Hierarchical Clustering using an example

Deboky Saha · Follow

3 min read · Feb 13, 2022

Hierarchical clustering algorithms create a hierarchy of clusters in which each node is made up of the clusters of its daughter nodes. Hierarchical clustering can be divided into two types: divisive and agglomerative. Divisive works from the top down; thus, it takes all observations in a huge cluster and breaks them down into smaller chunks.

Agglomerative is the opposite of divisive. As a result, it works from the bottom up, with each observation starting in its cluster and pairs of clusters merging as they progress up the hierarchy. Agglomeration refers to the act of gathering or amassing items, which is what the cluster does. Among data scientists, the agglomerative technique is more prevalent.

This approach creates the hierarchy from the individual pieces by gradually merging clusters. Let's imagine we wish to group six cities in England together based on their distances from one another and these are London, Manchester, Birmingham, Leeds, Cambridge, and Bath.

We create a distance matrix, with the distance between cities represented by the numbers. The distances between each pair of cities are shown in this table. Each city is assigned to its cluster at the start of the process. So if we have six cities, we have six clusters each containing just one city.

	London	Manchester	Birmingham	Leeds	Cambridge	Bath
London		163.05	113	195	49	115
Manchester			71	36	132	146
Birmingham				92	87	79
Leeds					130	171
Cambridge						120
Bath						

Table 1: Distance between cities (in miles)

The first stage is to decide which cities should be merged into a cluster (we'll call them clusters from now on). We wish to take the two nearest clusters based on the distance in most cases. We build a cluster out of Manchester and Leeds because these are the nearest clusters on the distance matrix. Here, we are only using a one-dimensional distance feature. However, our object can be multidimensional, and distance measurement can be Euclidean, Pearson, average distance, or other methods, depending on the data type and domain knowledge.

In the distance matrix, we must also merge these two closest cities. As a result, rows and columns are consolidated as the cluster is built. For example, how do we compute the distance between London and the Manchester/Leeds cluster? There are a variety of techniques, but let's say we use the distance from the London cluster's centre to Manchester/Leeds as an example. We now have one fewer cluster after updating the distance matrix.

	London	Manchester/ Leeds	Birmingham	Cambridge	Bath
London		164	113	49	115
Manchester/Leeds			81	128	159
Birmingham				87	79
Cambridge					120
Bath					

Table 2: Distance between cities after 1st clustering

Next, we look for the closest clusters once again. In this case, Cambridge and London are the closest ones which creates another cluster.

	London/ Cambridge	Manchester/Leeds	Birmingham	Bath
London/ Cambridge		133	92	97
Manchester/Leeds			81	159
Birmingham				79
Bath				

Table 3: Distance between cities after 2nd clustering

In the next step, the closest distance is observed between the Birmingham cluster and the Bath cluster. Forming a new cluster, the data in the matrix table gets updated. The rows and columns are merged as the clusters are merged and the distance updated. This is a common way to implement this type of clustering and benefit from caching distances between clusters.

	London/ Cambridge	Manchester/Leeds	Birmingham/Bath
London/ Cambridge		133	74
Manchester/Leeds			106
Birmingham/Bath			

Table 4: Distance between cities after 3rd clustering

	London/ Cambridge/ Birmingham/Bath	Manchester/Leeds
London/ Cambridge /Birmingham/Bath		42.2
Manchester/Leeds		

Table 5: Distance between cities after 4th clustering

In the same way, the agglomerative algorithm proceeds by merging clusters, and we repeat it until all clusters are merged, and the tree becomes completed. It means until all cities are clustered into a single cluster of size six. Hierarchical clustering is typically visualised as a dendrogram, and a horizontal line represents each merge.

References:

Machine Learning with Python

This course dives into the basics of machine learning using an approachable, and well-known programming language...

www.coursera.org

Zhao, Y. and Karypis, G., 2002, November. Evaluation of hierarchical clustering algorithms for document datasets. In *Proceedings of the eleventh international conference on Information and knowledge management* (pp. 515–524).

Data Science | Hierarchical Clustering | Machine Learning | Unsupervised Learning | Clustering