



LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034

M.Sc. DEGREE EXAMINATION – COMPUTER SCIENCE

FIRST SEMESTER – NOVEMBER 2015

CS 1822 - DATA MINING

Date : 03/11/2015

Dept. No.

Max. : 100 Marks

Time : 01:00-04:00

SECTION-A

ANSWER ALL THE QUESTIONS:

(10 x 2 = 20)

1. What is root mean square error?
2. Define Euclidian distance.
3. Define Classification.
4. What is back propagation technique?
5. What is clustering?
6. What is squared error for clustering?
7. What is fast update?
8. What are the rules applied to reduce minimum support?
9. What are crawlers?
10. What are HITS? Specify its components.

SECTION-B

ANSWER ALL THE QUESTIONS:

(5 x 8 =40)

11. a) Compare data mining with knowledge discovery in databases.
(OR)
b) Explain Baye's theorem. Illustrate with an example.
12. a) Describe the issues in classification.
(OR)
b) Explain CART with an example.
13. a) Explain PAM algorithm with an example
(OR)
b) Explain agglomerative clustering algorithm with an example
14. a) What are large Item-sets? Specify the algorithm to find *support* and *confidence*.
(OR)
b) Explain quantitative association rules and the algorithm to find the same.
15. a) Explain personalization with web content mining
(OR)
b) Discuss about spatial queries.

SECTION-C

ANSWER ANY TWO QUESTIONS:

(2 x 20 = 40)

16. i) Describe the basic data mining tasks with categorization.
ii) Explain the role of neural networks in solving classification problems.
Mention the issues, advantages and disadvantages.
17. i) Explain K-Means algorithm and explain how it is applied to the following datasets where the number of clusters are 2.

Items	Observations	
	X1	X2
A	5	3
B	-1	1
C	1	-2
D	-3	-2

- ii) Explain data parallelism and task parallelism with relevant algorithms
18. i) Discuss about Web usage mining.
ii) Explain Neural network models and its activation functions.
