قائمة الاسئلة
التنقيب في البيانات - ( )- المستوى الثالث -قسم نظم المعلومات - - كلية الحاسوب -جامعة صنعاء - الفترة - درجة الامتحان (40)
ا.د/ احمد سلطان الهجامي

1) Which of the following algorithm suits well for application such as customer segmentation, image segmentation and community detection
- 1) + Clustering
- 2) - association
- 3) - classification
- 4) - neural networks

2) Given a rule of the form IF X THEN Y, rule confidence is defined as the conditional probability that
- 1) + Y is true when X is known to be true
- 2) - Xis true when Y is known to be true
- 3) - Y is false when X is known to be false
- 4) - X is false when Y is known to be false.

3) It is the process of finding a model that distinguishes and labels
- 1) - . Data Characterization
- 2) + Data Classification
- 3) - Data discrimination
- 4) - Data Selection

4) Data transformation includes which of the following:
- 1) + A process to change data from a detailed level to a summary level
- 2) - A process to change data from a summary level to a detailed level
- 3) - Joining data from one source into various sources of data
- 4) - Separating data from one source into various sources of data

5) High entropy means that the partitions in classification are pure
- 1) - TRUE.
- 2) + FALSE.

6) In KDD and data mining, noise is referred to as random errors in database
- 1) - TRUE.
- 2) + FALSE.

7) Which statement is true about the K-Means algorithm?
- 1) - All attribute values must be categorical.
- 2) - The output attribute must be cateogrical
- 3) - Attribute values may be either categorical or numeric.
- 4) + All attributes must be numeric.

8) For questions given below, consider the data Transactions:
   T1: I10, I2, I4, I11, I5
   T2: I1, I9, I10, I4, I6
   T3: I1, I8, I4, I5
   T4: I7, I2, I3, I4, I5, I6
   T5: I1, I2, I3, I4, I5, I6 With support as 0.6 find all frequent itemsets?
- 1) - {i2},{i4},{i5},{i2,i4}, {i2,i5}, {i4,i5}, {i2,i4,i5}
- 2) - {i1},{i4},{i5},{i6}, {i1,i4}, {i5,i4}, {i1,i5}, {i4,i6}, {i2,i4,i5}
- 3) + {i1},{i2},{i4},{i5},{i6}, {i1,i4}, {i2,i4}, {i2,i5}, {i4,i5}, {i4,i6}, {i2,i4,i5}
- 4) - None of above

9) What is the effect of reducing min confidence criteria?
- 1) - Number of association rules remains same
- 2) + Some association rules will be added to the current set of association rules

3)   -   Some association rules will become invalid while others might become a rule.

4)   -   None of above

10)   Association rule mining is an unsupervised learning technique.

1)   **+**   TRUE.

2)   -   FALSE.

11)   Data warehouses support OLTP technology for many business intelligence

1)   -   TRUE.

2)   **+**   FALSE.

12)   This step of the KDD process model deals with noisy data.

1)   -   Creating a target dataset

2)   **+**   data preprocessing

3)   -   data transformation

4)   -   data mining

13)   The primary objective of Association Rule Mining is:

1)   -   To cluster data into groups

2)   **+**   To find hidden patterns or relationships between variables in a dataset

3)   -   To classify data into predefined categories

4)   -   To reduce the dimensionality of data

14)   Business intelligence (BI) can be characterized as a transformation of

1)   **+**   data to information to decisions to actions

2)   -   Big Data to data to information to decisions

3)   -   data to processing to information to actions

4)   -   actions to decisions to feedback to information

15)   What is Information Gain?

1)   -   It is a type of index structure

2)   **+**   It is a measure of purity

3)   -   Both options of A & B

4)   -   None of the above

16)   How do you calculate Confidence(A → B)?

1)   -   Support(A ∧ B) / Support (A)

2)   -   Support(A ∧ B) / Support (B)

3)   **+**   Support(A ∨ B) / Support (A)

4)   -   Support(A ∨ B) / Support (B)

17)   A good introduction to machine learning is the idea of _____.

1)   **+**   . concept learning.

2)   -   content learning.

3)   -   theory of falsification

4)   -   nothing from the above

18)   The _____ data are stored in data warehouse.

1)   -   . operational.

2)   **+**   historical.

3)   -   transactional.

4)   -   optimized

19)   Which of the following is not a factor in data quality:

1)   -   Accuracy

2)   -   Completeness

3)   **+**   Relevance

4)   -   Timeliness

20)   Suppose that a data mining task is to cluster the following 6 points with (x, y) representation into 2 clusters:

A1(4,6), A2(2, 5), A3(9, 3), A4(6, 9), A5(7, 5), A6(5, 7).

Consider A1 and A2 as the initial centroids of the 2 clusters. Use the K-means method to show the after the first iteration?

1)   -   2,5
2)   +   6,6.2
3)   -   6.2,6
4)   -   none of the above

21)   _Suppose that a data mining task is to cluster the following 6 points with (x, y) representation into 2 clusters: A1(4,6), A2(2, 5), A3(9, 3), A4(6, 9), A5(7, 5), A6(5, 7).

Consider A1 and A2 as the initial centroids of the 2 clusters. Use the K-means method to show the centroid value of the second cluster C2 after the first iteration?

1)   -   3,6
2)   +   2,5
3)   -   4,8
4)   -   None of the above

22)   __Suppose that a data mining task is to cluster the following 6 points with (x, y) representation into 2 clusters: A1(4,6), A2(2, 5), A3(9, 3), A4(6, 9), A5(7, 5), A6(5, 7).

Consider A1 and A2 as the initial centroids of the 2 clusters. Use the K-means method to show the points of cluster C1:

1)   -   A3,A4,A5,A6
2)   -   A2
3)   +   A1,A3,A4,A5,A6

23)   True or False: A high support value guarantees a strong association rule.

1)   -   TRUE.
2)   +   FALSE.

24)   True or False: Objective measures are based on stucture of patterns

1)   +   TRUE.
2)   -   FALSE.

25)   True or False: Subjective measures are based user who examine the patterns

1)   +   TRUE.
2)   -   FALSE.

26)   True or False: Objective measures are data dependent

1)   +   TRUE.
2)   -   FALSE.