



Hierarchical Clustering

Introduction: It is a specific type of clustering, or it can be said that it is a specific way to perform clustering. This type of clustering is good for small sized datasets, and they are very intuitive.



Harshit Dawar · [Follow](#)

Published in Analytics Vidhya

3 min read · May 24, 2020

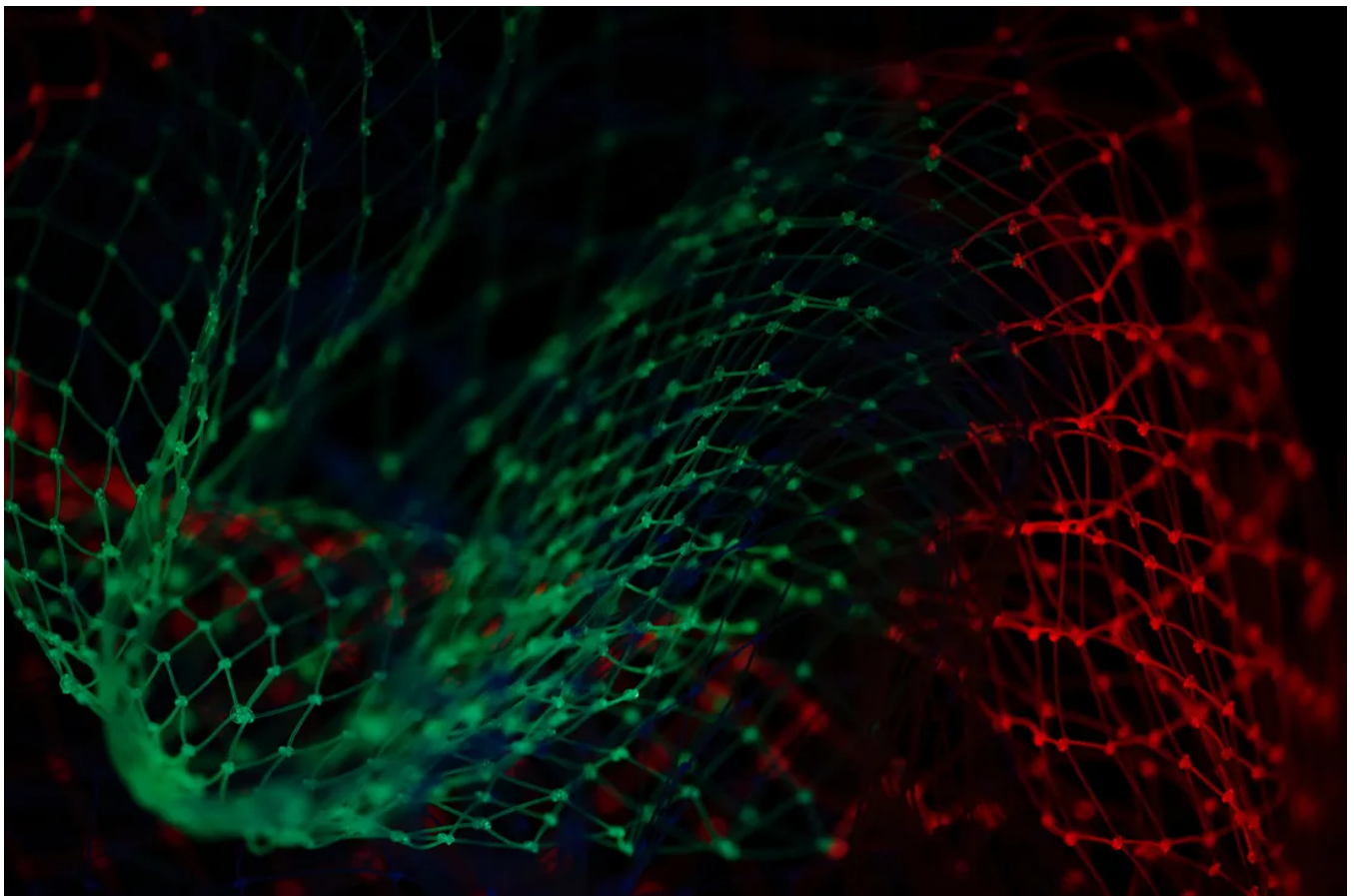


Listen



Share

... More



Source: [Pietro Jeng](#) via [Unsplash](#)

I will suggest that before this reading this article, please go through my other article which explains clustering in detail, because then you will have a strong base of the clustering.

--	--

What is Clustering?

Introduction: Ever thought of arranging data based on similar features without having the actual labels/classes/targets...

medium.com

Once completed reading the above article, now it is a good time to start reading this article.

Explanation of Hierarchical Clustering

This type of clustering starts to make clusters of data in **Tree format**.

Internally Hierarchical clustering try to construct a tree where each node can be considered as a cluster of its child nodes.

Hierarchical Clustering can be further divided into 2 categories which are:

1. Agglomerative
2. Divisive

Agglomerative Hierarchical Clustering

It is based on the bottom to top approach of tree formation.

This approach starts at the bottom of the tree & keep on constructing the cluster Tree upwards. It, first of all, finds the 2 or more features which are most similar and combine them as one by creating a node of the Cluster Tree, they can be found out let's say by the distance between them, let's say there are multiple points, then the 2 features having the minimum distance will be considered as the down-most node of the Tree.

By continuing the above-explained approach recursively, a complete tree is formed which can be easily interpreted & multiple insights can be easily extracted from that Tree.

Divisive Hierarchical Clustering

It is opposite to the Agglomerative approach & works oppositely as compared to the Agglomerative approach.

It is based on the top to bottom approach of cluster tree formation. First of all it creates a cluster of the items which are far apart, & then move downwards in the cluster tree & continues the above step until the complete tree is formed.

Generally most popular approach among the two is Agglomerative one, as it is more relatable to form clusters of the items first which are similar to each other.

Example of the Agglomerative Approach

In this example, I have shown the practical example of Hierarchical Clustering on fruits. Generally, the output of Hierarchical clustering is represented as a “Dendrogram”, which is a way to represent trees, or we can say diagrammatic representation of trees.

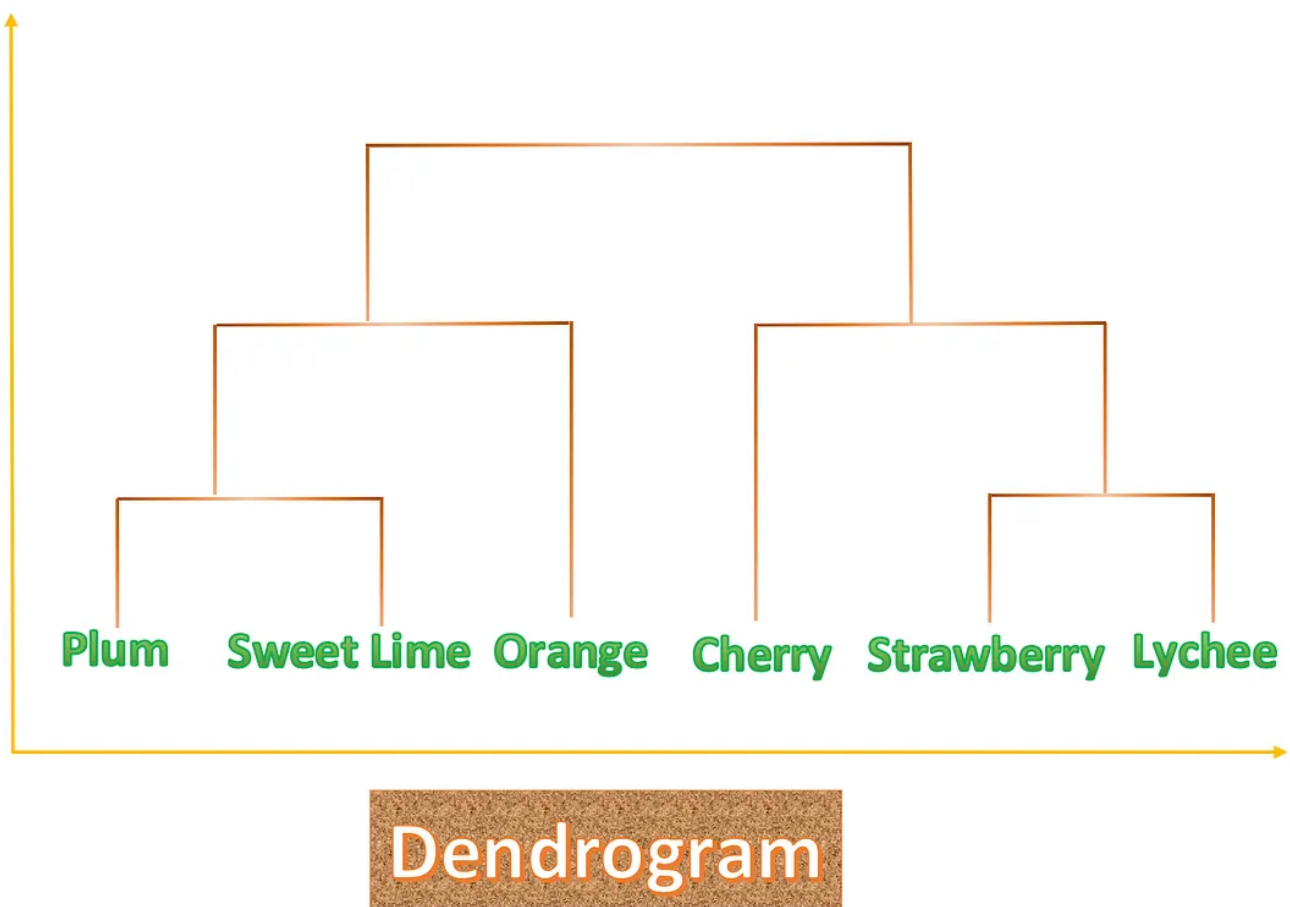


Image by Author

In the above figure, the “y” axis represents the similarity between the clusters, as the value of y increases, dissimilarity increases. “x” axis represents the clusters.

Explanation of the above Dendrogram

Clusters of pairs of (“Plum” and “Sweet Lime”) & (“Strawberry” and “Lychee”) are formed first (using agglomerative approach) because they are most similar.

Now, both of the above pairs can be treated as an individual entity, and the similarity of other fruits will be compared to them & further cluster tree grows. In the present example, pair of (“Plum” and “Sweet Lime”) is clustered with “Orange”, on the other hand, pair of (“Strawberry” and “Lychee”) is clustered with “Cherry”, based on the similarity between them.

In this fashion, hierarchical clustering executes and makes clusters of the data.

I hope my article explains each and everything related to hierarchical clustering along with the interpretation of the Dendrogram. Thank you so much for investing your time in reading my article and boosting your knowledge!

Machine Learning

Clustering

Data Science

Big Data

Deep Learning



Follow

Written by Harshit Dawar

346 Followers · Writer for Analytics Vidhya

AIOPS Engineer, have a demonstrated history of delivering large and complex projects. 14x Globally Certified. Rare & authentic content publisher.
