يوم الامتحان: الاربعاء

تاريخ الامتحان: 22 / 5 / 2019 م

المادة : تنقيب بيانات (464 رس)

الممتحن: د/ مصعب عبد الحميد محمد حسان

مدرس بقسم الرياضيات بكلية العلوم

الاسئلة و نموذج الإجابة

ورقة كاملة

Benha University
Faculty of Science
Dept. of Mathematics

Time: Two Hours
Second Semester 2018-2019
Date : 22/5/2019

Data Mining (MC464) for fourth Level Students (Computer Science)

Choose the correct answer for each of the following:[24 x 1.5+ 3 x 4 = 48 Marks]

1- Knowing all maximal patterns allows us to …..
 (A) reconstruct the set of all frequent patterns with their support counts.
 (B) reconstruct the set of all frequent patterns without their support counts.

2- In Eclat algorithm, determine support of any k-itemset by intersecting tid-lists of two of its k subsets
 (A) Yes   (B) No

3- The closed patterns are ….. of maximal patterns.   (A) superset   (B) subset

4- In Eclat algorithm, the Advantage is ……
 (A) tid-lists become too small for memory
 (B) very fast support counting

5- Suppose that a transaction database has only three transactions:

 $\{\{a1, a2, a3\}, \{ a1, a2\}, \{a1\}\}$ and let min_sup = 1. There are …. closed itemsets
 (A) two   (B) three   (C) four

6- Apriori property is ….
 (A) any subset of frequent itemset must be frequent.
 (B) any subset of maximal itemset must be frequent.
 (C) any subset of closed itemset must be frequent.

7- In generating association rules, we generate all nonempty subsets of each frequent itemset L and for every nonempty subset s of L, we output the rule ….
 (A) "s $\rightarrow$ (L - s)" that satisfy min_conf   (B) "s $\rightarrow$ (L - s)" that satisfy min_sup
 (C) "L $\rightarrow$ (L - s)" that satisfy min_conf   (D) "L $\rightarrow$ (L - s)" that satisfy min_sup

8- FP-growth is an order of magnitude faster than Apriori.
 (A) Yes   (B) No

9- In construction of FP-Tree, a branch is created for each transaction with items having their support count separated by colon. Whenever the same node is encountered in another transaction, we just increment the support count of ….
 (A) the common node or prefix   (B) the common node or suffix

10- In construction of FP-Tree, to facilitate tree traversal, an item header table is built so that each item points to ….
 (A) its occurrences in the tree via a chain of node-links
 (B) its prefix in the tree via a chain of node-links

11- In FP-Growth algorithm, the pattern growth is achieved by concatenation of …. with the frequent patterns generated from a conditional FP-Tree.
 (A) prefix pattern   (B) suffix pattern

12- In FP-Growth algorithm, construct its ….
 (A) conditional pattern base then its conditional FP-Tree.
 (B) conditional FP-Tree then conditional pattern base.

13- In classification, the model is represented as ….
 (A) classification rules   (B) decision trees
 (C) mathematical formulae   (D) all of them

14- In classification, the model is used for classifying …..
 (A) unknown samples.   (B) known samples.   (C) all samples

15- In classification, test set is …. of training set
 (A) independent   (B) dependent

16- 14- Accuracy rate is ….. that are correctly classified by the model
 (A) the percentage of training set samples   (B) the percentage of all samples

(C) the percentage of test set samples

17- In decision tree, ….. represents class labels
   (A) root   (B) internal node   (C) leaf node
18- In decision tree, information gained by branching on attribute A is equals to ...
   (A) Gain (A) = $Info(D) - Info_A(D)$ (B) Gain (A) = $Info_A(D) - Info(D)$
19- In KNN algorithm, K is the number of classes.   (A) Yes   (B) No
20- Which of the following is correct itemset?
   (A) aabc   (B) abc   (C) abcc
21- To measure the quality of the clustering process we use ….
   (A) the sum of the squared error   (B) the product of the squared error
22- Conditions for stopping partitioning in decision tree:
   (A) All samples for a given node belong to the same class.
   (B) There are no remaining attributes.   (C) There are no samples left.
   (D) all of them
23- ….. partitions the dataset into subsets or groups such that elements of a group
   share a common set of properties
   (A) association rules   (B) classification   (C) clustering
24- In KNN algorithm, k …. of the number of classes
   (A) must be a multiple   (B) must not be a multiple

25- The following is a set of two-dimensional points: {(1, 1), (2, 1), (4, 3), (5, 4)}. Use
   K-means with k = 2 and {(1, 1), (2, 1)} is the set of initial centroids then G^1
   matrix equals …..
   (A) 1   1   0   0   (B) 0   1   1   0
        0   0   1   1   1   0   0   1

26- The following is the pseudo-code of generating decision tree
   GenDecTree(Sample S, Attlist A)
   1. create a node N
   2. If all samples are of the same class C then label N with C; terminate;
   3. If A is empty then label N with the most common class C in S (majority
      voting); terminate;
   4. Select a ∈ A, with the highest information gain; Label N with a;
   5. For each value v of a:
       a. Grow a branch from N with condition a=v;
       b. Let S_v be the subset of samples in S with a=v;
       c. If S_v is empty then attach a leaf labeled with the most common class in S;
       d. Else attach the node generated by GenDecTree(S_v, A)
   (A) Yes   (B) No

27- The following is the pseudo-code of Apriori algorithm
   C_k: Candidate itemset of size k
   L_k : frequent itemset of size k
   L_1 = {frequent items};
   for (k = 1; L_k !=∅; k++) do begin
     for each transaction t in database do
        increment the count of all candidates in C_{k+1} that are contained in t
        L_{k+1} = candidates in C_{k+1} with min_support
      end
   end
   return ∪$_k$ L_k;
   (A) Yes   (B) No

Best Wishes

Model Answer

1- B
2- B
3- A
4- B
5- B
6- A
7- A
8- A
9- A
10- A
11- B
12- A
13- D
14- A
15- A
16- C
17- C
18- A
19- B
20- B
21- A
22- D
23- C
24- B
25- A
26- B
27- B