Roll No:

STATE OF THE PERSON NAMED IN	Maria Maria Maria	Sand from the same of	Andrew Street, Street, St. St.	Property of the Park of the Pa	
100	COLUMN TWO IS NOT THE OWNER.	THE PERSON NAMED IN	CONTROL CONTROL	STREET, STREET, GARNING	Manager Comments In
Share Market St.	THE RESERVE	1000	THE PARTY OF THE PARTY OF	STREET, STREET, SQUARE,	Married Committee
	CF 128	100000	100		A CASE CONTRACTOR
DOMESTIC OF		DESCRIPTION OF THE PERSON	000	100000000000000000000000000000000000000	2000 -0000
Name and Address of the		COLUMN TOWNS	THE RESERVE	COLUMN TWO IS NOT THE OWNER.	Control of the last of the las
Witness Street, or other Designation of the last	Married Woman	STREET, SQUARE	Maria San State of	and the same of the same of	A STATE OF THE PARTY OF

NOIDA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GREATER NOIDA (An Autonomous Institute)

Affiliated to Dr. A.P. J. Abdul Kalam Technical University, Uttar Pradesh, Lucknow

Course: B.TECH

Branch: CSBS

Semester: III

Sessional Examination III

Year: (2021 - 2022)

Subject Name: Computational Statistics

Time: 1.15Hours

[SET-B]

Max. Marks:30

General Instructions:

> This Question paper consists of 2 pages & 5 questions. It comprises of three Sections, A, B, and C

Section A Question No- 1 is objective type questions carrying 1 mark each, Question No- 2 is very short answer type carrying 2 mark each. You are expected to answer them as directed.

➤ <u>Section B - Question No-3</u> is Short answer type questions carrying 5 marks each. Attempt any two out of three questions given.

Section C Question No. 4 & 5 are Long answer type (within unit choice) questions carrying 6 marks each. Attempt any one part <u>a or b.</u>

	D'	SECTION – A	[08Marks]	
1.	All q	uestions are compulsory	(4×1=4)	
	2	Which of the following clustering requires merging approach? i) Partitional ii) Hierarchical iii) Naïve Bayes iv) None of the above	(1)	COS
	b.	Which of the following techniques would perform better for reducing dimensions of a data set? i) Removing columns which have too many missing values ii) Removing columns which have high variance in data iii) Removing columns with dissimilar data trends iv) None of the above	(1)	CO3
	c.	Which of the following is not the part of the exploratory factor analysis process? i) Extracting factors ii) Determining the number of factors before the analysis iii) Rotating the factors iv) Refining and interpreting the factors	(1)	CO4
	d.	The most popularly used dimensionality reduction algorithm is Principal Component Analysis (PCA). Which of the following is/are true about PCA? 1. PCA is an unsupervised method 2. It searches for the directions that data have the largest variance 3. Maximum number of principal components <= number of features 4. All principal components are orthogonal to each other Which is True:	(1)	CO3

		ii) 1,2 iii) 1,2	only 2 2,3 Il of the a	above				Oloc		(2×2=4)	
		estions are co					0		- /	(2)	CO5
	a.	What is dista	ance mea	sure in cl	lustering	?		•		(2)	CO3
	b.	What do you	underst	and by D	imension	nality Re	duction	1?		[10Marks]	
				SECTIO	N-B	1. 1. 1. 1.	1 1 1/2	2, 5, 5, 5,		$(2\times 5=10)$	
	Answ	er any two of	the follo	wing-	Vale			• • •	-ainto	(5)	CO5
	a.	Cluster the three cluster A1 (2, 10), A A8(4, 9).	L	moone	45 41 4 4 4				and the second second		
	b.	1 41	What does a Principal Component in a PCA signify? How can we represent them mathematically?							(5)	CO3
	c.	Factor Load	ding is a	data redu	iction me	thod des	igned t	o explain t	he	(5)	CO4
		correlations	s. Explair	the state	ement.		Araban a			[12Marks]	
	150			SECTIO	DN - C					$(1\times6=6)$	
4	Ans	Compute the	of the follo	owing-			40 . 1-1			(6)	CO3
S _C		$ \begin{array}{l} CLASS 1 \\ X = 2, 3, 4 \\ Y = 1, 5, 3 \\ CLASS 2 \\ X = 5, 6, 7 \\ Y = 6, 7, 8 \end{array} $						0	16	(6)	
	b.	Differentia	Differentiate between Exploratory Factor analysis and Confirmatory								CO4
			Factor analysis.								
5.	Ans	iswer any one of the following-								(1×6=6) (6)	COS
	a.	Perform DBSCAN on the given problem with €=2and minimum								(0)	000
		point=2.					7 y				
				2		AU	10		***		
		A1		2		()\	5				
		A2 A3		8			4				
		A3		5			8				
		A5		7			5				
		A6	AU								
		A7								A STATE OF	
		A8									
	h		Using the given distance matrix find the clusters using average link							(6)	CO5
1	1 2 1		technique and draw the dendrogram (Agglomerative approach).								
			P1	P2	P3	P4	P5	P6			
	1 . 1	AT A STATE OF THE PARTY OF THE	0								
	* /	P1	0	A CONTRACTOR OF THE PARTY OF TH		4 4 1 2	1 1	C. Carrie	100000000000000000000000000000000000000		Pr. 10 10 10 10 10 10 10 10 10 10 10 10 10
	*/	P1 P2	0.23	0.		100					
			15 18 18	0 0.15	0					1	
	c	P2	0.23	0.15	0.15	0			1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	COL	
	G	P2 P3	0.23	0.15	1	0 0.29 0.22	0 0.39	0		COL	