ENEE 459M: Topics in Computer Engineering – Machine Learning and Data Mining

Spring 2010
Midterm I

1. **[5pt]** What is the difference between a nominal and a numeric attribute?

   [15pt] Consider the following dataset consisting of a single attribute Z and a class attribute Class.

Z	Class
1.0	Yes
6.0	Yes
5.0	No
4.0	Yes
7.0	No
3.0	No
8.0	No
7.0	Yes
5.0	No

   Describe and apply a method to discretize the attribute Z assuming a minimum of 3 in majority class per interval. What are the corresponding intervals?

2. Consider the following dataset with three attributes and a class attribute

X	Y	W	Class
true	true	narrow	Yes
true	true	wide	Yes
true	false	narrow	No
false	false	narrow	Yes
false	true	wide	No
false	true	narrow	No
false	false	wide	No
true	false	wide	Yes
false	true	narrow	No

(a) [20pt] Apply the 1R algorithm to this dataset to generate a set of rules. Show your work step by step.

(b) [25pt] Apply the covering algorithm (PRISM) to derive two classification rules for Class = No. Show your work step by step.

3. [35pt] Consider the following dataset consisting of training examples to learn the attribute Class. Determine the splitting attribute used at the root of the decision tree corresponding to this dataset using the notion of **information gain**. Show step by step the method used to determine the attribute and the basis for selecting the splitting attribute.

A	B	Class
T	F	Yes
T	T	Yes
T	T	Yes
T	F	No
T	T	Yes
F	F	No
F	F	No
F	F	No
T	T	No
T	F	No