COMP 221, Fall 2007: Homework Assignment 3
Machine Learning. Due: Nov 20 in class

Outlook	Tempreature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

Consider the above dataset.

An 1-R rule has the following format: there is one rule for each attribute. For each attribute and each value of an attribute, a rule (Attribute=value)➔X, where X is P or N, is generated if the accuracy Pr(P|Attribute=value) is larger than or equal to Pr(N|Attribute=value).

Then for all values of an attribute, it generates a rule for each of its values. For example, for Outlook, the rules must have a branch for each of “sunny”, “rain” and “overcast” values.

1. For the above dataset, calculate the accuracy of 1-R rules for each attribute on the training data. Which attributes give the best rules?
2. Now split the data into two equal parts. The first part corresponds to the training data, and the second part corresponds to the test data. Repeat the process of building rules based on the training data. Will the best rule still be the best when tested on the test data? Show your calculations.