Homework #6
Algorithms II
Due: Monday November 4th , 2013

1. Consider the set of training data shown in the table below for a binary classification problem (Here + or -). Note that for each record there are three attributes, two of which are binary and one which is continuous.

Record	A	B	C	Class
1	T	T	1.0	+
2	T	T	6.0	+
3	T	F	5.0	-
4	F	F	4.0	+
5	F	T	7.0	-
6	F	T	3.0	-
7	F	F	8.0	-
8	T	F	7.0	+
9	F	T	5.0	-

(a) Determine the entropy of this collection of training examples.

(b) What are the information gains (based on entropy measure) for attributes A and B? Which one provides the largest information gain? To make sure you are counting correctly for each attribute make a table giving the split for the child nodes like

A	+	-
T	3	1
F	1	4

(c) If we split attribute C as ≤ 3, between 3 and 6, and > 6 determine the information gain.

(d) Determine a decision tree based upon choosing the largest information gain using the entropy measure and splitting attribute C as in (c). Justify your choices.