

APRIL 2024

51343/SZ26B

Time : Three hours :

Maximum : 75 marks

PART A — (10 × 2 = 20 marks)

Answer any TEN questions.

1. What are the advantages of time series analysis?
2. Write a note on prediction.
3. What is sequence discovery?
4. Define: "Hypothesis Testing".
5. Mention the advantages of neural networks.
6. Show the example of activation functions in neural networks.
7. What are the uses of ID3?
8. Give the purpose of C4.5 in decision tree.
9. State the clustering with genetic algorithms.
10. What do you mean by divisive clustering?
11. Define: "Data Parallelism".
12. What are generalized association rules?

PART B — (5 × 5 = 25 marks)

Answer any FIVE questions.

13. Summarize the major issues of data mining.
14. What are the data mining from a database perspectives? Explain.
15. Distinguish between the regression and correlation.
16. Describe the regression in statistical-based algorithms.
17. Elaborate the simple approach of distance-based algorithms.
18. Bring out the minimum spanning tree in partitional algorithms.
19. Write down the measuring the quality of rules in association rules.

PART C — (3 × 10 = 30 marks)

Answer any THREE questions.

20. Outline the basic data mining tasks in detail.
21. Discuss the Bayes theorem in statistical perspective on data mining.

22. Illustrate the K nearest neighbors in distance-based algorithms.
23. Demonstrate the PAM algorithm in partitional algorithms.
24. Examine the large itemsets in association rules.