7th Semester DATA MINING SESSIONAL EXAM -2, 2021

Time : 1 hour 15 Minutes (9:00 -10:00 A.M., Uploading 10:00 to 10:15 a.m. Max. Marks: 15

Answer all questions

Instructions: Show all calculations clearly. Late submissions will carry penalty.

Q.1. Suppose that the data mining task is to cluster points (with x, y representing location) into three clusters, where the points are $A1(4,11)$, $A2(3,5)$, $A3(8,4)$, $B1(5,8)$, $B2(7,5)$, $B3$ (6,4), $C1(1,2)$, $C2(4,9)$

The distance function is Euclidean distance. Suppose initially we assign *A2*, *B2*, and *C2* as the center of each cluster, respectively. Use the *k-means* algorithm to show
(a) The three cluster centers after the first round of execution.
(b) The final three clusters.

Q.2. Can K-means algorithms find the global optimum? Justify your answer using an example.

Q.3. A random sample of 400 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table:

	High School	Bachelors	Masters	Ph.d.	Total
Female	60	57	50	36	203
Male	40	45	55	57	197
Total	100	102	105	93	395

Are gender and education level dependent at 5% level of significance? Use standard chi-square table to determine the critical values.