CT – 3 || CSE 4891(A) - Data Mining

1. **(Marks 8)** Suppose that the data mining task is to cluster points (with (x,y) representing location) into three clusters, where the points are: A(2,10), B(2,5), C(8,4), D(1,2), E(4,9). If A and D are chosen as initial center points, find the cluster centers after the **second** round of ***K-means*** algorithm. You can consider Euclidean distance as the distance measure.
2. **(Marks 6)** Explain the differences between ***K-means*** and ***K-medoids*** algorithm? Which algorithm is best? Explain with an example.
3. **(Marks 6)** Briefly describe Single Link and Complete Link similarity measures in Hierarchical Clustering. What are the advantages and disadvantages of each method?