DHARMSINH DESAI UNIVERSITY, NADIAD
FACULTY OF TECHNOLOGY
B.TECH. SEMESTER VII IT

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

Examination	**: Third sessional**	**Seat No.**	**: _____**
Date	**: 15/10/2016**	**Day**	**: Saturday**
Time	**: 2:15 to 3:30**	**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) What is a supervised learning? Give an appropriate example. **[2]**
(b) Explain what a tree-pruning is. Also explain why it should be used. **[2]**
(c) What is an activation function in a neural network? **[2]**
(d) Explain the various operators used in a genetic algorithm. **[2]**
(e) "In clustering k-medoids should be preferred over k-means clustering algorithm." – State true or false. Justify your answer. **[2]**
(f) Explain relative inter connectivity and relative closeness. **[2]**

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

(a) Explain the algorithm of decision tree induction with example. **[6]**
(b) Explain the DBSCAN & OPTICS method of clustering. **[6]**
(c) Explain BIRCH & CURE method of clustering approach. **[6]**

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

RID	age	income	student	cred_rating	Class: buy_comp
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) The following data shows the number of Lincoln Continental cars sold by a dealer in Queens during the 12 months of 1994. **[6]**

Month	Number sold	Month	Number sold
Jan	52	July	54
Feb	48	Aug	65
Mar	57	Sept	70
Apr	60	Oct	80
May	55	Nov	90
June	62	Dec	75

a) Calculate the three month moving average for this data.
b) Calculate the five month moving average for this data.
c) Which one of these two moving averages is a better smoothing technique and why?

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. **[6]**
Initial input, weight and bias values:

X1	X2	X3	w14	w15	w24	w25
1	0	1	0.2	-0.3	0.4	0.1
w34	w35	w46	w56	Θ4	Θ5	Θ6
-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 table is given: **[6]**

Year	Rentals per quarter			
	I	II	III	IV
1991	370	320	470	420
1992	350	380	520	430
1993	390	370	530	460

Derive seasonal index using ratio to moving average method.