STELLA MARIS COLLEGE (AUTONOMOUS) CHENNAI –600 086
(For candidates admitted from the academic year 2011 – 12 & thereafter)

M. Sc. DEGREE EXAMINATION, APRIL 2015
BIOINFORMATICS
FOURTH SEMESTER

COURSE   : CORE
PAPER   : DATA MINING AND MACHINE LEARNING
TIME   : 3 HOURS   MAX. MARKS: 100

SECTION – A

ANSWER ALL QUESTIONS   (20 X 1=20)
I. CHOOSE THE RIGHT ANSWER

1. _____ is frequently referred to as k-means clustering.
    a. Non-hierarchical clustering
    b. Optimizing partitioning
    c. Divisive clustering
    d. Agglomerative clustering

2. Which of the following is not found in Genetic Algorithms?
        a . evolution   b. selection   c. reproduction   d. mutation

3. A computer based information system is needed because
        (i) The size of organization have become large and data is massive
        (ii) Timely decisions are to be taken based on available data
        (iii) Computers are available
        (iv) Difficult to get clerks to process data
        a (ii) and (iii)   b (i) and (ii)   c (i) and (iv)   d (iii) and (iv)

4. In apriori algorithm, the frequent item sets can be evaluated by _____ and _____
        i) support and conincidence
        ii) support and confidence
        iii) confidence and no.of items
        iv) no. of items and support

5. SOM stands for _____ _____ _____
        a. Smart operating Machines b. Self organizing maps
        c. Self operating machines   d. Smart operation module

6. The hierarchical clustering method is well applied to solve_____
        a. DNA replication   b. Gene expression
        c. DNA translation   d. gene function

7. _____ techniques can be used to reduce the number of values for a given continuous
    attribute, by dividing the range of the attribute into intervals.
        a. Discretization   b. Transformation   c. Smoothing d. none of the above

8. Where are Genetic Algorithms applicable?
        a. real time application   b. biology   c. Artificial Life   d. economics

9. Vertical lines represent clusters that are joined together. Such tree graph is a graphical cluster is called
        a. Dendrogram   b. scatter gram   c. scree plot   d. icicle diagram

10. Data mining concept operated by
        a .large quantities of operational data stored over a period of time   b. lots of tactical data
        c. several tape drives to store archival data   d .large mainframe computers

11. A genetic algorithm is a neural network that mimics the evolutionary, survival-of-the-fittest process to generate increasingly better solutions to a problem.
                        a. True   b. false

12. Data management and model management are terms that can be used interchangeably.
        a. True   b. false

13. Which AI system can be "trained" to recognize patterns?
        a. Expert systems   b. Neural networks   c. Genetic algorithms   d. Intelligent agents

14. Which AI system generates increasingly better solutions to problems by generating many solutions?
        a. Expert systems   b. Neural networks
        c. Genetic algorithms   d. intelligent agents

15. What is the genetic algorithm concept of evolution that refers to the survival of the fittest?
        a. Selection   b. Crossover   c. Mutation   d. None of the above

16. Which of the following is a disadvantage of a neural network?
        a. Learn and adjust to new circumstances on their own
        b. Lend themselves to massive parallel processing
        c. Function with only complete or well-structured information
        d. Cope with huge volumes of information

17. What is the biggest problem with neural networks?
        a. The way the input layer functions   b. The way the output layer functions
        c. The way the hidden layer functions   d. All of the above

18. What are the three layers of a neural network?
        a. Input, output, and decision   b. Input, output, and hidden
        c. Input, output, and highlighted   d. None of the above

19. An _____ mimics the structure and function of the human brain.
        a. Expert systems   b. Neural networks   c. Genetic algorithms   d. Intelligent agents

20. An _____ mimics the evolutionary survival-of-the-fittest process to generate increasingly better solutions to a problem.
        a. Expert systems   b. Neural networks   c. Genetic algorithms   d. Intelligent agents

SECTION – B

ANSWER ANY FOUR QUESTIONS. EACH ANSWER SHOULD NOT EXCEED 500 WORDS. All ANSWERS CARRY EQUAL MARKS. DRAW DIAGRAMS WHEREVER NECESSARY   (4 X 10 = 40)

21. Write the salient features of density based clustering method.

22. Explain the functions of SOM techniques in data mining.

23. Define Apriori algorithm.

24. What is machine learning process?

25. Describe the methodology used for Genetic algorithm.

26. Explain about back propagation neural network and its application.

27. Briefly describe the major issues take place on data mining.

SECTION – C

ANSWER ANY TWO QUESTIONS.   EACH ANSWER SHOULD NOT EXCEED 1200 WORDS. All ANSWERS CARRY EQUAL MARKS. DRAW DIAGRAMS WHEREVER NECESSARY   (2 X 20 = 40)

28. Describe the cross over technique and its application on bioinformatics.

29. Write brief account on data integration, transformation and reduction.

30. Explain in details about the various types of clustering methods.

31. Describe the methods of evaluation in the neural network model with one example.

✱✱✱✱✱✱✱✱✱✱