

PACKAGE: MTU CLUSTER

ANALYSIS: 1. AGGLOMERATIVE HIERARCHICAL CLUSTERING
2. DIVISIVE HIERARCHICAL CLUSTERING

AUTHOR: - Revanth Chappidi, Dhairya Kothari, Akshay Gaonkar, Bharath Inti

CONTACT AT: - chappidi@mtu.edu, dmkothar@mtu.edu, adgaonka@mtu.edu, binti@mtu.edu

MANAGED BY: - Dr. Shane T. Mueller (shanem@mtu.edu) ; Revanth Chappidi, Dhairya Kothari, Akshay Gaonkar, Bharath Inti

HIERARCHICAL CLUSTERING

Hierarchical clustering is a type of clustering algorithm which aims to build a hierarchy of clusters. There are two ways:

1. Agglomerative clustering:

It is a bottom up process in which pairs of clusters having minimum intraclass distance are merged successively until all the clusters have been merged into a single cluster containing all the data. agnes is the agglomerative nesting method which can be deployed for executing agglomerative clustering in R. The function prototype is as follows:

x is the distance among the observations and it can be calculated using following options:

- Euclidean
- Manhattan
- Minkowski
- Maximum
- Canberra
- Binary

`agnes(x, method = "single", ...)`

method is the types of linkages. The linkages differ in the way they merge the clusters.

- a. Single linkage : The clusters are formed using minimum distance between components as the criteria.
- b. Complete linkage : The clusters are formed using maximum distance between components as the criteria.
- c. Average linkage : The clusters are formed using unweighted average distance between components as the criteria.
- d. Weighted linkage : The clusters are formed using weighted average distance between components as the criteria.
- e. Ward linkage : The criteria for clustering the components is the inner squared distance using the minimum variance algorithm.

2. Divisive Clustering:

It is a top down process in which the data is splitted continuously until all the data tuples are segregated. diana is the divisive hierarchical clustering method in R. The function prototype is as follows:

x is the distance among the observations and it can be calculated using following options:

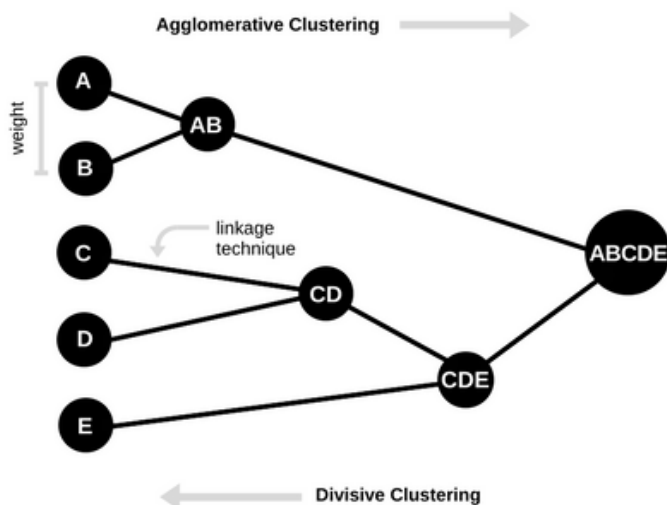
- Euclidean
- Manhattan
- Minkowski
- Maximum
- Canberra
- Binary

diana(x,)

USE CASE:

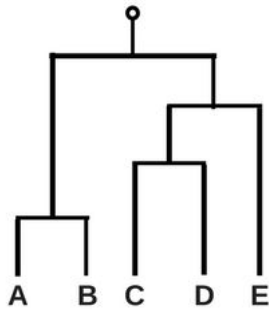
Consider five points: A, B, C, D, E. Each of these points are having some similarities between them - W (Weight between the components). The following image explains the workflow of clustering the components with hierarchical clustering:

- Agglomerative clustering (left to right)
- Divisive clustering (right to left)



The above (representative) image explains the clustering of components: A, B, C, D, E with some linkage technique. The steps also show how the formation of dendrogram takes place.

Dendrogram is shown in the image below.



Steps to use Hierarchical clustering (in Jamovi).

1. Install the MTU cluster package.
 2. Select dataset using file > open > browse OR file > open > example.
 3. From the analysis menu (Given above data), select MTU cluster > hierarchical clustering - agglomerative OR MTU cluster > hierarchical clustering - divisive.
 4. Select the columns which you are interested to analyze.
 5. Choose the appropriate distance measure from the distance dropdown menu and (applicable only for agglomerative clustering) also choose the linkage method from the linkage method dropdown menu.
 6. If you wish to normalize/ standardize the data, tick the “normalize” checkbox.
 7. Wait for the output to show on the Jamovi output screen displayed on right side.
 8. You may choose to export the output using file > export.
-