53 (IT 701) DMDW

2017

DATA MINING & DATA WAREHOUSE

Paper : IT 701

Full Marks : 100

Time : Three hours

The figures in the margin indicate full marks for the questions.

*Answer **any five** questions.*

1. *(a)* Explain the sampling (Non-parametric methods) in the Numerasity reduction and also the types of sampling.

   (b) What is data discretization and also mention the data discretization methods ?

   (c) What is concept hierarchy and specify the concept Hierarchy Generation for Nominal Data ?

   8+6+6

2. *(a)* What is role of smoothing in data cleaning and mention the techniques for data smoothing ? Given prime value (in INR) such as 160, 40, 32, 60, 50, 122, 125, 121, 128, 124, 150 and 145. Apply the Binning methods to partition the data.

(b) Use these methods to normalize the following group of data : 100, 400, 600, 800 and 1000.

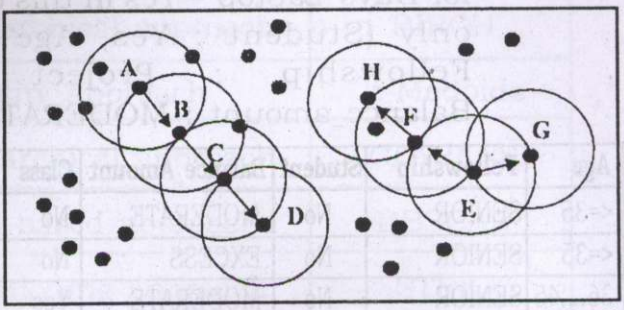
(i) min-max normalization by setting min = 0, max = 1

(ii) z-score normalization using standard deviation.

7+5+8

3. *(a)* Write down the name of clustering methods for large database. Explain the CURE approach with the algorithm. Mention its effectiveness in generating clusters.

(b) Consider the given below figure for a given ϵ represented by the radius of circles and Min Pts =2. Based on the above definitions, mention the label points that are in density-reachability and density connectivity.



12+8

4. (a) Write down the apriori algorithm steps and mention the 2-key steps in the implementation of apriori algorithm.

(b) A database has five transaction. Let min-sup= 50% and min-conf =60%. Find all frequent item sets using Apriori.

TID	Item sets
T100	{ D, E, B, O }
T200	{ B, A, S, U }
T300	{ G, A, O, N }
T400	{ B, A, S, U }
T500	{ G, R, A, M }
T600	{ J, H, A, R }
T700	{ B, E, S, U }

10+10

5. *(a)* Explain the Naive Bayesian Classification (NBC) with the five steps.

(b) Classify this training data set (given below) using Naive Bayesian classification and predicts this data set for Buys_Laptop = Yes in this condition only (Student : Yes, Age : > 45, Fellowship : Project fellow, Balance_amount : MODERATE).

Age	Fellowship	Student	Balance_Amount	Class : Buy_Laptop
<=35	SENIOR	No	MODERATE	No
<=35	SENIOR	No	EXCESS	No
36....45	SENIOR	No	MODERATE	Yes
>45	JUNIOR	No	MODERATE	Yes
>45	Project Fellow	Yes	MODERATE	Yes
>45	Project Fellow	Yes	EXCESS	No
36....45	Project Fellow	Yes	EXCESS	Yes
<=35	JUNIOR	No	MODERATE	No
<=35	Project Fellow	Yes	MODERATE	Yes
>45	JUNIOR	Yes	MODERATE	Yes
<=35	JUNIOR	Yes	EXCESS	Yes
36....45	JUNIOR	No	EXCESS	Yes
36....45	SENIOR	Yes	MODERATE	Yes
>45	JUNIOR	No	EXCESS	No

10+10

(iii) OLAP is major task of data warehouse system.

(iv) Data analysis and decision making are processed in OLAP.

(v) The operation of moving from finer-granularity data to a coarser granularity is called a Rollup.

(vi) The Roll-up operation navigates from more detailed data to less detailed data.

(vii) The class labels of training data are unknown in Un-supervised learning.

(viii) In Supervised learning, training data are accompanied by labels indicating the class of the observations.

(c) Write short notes on : ***(Any four)***

(i) Correlation Analysis (Numeric Data)

(ii) Neural network

(iii) Roll up and Dice operation

(iv) Naive Bayesian Classification

(v) Data Cube (Multidimensional data)

4+4+12
