HW8, ECE7252, Spring 2008

1. Exercise 5.1.

2. Assume the following 14 examples in the table below have been seen. Using information theory to measure the entropy, calculate which attribute is best to select for the root node of a decision tree. *For full credit, you must show all of your work.* Feel free to use textbook tables, calculator, etc. to figure out the logarithms base 2. Note that the attributes (features) of the instances are credit history, debt, collateral, and income. The point is to predict an individual's credit risk, abbreviated "risk." Note that in this case there are three classes, instead of the usual two classes "positive" and "negative." (1) You can do the first part of the problem by assuming all three "moderate" risk cases are labeled "high" risk to construct a decision tree; (2) Then in the second part assuming all three "moderate" risk cases are now labeled as "low" risk instead and repeat the construction; (3) Compare the above two sets of results; (4) If you need to consider three classes instead of the binary cases we discuss so far, what would you do to modify the decision tree construction algorithms?

CREDIT HISTORY	DEBT	COLLATERAL	INCOME	RISK
bad	high	none	$0 to $15K	high
unknown	high	none	$15 to $35K	high
unknown	low	none	$15 to $35K	moderate
unknown	low	none	$0 to $15K	high
unknown	low	none	over $35K	low
unknown	low	adequate	over $35K	low
bad	low	none	$0 to $15K	high
bad	low	adequate	over $35K	moderate
good	low	none	over $35K	low
good	high	adequate	over $35K	low
good	high	none	$0 to $15K	high
good	high	none	$15 to $35K	moderate
good	high	none	over $35K	low
bad	high	none	$15 to $35K	high

- This exercise is taken from **Machine Learning, Volume I**, edited by Michalski, Mitchell, and Carbonell, 1983.