



UNIVERSITY OF COLOMBO, SRI LANKA



UNIVERSITY OF COLOMBO SCHOOL OF COMPUTING

IS4003/SCS4104/CS4104 – Data Analytics

Academic Year 2022 - Semester 2

Assignment -Clustering

(Submit on or before September 23, 2022)

Answer All Questions

(1)

(i) Discuss the characteristics of the data sets for the suitability of the following inter-cluster dissimilarities used in agglomerative hierarchical clustering algorithms.

- (a). Single link
- (b). Complete link
- (c). Average Link

(marks 40)

(ii) Use the agglomerative hierarchical clustering algorithm with the **Manhattan distance** function for the following data set with two variables, x and y shown in Table 1 to obtain the best hierarchy of clusters for the above three inter-cluster dissimilarities.

objects	x	y
1	2	9
2	2	5
3	8	4
4	5	8
5	7	5
6	6	4
7	1	2
8	4	9
9	7	3
10	2	4
11	8	5

Table 1: coordinates of 11 objects

(a) Draw separate Dendrograms to show the hierarchies obtained for **each** inter-cluster dissimilarity given above. Note that you are required to show all the necessary calculation in determining the best pair of clusters at each iteration.

(marks 40)

(b) How do you use the Dendrogram to compare the similarities of clusters at different levels of hierarchies?

(marks 20)