UNIVERSITY OF MANITOBA
Midterm

Winter 2007

COMPUTER SCIENCE

Machine Learning

Date: **Friday, 2nd March 2007**
Time: **15:30 - 16:20**
Room: **EITC E2-165, 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:

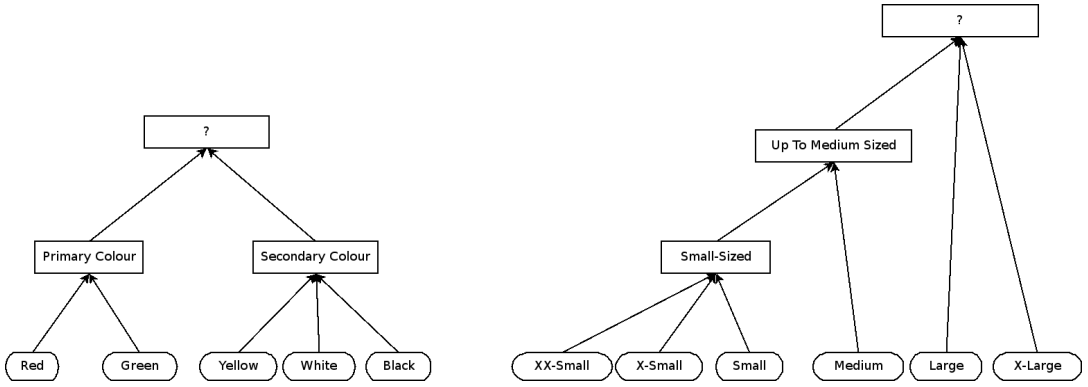
A	B	C	Total
34	35	33	102

Surname: _____

Forename(s): _____

Section A: Candidate Elimination

1. Given below is a conjunctive hypothesis space for a clothing domain D. What is the maximum size of the G-set that can be created with a single example. Show the resulting S and G-sets and the instance that results in the largest possible G-set. Simply select one of the possible examples if there is more than one such example.



[10 marks]

The maximum size of the G-set that can be created by a single negative example is 8

Instance: `<Black,Small>` -

Resulting S-set: `<empty>`

Resulting G-set: `<White,?>,<Yellow,?>,<Primary,?>, <?,XX-Small>,<?,X-Small>,<Medium>,<?,Large>, <?,X-Large>`

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

[13 marks]

Surname: _____

Forename(s): _____

```
<White,Medium>    +  S-Set = <White,Medium>
                     G-Set = <?,?>
<Green,Large>     -  S-Set = <White,Medium>
                     G-Set = <Secondary Colour,?>,<?,Up To Medium Sized>
<Blue,X-Large>    -  S-Set = <White,Medium>
                     G-Set = <Secondary Colour,?>,<?,Up To Medium Sized>
<White,X-Small>   +  S-Set = <White,Up To Medium Sized>
                     G-Set = <Secondary Colour,?>,<?,Up To Medium Sized>
<Yellow,Medium>   -  S-Set = <White,Up To Medium Sized>
                     G-Set = <White,?>
<White,Small>     +  S-Set = <White,Up To Medium Sized>
                     G-Set = <White,?>
<White,X-Large>   +  S-Set = <White,?>
                     G-Set = <White,?>
<White,XX-Small>  -  S-Set = nil
                     G-Set = nil
```

3. What is the minimum number of examples needed to learn the target concept `<Green, Small-Sized>` in the clothing domain shown above.

   Show the minimum training set such that the candidate elimination algorithm will learn the target concept. For each sample in the training set show the classification, the resulting S-set and G-set. One entry of the training set is already given in the answer box below.

   If it is impossible to determine a minimum training set such that the candidate elimination algorithm is able to learn the concept `<Green, Small-Sized>`, then say so in your answer and explain why.

   [11 marks]

The minimum size of the training set is 4 samples.

$D =$

Instance	Classification	S/G-set
<Green,X-Small>	+	S-set: <Green,Small> G-set: <?,?>
<Green,Small>	+	S-set: <Green,Small-Sized> G-set: <?,?>
<Red,Small>	-	S-set: <Green,Small-Sized> G-set: <Green,?>
<Green,Medium>	-	S-set: <Green,Small-Sized> G-set: <Green,Small-Sized>

Surname: _____

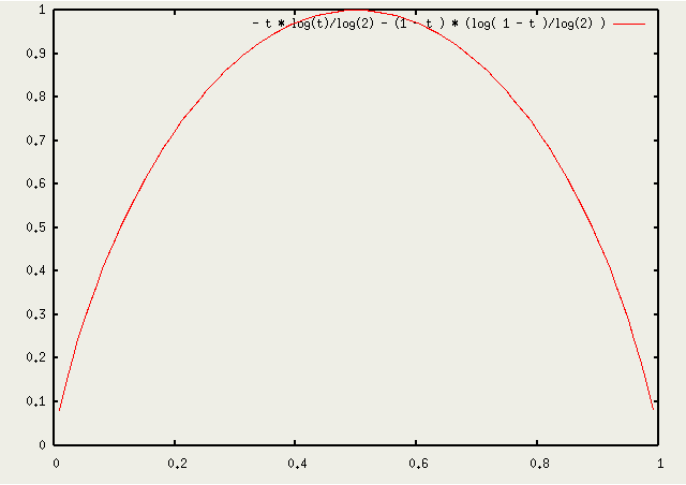
Forename(s): _____

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.



4. Given below is a set of instances from a medical diagnosis domain with two attributes blood pressure and height and whether the person suffered from a disease.

   Given the set of instances shown below, calculate the information gain for the attributes Blood and Height.

Instance	Blood	Height	Disease
x1	Normal	Normal	Yes
x2	High	Tall	No
x3	Normal	Small	Yes
x4	Normal	Tall	No
x5	High	Normal	Yes
x6	Low	Tall	No
x7	Low	Normal	No
x8	High	Small	No
x9	High	Small	No
x10	Low	Small	Yes

CONTINUED

Surname: _____

Forename(s): _____

[10 marks]

Information Gain(Blood) = 0.09

Information Gain(Height) = 0.29

```
All: ent(4,6) = 0.97
Attribute Blood:
All: 0.97 -
Blood = Low: 3/10 * ent( 1, 2 ) = 0.27 -
Blood = Normal: 3/10 * ent( 2, 1 ) = 0.27 -
Blood = High: 4/10 * ent( 1, 3 ) = 0.32
= 0.97 - 0.27 - 0.27 - 0.32 = 0.09

All: 0.97 -
Height = Small: 4/10 * ent( 2, 2) = 0.40
Height = Normal: 3/10 * ent( 2, 1) = 0.27
Height = Tall: 3/10 * ent(0,3) = 0.0
= 0.97-0.4-0.27=0.29
```

5. Given below is an instance set for a manufacturing domain. Attribute plant has the values PA, PB, and PC and attribute model has the values MA, and MB. The attribute failure specifies whether the part failed within the first 3 month.

Instance	Plant	Model	Failure
x1	PA	MA	No
x2	PA	MB	Yes
x3	PB	MB	No
x4	PC	MA	Yes
x5	?	MA	?

Unfortunately, some of the attribute values and classification for instance x5 were lost during transmission. However, from a previous run of the ID3 algorithm you know that the information gains for the attribute Plant and Model over the entire sample set x1..x5 were:

Information Gain(Plant)= 0.57
Information Gain(Model)= 0.20

Is it possible to *uniquely* determine the missing attribute values and classifications given the information above? If so, then show the missing values in the answerbox. If not, then state so in your answer and explain why it is impossible.

[15 marks]

Surname: _____

Forename(s): _____

It is impossible to find an instantiation for x5, since in both cases No or Yes the information gain for the attribute model would only be 0.02.

6. Given below is a set of rules for a medical diagnosis domain. The attribute Blood has values (Normal,High) refers to the blood pressure of the patient, the attribute Stress has values (Low,High) refers to the reported stress level.

   The rules in the rule set are ordered, that is 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 No, because of rule 2 which is evaluated before rule 3.

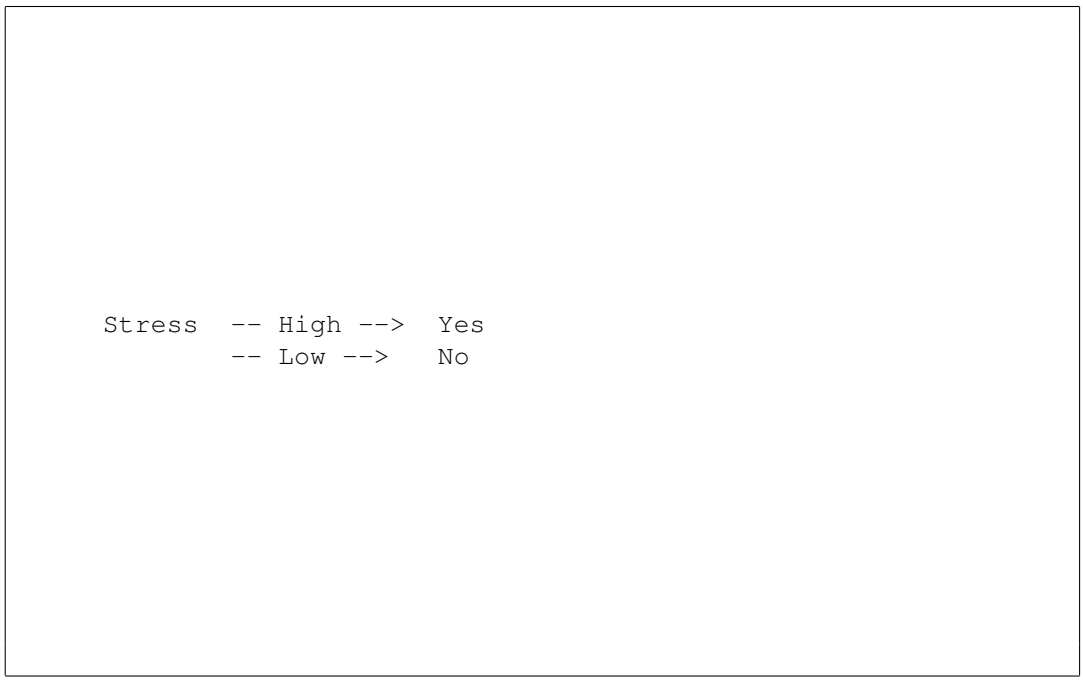
   Is it possible to convert these rules into a decision tree? If so, show the reduced decision tree (i.e., *the decision tree with a minimal number of nodes*) for this problem domain, otherwise explain why these rules can not be converted into a decision tree.

   Rule 1: *if* Blood=Normal and Stress=Low *then* No
   Rule 2: *if* Stress=Low *then* No
   Rule 3: *if* Blood=High *then* Yes
   Rule 4: *if* Stress=Low and Blood=High *then* Yes
   Rule 5: *if* Stress=High *then* Yes

   [10 marks]

Surname: _____

Forename(s): _____



Surname: _____

Forename(s): _____

Section C: Neural Nets

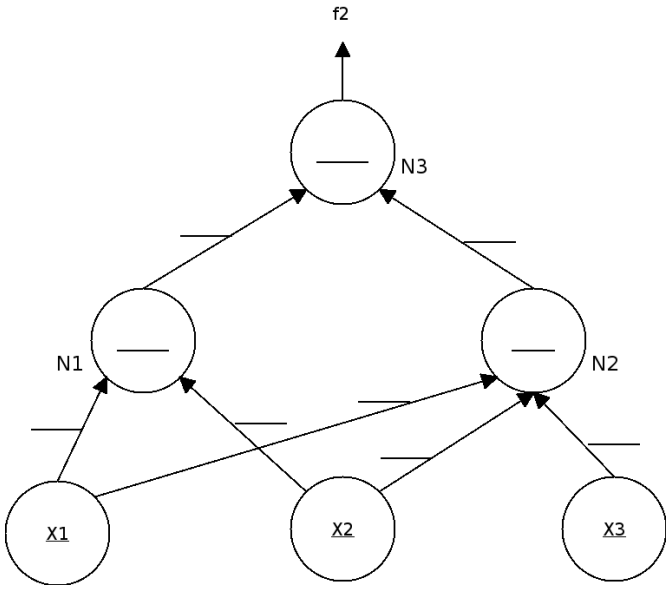
7. Given below is an artificial neural network (ANN) with three input nodes (X1,X2,X3), two hidden nodes, and one output node.

   The network uses simple threshold nodes (i.e., the node will output 1.0 if the sum of the weighted inputs is greater than or equal to 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.

   [18 marks]



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

It is impossible to represent the given function using the network architecture above. Any decomposition leads to a situation where a single network has to implement an XOR-function (non linearly seperable function). Since the first node only has weights from X1 and X2, Node N1 will split the function cube with a plane that is perpendicular to the X1 - X2 plane. It is impossible to seperate the other instances with just a single plane.

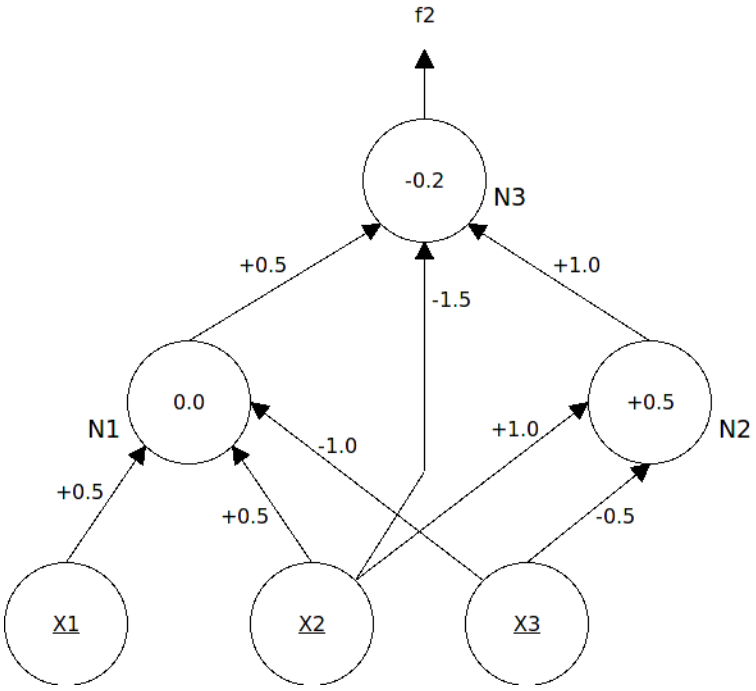
Surname: _____

Forename(s): _____

8. Given below is a neural network with three nodes and three inputs $X1$,$X2$, and $X3$. The network uses threshold neurons, that is a neuron will generate an output of 1, if the sum of the weighted inputs is greater than or equal to the threshold, 0 otherwise.

   Show one set of inputs $X1$, $X2$, and $X3$ that will generate a target output of 0 on the given network. Legal values for Xi are 0 and 1 only.

   If it is impossible to generate the target output with the given network than say so in your answer and explain why.



[15 marks]

To generate an output of 0, the inputs have to be as follows:
$X1$=0

$X2$=1

$X3$=1

Surname: _____

Forename(s): _____

Additional work pages

Surname: _____

Forename(s): _____

Additional work pages