CSCI/ARTI 8950 Machine Learning
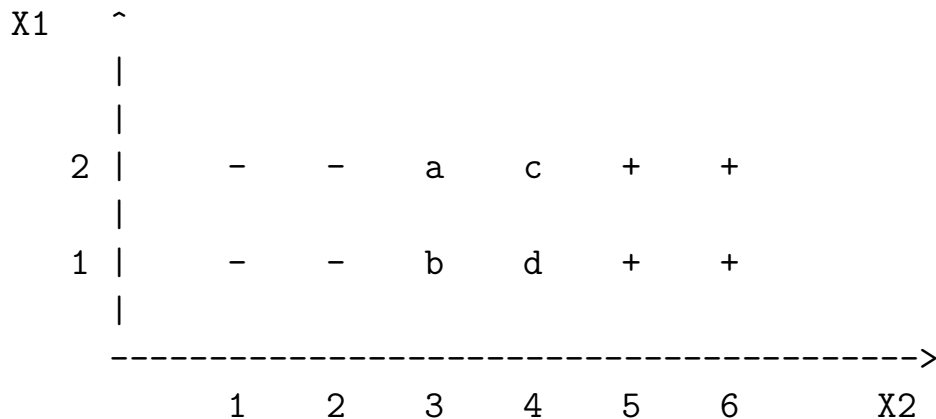
Assignment Number 3: Due 2/25/2016 (in class)

1. [**10 points**] Solve problem 4.9 on page 125 of the text book.

2. [**10 points**][**MID**] 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.

3. [**10 points**][**MID**] Consider the following diagram of a set of samples for machine learning:

```
X1    ^
      |
      |
   2  |    -    -    a    c    +    +
      |
   1  |    -    -    b    d    +    +
      |
      -------------------------------------->
           1    2    3    4    5    6        X2
```

(a) For the label assignment (a=+,b=+,c=-,d=-) can all the given samples (including a, b, c, and d) be correctly classified by a properly trained (or computed in any possible way) Perceptron, whose inputs are X1,X2 and has 2 variable weights and a variable threshold? If your answer is YES, sketch one such Perceptron. If your answer is NO, briefly explain why.

(b) For the label assignment (a=+,b=-,c=+,d=-) can all the given samples (including a, b, c, and d) be correctly classified by a properly trained (or computed in any possible way) binary decision tree with at most 2 levels (2 decisions along each path)? The decision at each level will be of the form $X_i \leq V$ where i is 1 or 2 and V is a variable threshold. If your answer is YES, sketch one such decision tree. If your answer is NO, briefly explain why.

4. **[10 points][MID] Short answers please**

(a) Give the name of one algorithm that searches an incomplete space of hypotheses but searches it completely.

(b) Give the name of one algorithm that searches a complete space of hypotheses but searches it incompletely.

(c) Why should the initial weights of a backpropagation neural network be set to very small values?

(d) Give one advantage to using the information GainRatio measure over the information Gain measure for constructing decision trees.

(e) Give one advantage for using two-fold cross-validation over ten-fold cross-validation.

5. **[10 points][MID]** Suppose that in order to estimate the true error $error_D(h)$ for some hypothesis h, we tested it on 100 instances that were not used for training in any possible way. We observed that h correctly classified 85 of these instances.

(a) Based on this information only, give the most probable value for $error_D(h)$, the standard deviation in this error estimate and 95% two-sided confidence interval. You may leave all of these answers as expressions (no need to use the calculator or do the arithmetic!)

(b) Suppose you are told that when h was tested on the 1000 instances of training data that were used to create it, it correctly classified only 800 of these! Use this additional information to come up with better answers than the ones you gave in part (a) above if you can. (again no need to use the calculator or do the arithmetic!)