KUMARAGURU college of technology
character is life

Register Number:……..…

MCA DEGREE EXAMINATIONS: DEC 2015

(Regulation 2014)

Third Semester

MASTER OF COMPUTER APPLICATIONS

P14CAT301: Data warehousing And Data Mining

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. Given the following activities of knowledge discovery in databases. Arrange them in the correct order of execution. [K_2]

   i) Data integration ii) Data cleaning iii) Data transformation
   iv) Data selection v) Pattern evaluation vi) Data mining

   a) i – iv – iii – v – ii- vi
   b) ii – iii - i - v – vi - iv
   c) ii – iv – iii – vi – i - v
   d) ii – i – iv – iii – vi - v

2. A cell in the base cuboid is called a _____ cell and from a nonbase cuboid is called a _____ cell. [K_2]

   a) base, nonbase
   b) apex, aggregate
   c) base, aggregate
   d) apex, descendant

3. FP growth algorithm adopts _____ strategy. [K_2]

   a) divide & conquer
   b) brute force
   c) backtracking
   d) branch & bound

4. Match the items in List I with that of List II [K_3]

List I	List II
A. Data integration	i. Binning
B. Data reduction	ii. Normalization
C. Data transformation	iii. Principal Component Analysis
D. Data cleaning	iv. Correlation

   a) A – i , B – iii , C – ii , D -iv
   b) A – iv , B – ii , C – iii , D - i
   c) A – iv , B – iii , C – ii , D -i
   d) A – iii , B – i , C – ii , D -iv

5. If the class label for each training tuple is not known, then it is termed as _____ learning. [K_2]

   a) supervised
   b) unsupervised
   c) predictive
   d) classified

6. The time complexity of 'k' means algorithm with 'n' objects, 'k' clusters and 't' iterations is _____ . [K_2]

a) O(nkt) b) O(nk)

c) O(nk/t) d) O(t)

7. Assertion (A): Outliers are different from noisy data. [K_4]

Reason (R): Noise is a random error or variance.

a) Both A and R are individually true and R is the correct explanation of A. b) Both A and R are individually true but R is not the correct explanation of A.

c) A is true but R is false. d) A is false but R is true.

8. A star schema has ____ type of relationship between a dimension and fact table. [K_2]

a) many –to-many b) one-to-one

c) many-to-one d) one-to-many

9. Which of the following are the types of numerical attributes? [K_3]

i) Interval scaled
ii) Ordinal
iii) Ratio-scaled

a) i) and ii) only b) i) and iii) only

c) ii) and iii) only d) i) , ii) and iii) only

10. Web _____ mining is the process of extracting useful information from server logs. [K_2]

a) content b) structure

c) data d) usage

PART B (10 x 2 = 20 Marks)

11. Use smoothing by bin means to partition the following data into three bins: [K_3]

4, 15, 21, 21, 24, 25, 28, 34

12. State the need for OLAP. [K_2]

13. What is the relationship between nominal and binary attributes? [K_4]

14. A cube has 10 dimensions and each dimension has 5 levels. How many total cuboids can be generated? [K_3]

15. State the Apriori property. [K_2]

16. Differentiate between a data mart and a data warehouse. [K_4]

17. Write any two applications of data mining. [K_2]

18. How do you choose the best split while constructing a decision tree? [K_4]

19. Why are concept hierarchies useful in data mining? [K_5]

20. A customer has one more cup of coffee than usual in a day. Is it an outlier or not? Justify. [K_5]

PART C (6 x 5 = 30 Marks)

21. Given the data 200,300,400,600, 1000 . Normalise it using [K_3]
    i) Minmax normalization
    ii) Z score normalization
    iii) Normalization by decimal scaling

22. What are the different types of OLAP operations that can be performed on a data cube? Give examples. [K_2]

23. Explain the different types of attributes with examples. [K_2]

24. Describe the different functionalities that can be performed using data mining. [K_3]

25. Write the 'K'-medoids algorithm. [K_2]

26. Suppose a market shopping data warehouse consists of four dimensions: customer, date, product, and store, and two measures: count, and avg sales, where 'avg sales' stores the real sales in rupees at the lowest level and 'count' stores the number of items sold. Draw a snowflake schema diagram. [K_5]

PART D (4 x 10 = 40 Marks)

27. Explain the architecture of data mining with a neat diagram. [K_2]

28. Given the data about a fictitious marketing strategy where a company sent out some promotion to various houses and recorded a few facts about each house and also whether the people responded or not. Construct a decision tree for the following data and generate the rules to find out which type of people responded to the promotion scheme. [K_5]

District	**House Type**	**Income**	**Previous Customer**	**Outcome**
Suburban	Detached	High	No	Nothing
Suburban	Detached	High	Yes	Nothing
Rural	Detached	High	No	Responded
Urban	Semi-detached	High	No	Responded
Urban	Semi-detached	Low	No	Responded
Urban	Semi-detached	Low	Yes	Nothing
Rural	Semi-detached	Low	Yes	Responded

Suburban	Terrace	High	No	Nothing
Suburban	Semi-detached	Low	No	Responded
Urban	Terrace	Low	No	Responded
Suburban	Terrace	Low	Yes	Responded
Rural	Terrace	High	Yes	Responded
Rural	Detached	Low	No	Responded
Urban	Terrace	High	Yes	Nothing

29. How is data mining carried out in WWW? [K_4]

30. Apply Apriori algorithm on the following data set and generate strong association rules (with 30% support and 40% confidence). [K_3]

Let minimum support level = 4.

TID	Items
1	{1,2,3,4}
2	{1,2,3,4,5}
3	{2,3,4}
4	{2,3,5}
5	{1,2,4}
6	{1,3,4}
7	{2,3,4,5}
8	{1,3,4,5}
9	{3,4,5}
10	{1,2,3,5}

****