Indian Institute of Space Science and Technology – Thiruvananthapuram

MA613 Data Mining

Assignment-III

Date: 22-10-2015

1. Consider the following data:

A1	A2	A3	A4	Y
0	F	F	0	1
0	F	T	1	1
0	T	F	1	1
1	T	T	0	1
1	F	F	0	0
1	T	T	1	0
1	T	T	0	0

   (a) Construct a decision tree.

   (b) Frame the rules from the tree.

   (c) Find the output for the following data: $x_1 = (0, F, F, 1)^T, x_2 = (1, F, T, 0)$

   (d) Check whether the tree suffers from overfitting. If so, modify the tree with suitable techniques.

2. A database has five transactions. Let $min_{sup} = 60\%$ and $min_{conf} = 80\%$. Find all

TID	Items bought
TI1	(Milk, Orange, Pomegranate, Lemon, Egg, Bread)
TI2	(Doughnut, Orange, Pomegranate, Lemon, Egg, Bread)
TI3	(Milk, Apple, Lemon, Egg)
TI4	(Milk, Orange Juice, Cheese, Lemon, Bread)
TI5	(Cheese, Orange, Orange, Lemon, Banana, Egg)

   frequent itemsets using Apriori & FP-growth. List all the strong association rules.

3. Download Credit Approval Data Set from UCI website.

(a) Apply k nearest neighbor classifier. Apply bootstrap validation technique.

(b) Using Weka software, apply decision tree.

4. Write short notes on Mahalanobis distance.

Notes

- All the files related with the assignment should be saved in a single folder and send to sumitra@iist.ac.in.

- Last date of submission: 29-10-2015.

- **As far as assignments are concerned, students are expected to observe academic honesty and integrity. Though the students can collaborate and discuss, copying directly other students' assignment or allowing your own assignment to be copied constitute academic dishonesty and is highly discouraged.**