Assignment 1(2024)

All Questions are of 1 mark.

1. The earliest step in the data mining process is usually?

a) Visualization
b) Preprocessing
c) Modelling
d) Deployment

Ans: b

Explanation: Preprocessing is the earliest step in data mining.

2. Which of the following is an example of continuous attribute?:

a) Height of a person
b) Name of a person
c) Gender of a person
d) None of the above

Ans: a

Explanation: Height of a person is Real Number.

3. Friendship structure of users in a social networking site can be considered as an example of:

a) Record data
b) Ordered data
c) Graph data
d) None of the above

Ans: c

Explanation: Friendship is an edge in a graph with users as nodes.

4. Name of a person, can be considered as an attribute of type?

a) Nominal
b) Ordinal
c) Interval
d) Ratio

Ans: a

Explanation: Nominal-related to names. The values of a Nominal attribute are name of things, some kind of symbols. There is no order (rank, position) among values of nominal attribute.

5. A store sells 15 items. Maximum possible number of candidate 2-itemsets is:

a) 120
b) 105
c) 150
d) 2

Ans: b

Explanation: Number of ways of choosing 2 items from 15 items is 15C2 = 105

6. If a record data matrix has reduced number of rows after a transformation, the transformation has performed:

a) Data Sampling
b) Dimensionality Reduction
c) Noise Cleaning
d) Discretization

Ans: a

Explanation: Sample is the subset of the population. The process of selecting a sample is known as sampling.

Answer Q7-Q10 based on the following table:

Customer ID	Transaction ID	Items Bought
1	1	{a,d,e}
1	2	{a,b,c,e}
2	3	{a,b,d,e}
2	4	{a,c,d,e}
3	5	{b,c,e}
3	6	{b,d,e}
4	7	{c,d}
4	8	{a,b,c}
5	9	{a,d,e}
5	10	{a,b,e}

7. Taking transaction ID as a market basket, support for each itemset {e}, {b,d}, and {b,d,e} is:

a) 0.8, 0.2, 0.2
b) 0.3, 0.3, 0.4
c) 0.25, 0.25, 0.5
d) 1,0,0

Ans: a

Explanation: support of {e} = 8/10, {b,d} = 2/10, {b,d,e} = 2/10.

8. Based on the results in (7), confidence of association rules {b,d}->{e} and {e}->{b,d} are:

a) 0.5, 0.5
b) 1, 0.25
c) 0.25, 1
d) 0.75, 0.25

Ans: b

Explanation: Confidence(X->Y) = support({X,Y})/support({X}).

Confidence({b,d}->{e}) = support({b,d,e})/support({b,d}) = 0.2/0.2 = 1.

Confidence({e}->{b,d}) = support({b,d,e})/support({e}) = 0.2/0.8 = 0.25.

9. Repeat (7) by taking customer ID as market basket. An item is treated as 1 if it appears in at least one transaction done by the customer, 0 otherwise. Support of itemsets {e}, {b,d}, {b,d,e} are:

a) 0.3, 0.5, 0.2
b) 0.8, 1, 0.2
c) 1, 0.2, 0.8
d) 0.8, 1, 0.8

Ans: d

Explanation: Treating each customer id as a market basket.

Customer ID	Items Bought
1	{a,d,e}, {a,b,c,e}
2	{a,b,d,e}, {a,c,d,e}
3	{b,c,e}, {b,d,e}
4	{c,d}, {a,b,c}
5	{a,d,e}, {a,b,e}

Support({e}) = 4/5 = 0.8

Support({b,d}) = 5/5 = 1

Support({b,d,e}) = 4/5 = 0.8

10. Based on the results in (9), confidence of association rules {b,d}->{e} and {e}->{b,d} are:
a) 0.8, 1
b) 1, 0.8
c) 0.25, 1
d) 1, 0.25

Ans: a

Explanation: Confidence(X->Y) = support({X,Y})/support({X}).

Confidence({b,d}->{e}) = support({b,d,e})/support({b,d}) = 0.8/1 = 0.8.

Confidence({e}->{b,d}) = support({b,d,e})/support({e}) = 0.8/0.8 = 1.