

CSCI/ARTI 8950 Machine Learning

Assignment Number 4: Due Thursday 2/22/2007 (in class)

1. [20 points] Consider the following training set of samples for machine learning:

Example	A1	A2	A3	A4	label
1	1	2	2	2	-
2	1	1	1	1	a
3	2	3	2	1	b
4	1	3	3	3	c
5	3	1	2	1	d
6	1	1	1	2	-

The attributes **A1** through **A4** are integers with values in the range [1,2,3] each.

- (a) For the label assignment (a=-,b=+,c=-,d=+) give a minimal size (measured by the total number of nodes) decision tree that can correctly classify all the training examples.
 - (b) How would the tree given in Part (a) above classify the following examples: (1,2,2,3) and (3,2,1,1)?
 - (c) Propose a label assignment for a, b, c, and d that will make attribute **A4** better than attribute **A3** according to the ID3 information gain measure.
2. [10 points] Solve problem 4.9 on page 125 of the text book.
3. [10 points] Solve problem 5.2 on page 152 of the text book.
4. [10 points] Solve problem 6.1 on page 198 of the text book.