Total No. of Questions :8]

SEAT No. :

P2283

[Total No. of Pages :4

[5333] - 2003

M.Sc.

COMPUTER SCIENCE

CS 203 : Data Mining and Warehousing

(2013 Pattern) (Semester - II)

Time : 3 Hours] ***[Max. Marks :50***

Instructions to the candidates:

1) ***Answer any Five questions.***
2) ***Figures to the right indicate full marks.***

Q1) Solve

a) What are the social implication of data mining? **[4]**

b) Explain OLAP. **[4]**

c) Define : Precision and Recall. **[2]**

Q2) Solve

a) Explain overfitting with example. **[4]**

b) What are the challenges in web mining? **[4]**

c) Why data processing is required? **[2]**

Q3) Solve

a) Explain data warehouse architecture with the help of neat diagram. **[4]**

b) Construct an FP-Tree on the following data **[4]**

P.T.O.

TID	Item
1	E, A, D,B
2	D, A, C, E, B
3	C, A, B, E
4	B, A, D
5	D
6	D, B
7	A, D, E
8	B, C

c) What is chi square Test? **[2]**

Q4) Solve

a) Generate Frequent item sets using Apriony For the Following transactions with minimum support = 3 **[4]**

Transaction ID	Items
T 10	M, O, N, K, E, Y
T20	D, O, N, K, E, Y
T30	M, A, K, E
T40	C, A, K, E
T50	C, O, K, E
T60	D, A, Y
T70	B, R, E, A, D

b) Explain linear regression in detail. **[4]**

c) What is Boot strap? **[2]**

Q5) Solve

a) Write short note on WEKA. **[4]**

b) Explain Data integration and data transformation. **[4]**

c) What is pattern discovery in web data mining. **[2]**

Q6) Solve

a) Explain F-measure and confusion matrix. **[4]**

b) Explain frequent subgraph mining. **[4]**

c) List methods to handle the misssing values. **[2]**

Q7) Solve

a) Suppose that the data mining task is to cluster the following eight points (with (x, y) representing location into three clusters : A1(2, 10), A2 (2,5), A3(8,4), B1(5, 8), B2(7, 5), B3(6,4), C_1(1, 2), C_2(4, 9). The distance Function is Eaclidean distance. Suppose, Initially we assign A1, B1, and C1 as center of each cluster. Apply K-means Algorithm. **[5]**

b) Write note on Text mining. **[5]**

Q8) Solve

a) Consider the data from employee database. **[5]**

Department	Status	Age	Salary	Count
Sales	Senior	31..35	46k..50k	30
Sales	Junior	26..30	26k..30k	40
Sales	Junior	31..35	31k..35k	40
Systems	Junior	21..25	46k..50k	20
Systems	Senior	31..35	66k..70k	5
Systems	Junior	26..30	46k..50k	3
Systems	Senior	41..45	66k..70k	3
Marketing	Senior	36..40	46k..50k	10
Marketing	Junior	31..35	41k..45k	4
Secretary	Senior	46..50	36k..40k	4
Secretary	Junior	26..30	26k..30k	6

Give the data tuple having the values “systems, 26..30, 46k..50k”. For the attributes department, age and salary. Find out class label of given tuple wing naive Bayesian classification for status.

b) Write note on SVM classifier **[5]**