University Mohamed Boudiaf - M'sila
Faculty of MI, CS Department
Master1 AI

May 13th, 2024
Duration : 1h30

Final Exam

Machine Learning & Data Mining

Exercice 1 (4 pts): Given a decision tree, you have the option of (a) converting the decision tree to classification rules and then pruning the resulting rules, or (b) pruning the decision tree and then converting the pruned tree to classification rules. What advantage does (a) have over (b)?

Exercice 2 (3 pts): The following table consists of training data from an employee database. The data have been generalized. For example, "31 : : : 35" for *age* represents the age range of 31to 35. For a given row entry, *count* represents the number of data tuples having the values for *department*, *status*, *age*, and *salary* given in that row.

a- If we want to apply association rules algorithm, what modifications must be made on the dataset?

b- what is the real size of the dataset (number of tuples)?

Department	Status	age	salary	count
Sales	Senior	31...35	46K...50K	3
Sales	Junior	26...30	26K...30K	2
Sales	Junior	31...35	31K...35K	2
Systems	Junior	21...25	46K...50K	2
Systems	Senior	31...35	66K...70K	3
Systems	Junior	26...30	46K...50K	2
Systems	Senior	41...45	66K...70K	3
Marketing	Senior	36...40	46K...50K	1
Marketing	Junior	31...35	41K...45K	2
Secretary	Senior	46...50	36K...40K	3
Secretary	Junior	26...30	26K...30K	2

Exercice 3 (3 pts): With decision trees using Gini Index, we must split numeric attribute to two (02) subsets. Explain how k-means can be used to determine the best point split.

Exercice 4 (10 pts): Consider the following Boolean database with 5 items and 10 transactions**:**

Run the algorithm a priori with Minimal Support=0.3, taking care not to consider the impossible associations in progress algorithm.

Find all the possible rules.

	X1	X2	X3	X4	X5
t1	0	1	0	0	1
t2	0	0	1	0	1
t3	0	0	1	0	0
t4	1	1	1	1	1
t5	1	1	1	1	1
t6	1	1	1	1	0
t7	1	0	1	1	0
t8	1	0	1	1	0
t9	1	0	0	0	1
t10	1	0	0	0	1

Dr. Tahar Mehenni

University Mohamed Boudiaf - M'sila
Faculty of MI, CS Department
Master1 AI

May 13th, 2024
Duration : 1h30

Solution of Final Exam

Machine Learning & Data Mining

Exercice 1 (4 pts):

The pruning is reducing the set of rules as well as the size of the tree. However, reducing the set of rules is more easier than reducing the size of the tree.

Exercice 2 (3 pts):

a- If we want to apply association rules algorithm, each tuple will be duplicated the number of times as the value of the attribute *count*. For example, the first tuple is duplicated with the same values of the attributes *department*, *status*, *age* , *salary* 3 times (*count*=3), as follows:

Department	*Status*	*age*	*salary*	*count*
Sales	Senior	31...35	46K...50K	3

→

Department	*Status*	*age*	*salary*
Sales	Senior	31...35	46K...50K
Sales	Senior	31...35	46K...50K
Sales	Senior	31...35	46K...50K

b- The final size of the dataset (number of tuples) is 25.

Exercice 3 (3 pts):

With decision trees using Gini Index, we must split numeric attribute to two (02) subsets. We can apply a clustering algorithm like k-means to find the best point split. With k-means, we put k=2, and apply the algorithm after sorting the values of the attribute. The split point is computed as follows:

Split point = (last value in the first cluster + first value in the 2^nd^ cluster) / 2

Exercice 4 (10 pts):

1-itemset (1pt)	Freq	Support
X1	7	0.7
X2	4	0.4
X3	7	0.7
X4	5	0.5
X5	6	0.6

2-itemset (1pt)

	X1	X2	X3	X4	X5
X1		3	5	5	4
X2			3	3	3
X3				5	3
X4					2

3-itemset (2 pts)	X1X2	X1X3	X1X4	X1X5	X2X3	X2X4	X2X5	X3X4	X3X5
X1X2		3	3	2				x	x
X1X3			3	2		x	x		
X1X4				2	x		x		x
X1X5					x	x		x	x
X2X3						3	2		
X2X4							2		x
X2X5								x	
X3X4									2

4-itemset (1 pt)

	X1X2X3	X1X2X4	X1X3X4	X2X3X4
X1X2X3		3		
X1X2X4				
X1X3X4				
X2X3X4				

Recap (1 pts)

2-itemset	X1X2	X1X3	X1X4	X1X5	X2X3	X2X4	X2X5	X3X4	X3X5
3-itemset	X1X2X3	X1X2X4	X1X3X4	X2X3X4					
4-itemset	X1X2X3X4								

RULES (4 pts)

X1X2	X1 →X2	X2 → X1
X1X3	X1 → X3	X3 → X1
X1X4	X1 → X4	X4 → X1
X1X5	X1 → X5	X5 → X1
X2X3	X2 → X3	X3 → X2
X2X4	X2 → X4	X4 → X2
X2X5	X2 → X5	X5 → X2
X3X4	X3 → X4	X4 → X3
X3X5	X3 → X5	X5 → X3

X1X2X3	X1 → X2X3	X2X3 → X1
	X2 →X1X3	X1X3 →X2
	X3 → X1X2	X1X2 → X3
X1X2X4	X1 → X2X4	X2X4 → X1
	X2 → X1X4	X1X4 → X2
	X4 → X1X2	X1X2 → X4
X1X3X4	X1 → X3X4	X3X4 → X1
	X3 → X1X4	X1X4 →X3
	X4 → X1X3	X1X3 → X4
X2X3X4	X2 → X3X4	X3X4 → X2
	X3 → X2X4	X2X4 → X3
	X4 → X2X3	X2X3 → X4

X1X2X3X4	X1 → X2X3X4	X2X3X4 → X1
	X2 → X1X3X4	X1X3X4 → X2
	X3 → X1X2X4	X1X2X4 → X3
	X4 → X1X2X3	X1X2X3 → X4
	X1X2 → X3X4	X3X4 → X1X2
	X1X3 → X2X4	X2X4 → X1X3
	X1X4 → X2X3	X2X3 → X1X4

Rédigé par T. Mehenni