Printed page: 2 Subject Code: ACSBS0301

Roll No: | | | | | | | | | | | | |

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.
- Section B - Question No-3 is Short answer type questions carrying 5 marks each. Attempt any two out of three questions given.
- Section C -Question No. 4 &5are Long answer type (within unit choice) questions carrying 6marks each. Attempt any one part *a* or *b*.

		SECTION – A	[08Marks]	
1.	*All questions are compulsory*		(4×1=4)	
	a.	Which of the following clustering requires merging approach? i) Partitional ii) Hierarchical iii) Naïve Bayes iv) None of the above	(1)	CO5
	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

		i) 1 only ii) 1,2 iii) 1,2,3 iv) All of the above		
2.	*All questions are compulsory*		(2×2=4)	
	a.	What is distance measure in clustering?	(2)	CO5
	b.	What do you understand by Dimensionality Reduction?	(2)	CO3
		SECTION – B	[10Marks]	
3.	*Answer any <u>two</u> of the following-*		(2×5=10)	
	a.	Cluster the following eight points with (x, y) representing locations into three clusters using K means. A1 (2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9).	(5)	CO5
	b.	What does a Principal Component in a PCA signify? How can we represent them mathematically?	(5)	CO3
	c.	Factor Loading is a data reduction method designed to explain the correlations. Explain the statement.	(5)	CO4
		SECTION – C	[12Marks]	
4	*Answer any <u>one</u> of the following-*		(1×6=6)	
	a.	Compute the principal component of following data- <u>CLASS 1</u> X = 2 , 3 , 4 Y = 1 , 5 , 3 <u>CLASS 2</u> X = 5 , 6 , 7 Y = 6 , 7 , 8	(6)	CO3
	b.	Differentiate between Exploratory Factor analysis and Confirmatory Factor analysis.	(6)	CO4
5.	*Answer any <u>one</u> of the following-*		(1×6=6)	
	a.	Perform DBSCAN on the given problem with €=2and minimum point=2.	(6)	CO5

	x	y
A1	2	10
A2	2	5
A3	8	4
A4	5	8
A5	7	5
A6	6	4
A7	1	2
A8	4	9

	b.	Using the given distance matrix find the clusters using average link technique and draw the dendrogram (Agglomerative approach).					(6)	CO5

	P1	P2	P3	P4	P5	P6
P1	0					
P2	0.23	0				
P3	0.22	0.15	0			
P4	0.37	0.20	0.15	0		
P5	0.34	0.14	0.28	0.29	0	
P6	0.23	0.25	0.11	0.22	0.39	0