

Birla Institute of Technology & Science – Pilani
Hyderabad Campus
1st Semester 2010-2011
Data Mining (CS C415) – Test2

Date: 28.10.2010
Duration: 50min.

Weightage: 20%
Type: Closed Book

Please note:

1. Answer **all** questions.
 2. Manage your time judiciously.
 3. Total no. of pages in the question paper: 3
-

1. Answer using only .(TRUE / FALSE)

[10 Marks]

- a) The Decision tree classifier is an eager learner.
 - b) The Rule ordering is done based on the rule accuracy only.
 - c) The Apriori scans the transactional database only once.
 - d) An item set is maximal frequent if none of its immediate supersets is infrequent
 - e) If the correlation measure is 0 then attributes in the either side of association rule are indendent.
 - f) The HUNT'S algorithm is used for growing the rules in rule based classifier.
 - g) The Decision tree uses Pruning techniques after the tree is built.
 - h) The Bayes classifier treats the attributes as independent.
 - i) In the Artificial Neural Network classifier the number of input nodes is equal to the number of classes.
 - j) A linear SVM searches for a hyper plane with largest margin.
-

2. NASA wants to be able to discriminate between Martians (M) and Humans (H) based on the following characteristics: Green {N,Y}, Legs{2,3}, Height{S,T}, Smelly{N,Y}.

Our available training data is as follows:

[20 Marks]

Speices	Green	Legs	Height	Smelly
M	N	3	S	Y
M	Y	2	T	N
M	Y	3	T	N
M	N	2	S	Y
M	Y	3	T	N
H	N	2	T	Y
H	N	2	S	N
H	N	2	T	N
H	Y	2	S	N
H	N	2	T	Y

Draw a decision tree for the class attribute "Species" using the information gain.

3. Suppose you are given the following set of data with three Boolean input variables a; b; and c, and a single Boolean output variable K. Assume we are using a naive Bayes classifier to predict the value of K from the values of the other variables.

[10 marks]

A	B	C	K
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

- According to the naive Bayes classifier, what is the probability of class K=1 when a=1, b=1, and c=0?
- According to the naive Bayes classifier, what is the probability of class K=1 when a=1, b=1, and c is unknown?

4. Consider the data set given in the table given below.

[25 marks]

Customer id	Transaction id	Items
418	234145	{X,Z}
345	543789	{U,V,W,X,Y,Z}
323	965157	{U,W,Y}
418	489651	{V,X,Z}
567	748965	{U,Y}
567	325687	{W,X,Y}
323	127895	{X,Y,Z}
635	617851	{U,Z}
345	824697	{V,Y}
635	102458	{V,W,X}

- Compute the support for item sets {Y}, {X,Z} and {X,Y,Z} by treating each transaction ID as a market basket. [6 marks]
- Use support calculated in(a) to compute the confidence for the rules $XZ \rightarrow Y$ and $Y \rightarrow XZ$. Draw conclusion about confidence as a measure based on the results. (Hint: this is one line answer think about rules $XZ \rightarrow Y$ and $Y \rightarrow XZ$.) [6 marks]
- Repeat part (a) by treating each customer ID as a market basket. Each item should be treated as a binary variable (1 if an item appears in at least one transaction bought by the customer, and 0 otherwise.) [6 marks]
- Use the results in part (c) to compute the confidence for the association rules $XZ \rightarrow Y$ and $Y \rightarrow XZ$. Draw conclusion about confidence as a measure based on the results. (Hint: this is one line answer think about rules $XZ \rightarrow Y$ and $Y \rightarrow XZ$.) [6 marks]

-----GOOD LUCK-----