प्रज्ञानं ब्रह्म

Manipal

INSPIRED BY LIFE

MANIPAL INSTITUTE OF TECHNOLOGY

(Constituent Institute of Manipal University)

MANIPAL-576104

KNOWLEDGE IS POWER

MANIPAL INSTITUTE OF TECHNOLOGY

SEVENTH SEMESTER BE(CSE) END SEMESTER EXAMINATION – DEC – 2013

ELECTIVE IV: DATA WAREHOUSING AND DATA MINING (CSE 433)

02-12 -2013

TIME : 3 HOURS MAX.MARKS : 50

Instruction to Candidates

- Answer **any five** full questions.

1A. The four key words - subject-oriented, integrated, time-variant, and nonvolatile - distinguish data ware house from other data repository systems. Justify

1B. Give lattice of cuboids making up a 4-D Sales data cube for time, item, location and supplier. Illustrate the OLAP operations: Roll up, Drill down, Slice and Pivot using Sales data cube.

1C. List the basic steps in attribute-oriented induction. Give an example to derive ‘Prime generated relation’ (2+(2+2)+(2+2))

2A. With necessary examples, explain the Iceberg Cube and Closed Cube options in Cube materialization.

2B. What is Data Mining? How it is different from Database processing?

2C. Briefly outline how to compute the dissimilarity between objects of following type:

i. Nominal attributes ii. Asymmetric binary attributes iii. Numeric attributes
iv Ordinal variables ((2+2)+2+(1X4))

3A. Find frequent item sets and association rules with min. confidence = 60% and min. support = 40% for the sales data given below using Apriori algorithm:

T ID	Item Set
100	Milk, Bread, Jam
101	Bread, Butter, Juice
102	Soda, Bread, Butter
103	Bread, Juice, Soda
104	Milk, Juice

3B. How correlation analysis is done using lift and X^2?

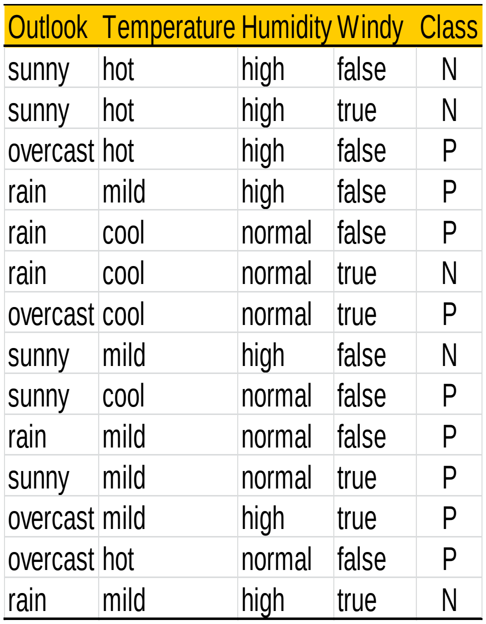
3C. Explain the pincer-search method to find frequent itemsets. (4+2+4)

4A. Build a decision tree for the training data set shown in the Fig a. to classify the tuple x= <rain, hot, high , false> with information gain as attribute selection measure.
4B. Explain Bagging ensemble generation to improve the classification accuracy.
4C. Calculate the precision and recall for the confusion matrix shown in Fig b.
(5+3+2)

5A. Give pseudo code for PAM- k-medoid partitioning algorithm to cluster the given data set. Compare k=medoid with k-means clustering technique with respect to robustness and computational complexity.
5B. Explain following clustering quality measuring parameters
i) Cluster homogeneity ii) Cluster Compactness iii) Rag bag.
5C. For the distance matrix shown in Fig c. draw the dendogram using single link and complete link (5+3+2)

6A. Explain following with respect to Back propagation algorithm
i) Propagate the inputs forward. Ii) back propagate the error.

6B. Briefly explain different types of web mining.
6C. With examples explain collective and contextual outlier. (5+3+2)



Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

Fig a. weather data set

Classes	yes	No
Yes	90	210
No	140	9560

Fig b. Confusion matrix

Item	P	Q	R	S	T
P	0	1	2	2	3
Q	1	0	2	4	3
R	2	2	0	1	5
S	2	4	1	0	3
T	3	3	5	3	0

Fig c. Distance matrix

****