**Hierarchical clustering**

This is an algorithm that groups similar objects into groups called *clusters*. The endpoint is a set of clusters*,*where each cluster is distinct from each other cluster, and the objects within each cluster are broadly similar to each other.

Hierarchical clustering can be performed with either a *distance matrix*or *raw data.* When raw data is provided, the software will automatically compute a distance matrix in the background.

Hierarchical clustering starts by treating each observation as a separate cluster. Then, it repeatedly executes the following two steps:

1. identify the two clusters that are closest together,
2. merge the two most similar clusters. This iterative process continues until all the clusters are merged together.

There are many distance metrics that have been developed. The choice of distance metric should be made based on theoretical concerns from the domain of study. That is, a distance metric needs to define similarity in a way that is sensible for the field of study.

For example, if clustering crime sites in a city, city block distance may be appropriate. Or, better yet, the time taken to travel between each location. Where there is no theoretical justification for an alternative, the distance between two points, commonly known as the Euclidean is generally used, as it is usually the appropriate measure of distance in the physical world.

*Ward’s method* is the sensible default. This method works out which observations to group based on reducing the sum of squared distances of each observation from the average observation in a cluster. This is often appropriate as this concept of distance matches the standard assumptions of how to compute differences between groups in statistics (e.g., *ANOVA*, *MANOVA*).

The strengths of hierarchical clustering are that it is easy to understand and easy to do. The weaknesses are that it rarely provides the best solution, it involves lots of arbitrary decisions, it does not work with missing data, it works poorly with mixed data types, it does not work well on very large data sets, and its main output, the dendrogram, is commonly misinterpreted. There are better alternatives, such as *latent class analysis.*

There are four types of clustering algorithms in widespread use: *hierarchical clustering*, *k-means cluster analysis, latent class analysis,*and *self-organizing maps.*The math of hierarchical clustering is the easiest to understand. It is also relatively straightforward to program. Its main output, the dendrogram, is also the most appealing of the outputs of these algorithms but it doesn’t always provide the best solution.