

**M. Sc. DEGREE EXAMINATION, APRIL 2014**  
**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. The analysis of data objects without consulting a known class label is called
  - a. characterization
  - b. clustering
  - c. evolution
  - d. tree
  
2. \_\_\_\_\_ methods smooth a sorted data value by consulting its “neighborhood”, that is, the values around it.
  - a. Binning
  - b. Clustering
  - c. Combined computer and human inspection
  - d. Regression
  
3. Data Mining is also referred to as \_\_\_\_\_ .
  - a. Knowledge discovery in databases
  - b. Data Cleaning
  - c. Data extraction
  - d. Data management
  
4. Volume of operational information is
  - a. Condensed
  - b. Detailed
  - c. Summarized
  - d. Irrelevant
  
5. Data mining requires
  - 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

6. Which of the following(s) is/are found in Genetic Algorithms?  
(i) evolution (ii) selection (iii) reproduction (iv) mutation
- a. i & ii only
  - b. i, ii & iii only
  - c. ii, iii & iv only
  - d. all of the above
7. Which of the following statement(s) is/are true?
- (i) Genetic Algorithm is a randomised parallel search algorithm, based on the principles of natural selection, the process of evolution.
  - (ii) GAs are exhaustive, giving out all the optimal solutions to a given problem.
  - (iii) GAs are used for solving optimization problems and modeling evolutionary phenomena in the natural world.
  - (iv) Despite their utility, GAs remain a poorly understood topic.
- a. i, ii & iii only
  - b. ii, iii & iv only
  - c. i, iii & iv only
  - d. all of the above
8. If crossover between chromosomes in search space does not produce significantly different offspring, what does it imply? (if offspring consist of one half of each parent)
- (i) The crossover operation is not successful.
  - (ii) Solution is about to be reached.
  - (iii) Diversity is so poor that the parents involved in the crossover operation are similar.
  - (iv) The search space of the problem is not ideal for GAs to operate.
- a. ii, iii & iv only
  - b. ii & iii only
  - c. i, iii & iv only
  - d. all of the above

Matching between terminologies of Genetic Algorithms and Genetics:

**Genetic Algorithms**

**Genetics (biology)**

- |                               |  |
|-------------------------------|--|
| (a) representation structures | (i) external disturbance, such as cosmic radiation |
| (b) crossover                 | (ii) chromosomes                                   |
| (c) mutation                  | (iii) survivability                                |
| (d) selection                 | (iv) sexual reproduction                           |

9. a. \_\_\_\_\_
10. b. \_\_\_\_\_
11. c. \_\_\_\_\_
12. d. \_\_\_\_\_

13. Which statement is not true about cluster analysis?
- a. Objects in each cluster tend to be similar to each other and dissimilar to objects in the other clusters.
  - b. Cluster analysis is also called classification analysis or numerical taxonomy.
  - c. Groups or clusters are suggested by the data, not defined a priori.
  - d. Cluster analysis is a technique for analyzing data when the criterion or dependent variable is categorical and the independent variables are interval in nature.
14. A \_\_\_\_\_ or tree graph is a graphical device for displaying clustering results. Vertical lines represent clusters that are joined together. The position of the line on the scale indicates the distances at which clusters were joined.
- a. Dendrogram
  - b. scatter gram
  - c. scree plot
  - d. icicle diagram
15. \_\_\_\_\_ is frequently referred to as k-means clustering.
- a. Non-hierarchical clustering
  - b. Optimizing partitioning
  - c. Divisive clustering
  - d. Agglomerative clustering
16. In cluster analysis, objects with larger distances between them are more similar to each other than are those at smaller distances.
- (True or False)
17. Which of the following is true?  
Single layer associative neural networks do not have the ability to:
- (i) perform pattern recognition
  - (ii) find the parity of a picture
  - (iii) determine whether two or more shapes in a picture are connected or not
- a. (ii) and (iii) are true
  - b. (ii) is true
  - c. all of them are true
  - d. all of them are false
18. \_\_\_\_\_ algorithms are trained on labeled examples, i.e., input where the desired output is known.
- a. Supervised learning
  - b. Unsupervised learning
  - c. Semi-supervised learning
  - d. Reinforcement learning
19. Which of the following statement(s) is/are true?  
Machine learning focuses on prediction, based on known properties learned from the training data.
- (True or False)
20. Data mining focuses on the discovery of (previously) unknown properties in the data.
- (True or False)

**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. Describe challenges to data mining regarding data mining methodology and user interaction issues.
22. Write advantages and disadvantage of machine learning uses.
23. Briefly explain the role of density based clustering method on bioinformatic.s
24. Define role of SOM techniques in data mining.
25. Explain in detail about the Apriori algorithm for data processing.
26. Explain the role of crossing over techniques for fitness function.
27. Write short notes on transactional databases.

**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 various types of data mining systems are used in the area of Bioinformatics.
29. Describe the KDD process in data mining with a neat diagram.
30. Write briefly about the features, anatomy and methodology of Genetic algorithm.
31. Describe the back propagation classification and association based classification.

\*\*\*\*\*