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

Spring 2010
Midterm II

1. Consider the following dataset that represents 8 transactions of a grocery store purchases.

Bread	Milk	Cheese	Beer	Eggs	Soda
1	1	0	0	1	0
1	0	1	1	0	0
0	1	1	0	0	1
1	1	0	1	1	0
0	0	1	0	1	1
1	1	0	1	0	1
0	0	0	1	0	1
1	1	0	0	1	0

(a) [15pt] Determine all the 2-itemsets and 3-itemsets with 25% support.
(b) [10pt] For each frequent 3-itemset, determine all the rules with 75% confidence.

2. [20pt] Consider the truth table of a Boolean function $Y=f(X_1,X_2,X_3)$

X_1	X_2	X_3	Y
true	false	false	false
true	false	true	true
true	true	false	true
true	true	true	true
false	false	true	false
false	true	false	false
false	true	true	true
false	false	false	false

Is there a perceptron that can realize this Boolean function, assuming the truth value is represented by +1 and the false value by -1? Show the details of how you would determine

whether there is such a perceptron; Show either the weights on the edges of the perceptron or prove that no such perceptron exists.

3. Consider the following dataset that represents whether a borrower has defaulted or not.

Home Owner	Marital Status	Annual Income	Defaulted Borrower
Yes	Single	High	No
No	Married	High	No
No	Single	Low	No
Yes	Married	High	No
No	Divorced	High	Yes
No	Married	Low	No
Yes	Divorced	High	No
Yes	Single	Low	Yes
No	Married	Low	Yes
No	Single	Low	Yes

(a) [20pt] Derive the naïve Bayesian networks corresponding to this dataset. Explain in details all the statistical information needed to fully define the Bayesian network.

(b) [10pt] Determine the likelihood that a homeowner who is divorced with a low income to be a defaulted borrower.

4. **(a) [10pt]** Given a dataset D with 100 instances, describe precisely how the bootstrap method would extract a training and a test dataset from D. What are the relative sizes of the training and test datasets?

(b) [15pt] After training a classifier on a dataset whose class attribute can take the values A, B, or C, we test the classifier on a test data with 90 instances and obtain the following confusion matrix (columns correspond to predicted class and rows correspond to actual class).

A	B	C
20	5	5
10	15	4
6	7	18

Using the confusion matrix, determine the True Positive Rate, False Positive Rate, Recall, and Precision corresponding to class attribute = B. Show in detail how you derived the values of these parameters.