← Go Back to Data Analysis & Visualization

Course Content

Hierarchical Clustering

Hierarchical Clustering is one of the many types of clustering algorithms.

There are 2 types of hierarchical clustering:

- Agglomerative clustering
- Divisive clustering

Let's see how each of these types of algorithms work:

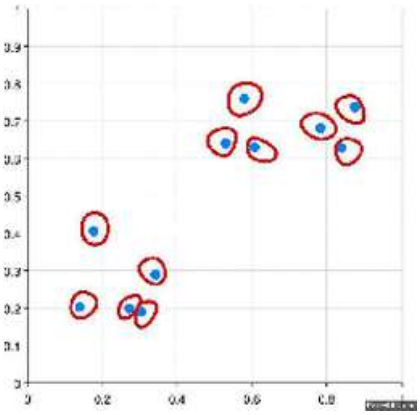
Agglomerative Clustering

It is a **bottom-up** clustering method.

Step 1: Let's say we have N data points, then each of the points will be considered as N different clusters.

Step 2: Find two clusters that are closest among the N clusters (from step 1) and combine them into a single cluster. This will result in having N-1 clusters now.

Step 3: Find the next 2 closest clusters and combine them into a single cluster and we have N-2 clusters now. This will be continued till we have only one cluster containing all the data points. The below graphic demonstrates this process.



Source

To calculate the distance between two clusters, we use the **linkage method**. There are different types of linkages that can be used to compare the distance between two clusters:

- **Complete-linkage**: The distance between two clusters is defined as the longest distance between two points in each cluster.

- **Average-linkage**: The distance between two clusters is defined as the average distance between each point in one cluster to every point in the other cluster. Average-linkage and complete-linkage are the two most popular distance metrics in

hierarchical clustering.

- **Centroid-linkage:** The distance between the centroids of two clusters. It finds the centroid of cluster 1 and centroid of cluster 2 and then calculates the distance between the two before merging.
- **Single-linkage**: The distance between two clusters is defined as the *shortest* distance between two points in each cluster. This linkage may be used to detect high values in your dataset which may be outliers as they will be merged at the end.

Divisive Clustering

In contrast to Agglomerative clustering, divisive clustering is a **top-down** clustering method. We work backward in this method.

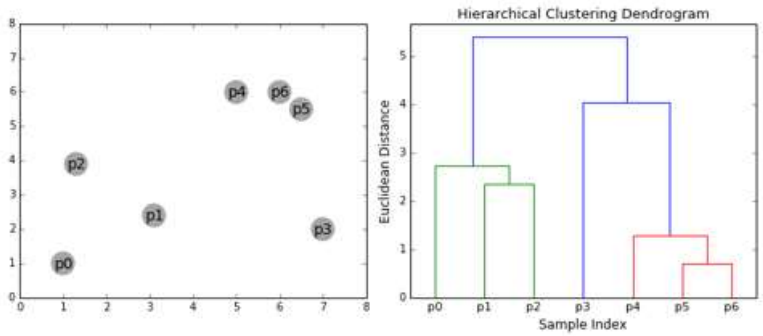
Step 1: All the points are grouped together into a single cluster.

Step 2: The single cluster is split into two different clusters.

Step 3: Step 2 is repeated till the number of clusters is equal to the number of data points.

Dendrogram

A dendrogram is a diagram that shows the hierarchical relationship between objects. It is used to see the output of hierarchical clustering. The below graphic explains, how the dendrogram is being formed as the clustering happens.



Source

Previous Next

Proprietary content.©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

© 2023 All rights reserved.

Help