



**DHARMSINH DESAI UNIVERSITY, NADIAD**  
**FACULTY OF TECHNOLOGY**  
**B.TECH. SEMESTER V [I.T]**

**SUBJECT: (IT-704) Data Analysis & Information Extraction**

**Examination : Third Sessional**      **Seat No. : \_\_\_\_\_**  
**Date : 10/10/2014**      **Day : Friday**  
**Time : 1:00 to 2:15**      **Max. Marks : 36**

**INSTRUCTIONS:**

1. Figures to the right indicate maximum marks for that question.
2. The symbols used carry their usual meanings.
3. Assume suitable data, if required & mention them clearly.
4. Draw neat sketches wherever necessary.

**Q.1 Do as directed.**

- (a) State the difference between k-means and k-medoids. [2]
- (b) Is Prediction same as Classification? If YES, justify by giving examples. If NO, highlight the differences. [2]
- (c) Explain what are lazy learners or instance-based classifiers with examples. [2]
- (d) What is the difference between PAM and CLARA? Which one is better? [2]
- (e) What are the difficulties of the hierarchical clustering? Explain in brief. [2]
- (f) What is an epoch? And state various types of an epoch. [2]

**Q.2 Attempt *Any Two* from the following questions. [12]**

- (a) Define the requirements of the clustering in the data mining. [6]
- (b) Explain the genetic algorithm and fuzzy logic approach in the classification. [6]
- (c) What are the basic steps for Training and Classification in Decision Tree Induction algorithm? [6]

**Q.3 (a) Training data tuples of XYZ Cos. customer database are as follows: [6]**

<i><b>RID</b></i>	<i><b>age</b></i>	<i><b>income</b></i>	<i><b>student</b></i>	<i><b>credit rating</b></i>	<i><b>Class: buys computer</b></i>
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31...40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31...40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31...40	medium	no	excellent	yes
13	31...40	high	yes	fair	yes
14	>40	medium	no	excellent	no

Using Bayesian Classifier to predict the class label of the following sample:  
X = (age = "<=30", income = "medium", student = "yes", credit\_rating = "fair").

- (b) A person has 5 cars. There is a linear relationship between a car and repair, where repairs are considered in a hundreds of rupees. The information for the repair costs collected for the last year is given in table. [6]

Car	1	2	3	4	5
Age	1	3	3	5	6
Repairs	4	6	7	7	9

A person wants to predict the repair costs for the next year for the car which is 5 years old now.

**OR**

- Q.3** (a) The following is the data for a multilayer feed-forward neural network. Let the learning rate be 0.9 and the first training sample,  $X = (1,0,1)$ , whose class label is 1. Initial input, weight and bias values: **[6]**

X1	X2	X3	w14	w15	w24	w25	w34	w35	w46	w56	$\Theta_4$	$\Theta_5$	$\Theta_6$
1	0	1	0.2	-0.3	0.4	0.1	-0.5	0.2	-0.3	-0.2	-0.4	0.2	0.1

Calculate the net input, output, error at each node, weight and bias updating.

- (b) The following data of customers visiting hotel “Supreme” quarterly is available. **[6]**  
The no. of customers visited quarterly:

Year	I	II	III	IV
2001	200	180	185	95
2002	220	188	173	83
2003	220	176	161	87

By using the ratio to moving average method, calculate the seasonal index for each quarter.