

Assignment 1- Chapter3. Decision Tree Learning

Due Date: 94/08/14

Send by Email to: azrabiee@gmail.com

3.1. Give decision trees to represent the following boolean functions:

(a) $A \wedge \neg B$

(b) $A \vee [B \wedge C]$

(c) $A \text{ XOR } B$

(d) $[A \wedge B] \vee [C \wedge D]$

3.2. Consider the following set of training examples:

Instance	Classification	a_1	a_2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

- (a) What is the entropy of this collection of training examples with respect to the target function classification?
- (b) What is the information gain of a_2 relative to these training examples?

3.3. ID3. In this question, you will write the decision tree code and perform experiments with it. You will observe and discuss the overfitting problem. Our data is a binary classification data set with discrete attributes, and we only require the decision tree to be able to process this kind of data.

Data files. All resources are provided in the file hw1.rar. We use a noisy mushroom data set for this problem. Using this data set, we will train decision trees to classify each mushroom as poisonous or not, using 22 discrete features such as cap shape, cap color and gill size. There are three data files in hw1.rar: noisy10 train.ssv, noisy10 valid.ssv, and noisy10 test.ssv. They are training set, validation set (i.e., pruning set), and testing set. The format of each file is: first three lines are data statistics (number of variables plus label, variable names, properties of each variable), and from the 4th line is the data, where each line is an example and each column is either the label (the first column) or a variable.