

**BIRLA INSTITUTE OF TECHNOLOGY, MESRA, RANCHI
(END SEMESTER EXAMINATION)**

**CLASS: M.TECH. (EXECUTIVE)
BRANCH: CSE**

**SEMESTER : II
SESSION : SP/2022**

SUBJECT: CS538 UNSUPERVISED LEARNING

TIME: 3 Hours

FULL MARKS: 50

INSTRUCTIONS:

1. The question paper contains 5 questions each of 10 marks and total 50 marks.
 2. Attempt all questions.
 3. The missing data, if any, may be assumed suitably.
 4. Tables/Data hand book/Graph paper etc. to be supplied to the candidates in the examination hall.
-

- Q.1(a) Why clustering is considered as an unsupervised learning algorithm? Mention at least four requirements of clustering process. [5]
- Q.1(b) What is the difference between symmetric and asymmetric binary dissimilarity, explain with suitable example? Consider two objects $O_1(22,1,42,10)$ and $O_2(20,0,6,8)$. Calculate Minkowski distance (L_p norm) between the two objects, (assume $p = 3$) [5]
- Q.2(a) State at least four limitations of traditional K-Means clustering algorithm. Which method is more robust k-Means or K-Medoid, justify your answer? [5]
- Q.2(b) Illustrate how CLARA works well for large data set? What do you mean by single linkage and complete linkage in hierarchical clustering? [5]
- Q.3(a) Explain the following terms used in density-based clustering with suitable example: (i) Core point (ii) Density reachable (iii) Density connected. [5]
- Q.3(b) What do you understand by the density of a grid cell? Why density factor is used in grid-based clustering? [5]
- Q.4(a) Mention and explain three major cluster validity measures. Why and how entropy-based measure works for the validation of clustering? [5]
- Q.4(b) Consider a data set consists of 150 instances and two attributes. The instances are labelled with three classes C_1, C_2, C_3 . Each class consists of 50 instances. A K-Means clustering is implemented on the data set with parameter $K = 3$. The result of the clustering process creates three clusters U_1, U_2, U_3 . U_1, U_2, U_3 contain 30, 24, 96 instances respectively. The contingency table is given below: [5]

	C1	C2	C3
U1	30	0	0
U2	20	4	0
U3	0	46	50

Compute the validity of the clustering process using the following measures:

- (i) F-Score (ii) Jaccard Coefficient

Depending on the values of the above-mentioned measures what is your conclusion about the clustering process?

- Q.5(a) How Beta CV measure is different from Dunn Index? Give mathematical explanation of each measure. [5]
- Q.5(b) Mention at least three issues to deal with anomalies in data set. How objective function can be used to assess the belongingness of an object in cluster-based anomaly detection method? [5]

:::18/07/2022:::