s): ______________________________



x_1	x_2	x_3	f_2
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	0
1	0	1	1
1	1	0	1
1	1	1	0

Section D: Naive Bayesian Learning

7. To determine the probability that a drill hole contains oil, the following tests can be performed by a geologist.

   (a) A sonar scan (T1) will indicate in 80% of all cases that oil is present if oil is present. The test will wrongly indicate that oil is present in 30% of all cases where there is no oil present.

   (b) A core sample (T2) will report in 90% of all cases that oil is present if oil is present. The test will wrongly indicate that oil is present in 60% of all cases where there is no oil present.

   From experience, the geologist knows that about 33% of all drill holes contain oil.

   What is the probability of oil given that a sonar scan reported a positive test result. [5 marks]

$P(\text{Oil}|\text{T1=Yes}) = \frac{P(\text{T1=Yes}|\text{Oil}) * P(\text{Oil})}{P(\text{T1=Yes})}$
$P(\text{T1=Yes}|\text{Oil}) = 0.8$
$P(\text{Oil} = 0.33$
$P(\text{T1=Yes}) = P(\text{T1=Yes}|\text{Oil}) + P(\text{T1=Yes}|\tilde{}\text{Oil})$ $P(\text{T1=Yes}) = (0.8 * 0.33 + 0.3 * 0.66) = 0.46$
$P(\text{Oil}|\text{T1=Yes}) = 0.8 * 0.33/(0.8 * 0.33 + 0.3 * 0.66) = 0.57$

Surname: ____________________

Forename(s): ____________________

8. What is the probability of oil given that both the sonar scan and the core sample indicate the presence of oil. You can assume that the two tests are conditionally independent.

[10 marks]

$P(\text{Oil}|\text{T1=Yes} \wedge \text{T2=Yes}) =$
$\frac{P(\text{T1=Yes}|\text{Oil})*P(\text{T2=Yes}|\text{Oil})*P(\text{Oil})}{P(\text{T1=Yes}|\text{Oil})*P(\text{T2=Yes}|\text{Oil})*P(\text{Oil})+P(\text{T1=Yes}|\tilde{}\text{Oil})*P(\text{T2=Yes}|\tilde{}\text{Oil})*P(\tilde{}\text{Oil})} =$
$= 0.8 * 0.9 * 0.33/(0.8 * 0.9 * 0.33 + 0.3 * 0.6 * 0.66) = 0.66$

9. Given the following data set (Table 2 for the size,color domain, what is the naive bayesian classification of the new instance `<XS,green>`.

Size	Color	Target 2
XS	green	Yes
L	green	No
XL	black	No
XS	black	No
XL	green	Yes
XS	white	Yes
M	black	No
M	green	Yes

Table 2: Training set for the Size, Color domain

[10 marks]

```
P(Yes) = 4/8 = 0.5
P(No) = 4/8 = 0.5
P(XS|Yes) = 2/4 = 0.5
P(XS|No) = 1/4 = 0.25
P(green|Yes) = 3/4 = 0.75
P(green|No) = 1/4 = 0.25
P(yes) = P(Yes) * P(XS|Yes) * P(green|Yes) = 1/2*1/2*3/4=3/16
P(no) = P(No) * P(XS|No) * P(green|No) = 1/2*1/4*1/4 = 1/32
```

Surname: ______________________________

Forename(s): ______________________________

Additional work pages

Surname: ______________________________

Forename(s): ______________________________

Additional work pages