[Total marks: 60

(03 hours)

N. B.:

1. Attempt any **one of the two Sections Data Mining or Operations Research**.

2. Attempt **any two questions from each Section**.

3. Figures to the right indicate full marks.

4. Simple non-programmable calculator is allowed.

SECTION I: Data Mining

Q.1 (a) Define Entropy in information theory by stating different issues involved in it. (05)

(b) For each attribute of the following table write three classification rules of your   (05)
choice and find the total error corresponding to each attribute using 1R classifier.
According to you which are the best attribute and corresponding rules. Justify
your answer.

Outlook	Temperature	Humidity	Windy	Class
Sunny	Hot	High	False	N
Sunny	Hot	High	True	N
Overcast	Hot	High	False	Y
Rain	Mild	High	False	Y
Rain	Cool	Normal	False	Y
Overcast	Cool	Normal	True	Y
Sunny	Mild	High	False	N
Sunny	Cool	Normal	False	Y

(c) Consider the following Car Insurance information with two attributes: age and   (05)
car type. Where age is an ordinal attribute, car type is a categorical attribute,
class L: low and H: high (risk). Use ID3 algorithm to arrive at a decision tree
in classification of data.

Age	Car type	Class
> 41	Maruti	L
> 41	Hyundai	H
< 41	Maruti	H
< 41	Indica	H
> 41	Maruti	L
> 41	Hyundai	H

73011 Page 1 of 4

Q.2 (a) Explain BIRCH algorithm by considering suitable example. (05)

(b) Explain k-mean clustering method for the set $\{11, 2, 20, 4, 10, 12, 3, 30, 25\}$ by (05)
dividing the set into three clusters. Initially begin with clusters with mean 3, 4
and 11.

(c) Represent the following distance matrix by dendogram after clustering using (05)
average linkage criteria of agglomerative method.

	M_1	M_2	M_3	M_4	M_5
M_1	0	1	2	2	3
M_2	1	0	2	4	3
M_3	2	2	0	1	5
M_4	2	4	1	0	3
M_5	3	3	5	3	0

Q.3 (a) Explain market basket analysis with example. (05)

(b) Define support and confidence to measure the strength of association rule. Cal- (05)
culate support and confidence for the rules jam $\Rightarrow$ (butter, bread) and
butter $\Rightarrow$ bread, from the following:
Basket 1: bread, butter, jam   Basket 2: bread, butter
Basket 3: beer, bread   Basket 4: bread, butter, milk
Basket 5: beer, milk

(c) Explain how to generate Frequent pattern (FP) tree for the following data sets (05)
with minimum support 3:

TID	Items bought
1	$\{f, a, c, d, g, i, m, p\}$
2	$\{a, b, c, f, l, m, o\}$
3	$\{b, f, h, j, o\}$
4	$\{b, c, k, s, p\}$
5	$\{a, f, c, e, l, p, m, n\}$

73011

SECTION II: Operations Research

Q.4 (a) Define assignment problem and write down stepwise procedure to solve assignment problem. (08)

(b) Solve the following assignment problem and determine the optimum assignment schedule. (07)

	I	II	III	IV	V
A	6	5	8	11	16
B	1	13	16	1	10
C	16	11	8	8	8
D	9	14	12	10	16
E	10	13	11	8	16

Q.5 (a) Explain Monte Carlo simulation method. (05)

(b) Explain Dr. Karmarkar's method. (05)

(c) Explain difference between simplex and revised method. (05)

Q.6 (a) Define PERT and CPM. Also explain advantages and differences between PERT and CPM. (07)

(b) Consider the following PERT system network. (08)
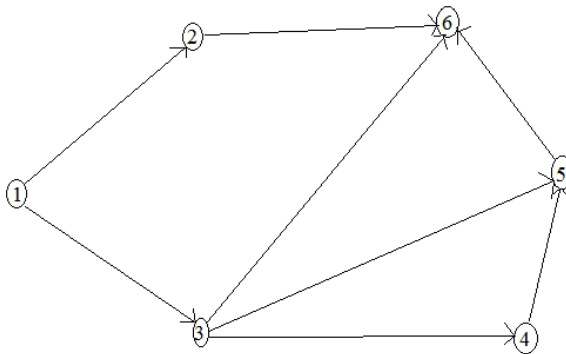
The usual three estimates for the time required (in months) for each these activities are as given below. Find earliest and latest expected times to reach each node and critical path.

Activity	(1, 2)	(1, 3)	(2, 6)	(3, 4)	(3, 5)	(3, 6)	(4, 5)	(5, 6)
t_0	8	6	9	5	8	11	3	5
t_m	9	7	12	5	10	15	4	6
t_p	10	9	15	5	11	20	6	8

SECTION III: Statistical Process Control

Q.7 (a) Show that the actual set point for the manipulatable variable at the end of   (07)
period t is simply the sum of all the adjustments through time t.

(b) Write a short note on the following.

   i. Bounded adjustment chart.   (03)

   ii. Rounded adjustment chart.   (02)

   iii. Discrete proportional Integral control equation and Discrete proportional   (03)
   integral derivative control equation.

Q.8 (a) Define controllable and uncontrollable variables by giving one example. Also   (07)
Explain how experimental design can improve the process performance?

(b) i. Explain the concept of Response Surface Methodology with an example of   (05)
method of steepest ascent.

   ii. What is design resolution? Explain Resolution III designs.   (03)

Q.9 (a) Bring out the relationship between the S/N ratio and quality loss after adjust-   (07)
ment. Also comment on signal to noise ratios for static type problems.

(b) i. Explain signal factors and noise factors.   (03)

   ii. Briefly explain the D-D type of quality characteristic for signal to noise   (05)
   ratios for dynamic problems.

———————————