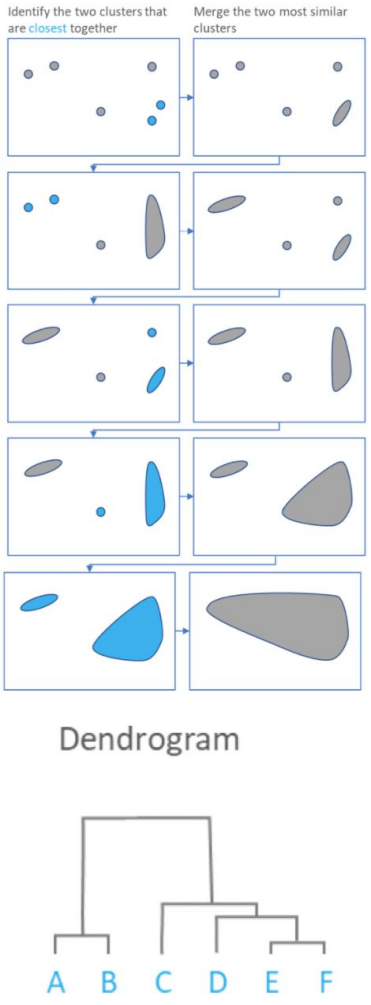
Classification vs Clustering

Classification belongs to supervised learning whereas Clustering can be seen as unsupervised learning. Clustering is a type of algorithm that seeks to provide information without any prior knowledge of output. Classification handles data that is input with labels and has prior knowledge of a possible output. Clustering can be seen in data analysis for retail and Classification is seen in fraud detection in banking.

Hierarchical clustering

Hierarchical clustering is an algorithm that groups data into clusters. Hierarchical clustering starts by treating each observation as a separate cluster, then repeats the same two steps over and over until it has gone through every cluster; (1) It identifies the two clusters that are the most similar and (2) brings them close together. This is a process that continues until every single cluster has been merged together as the illustration on the right shows. The main output for Hierarchical Clustering is a dendrogram which essentially shows the relationships between the clusters.

Hierarchical clustering is relatively easy to understand, the maths behind it is relatively simple and is also fairly straightforward to program, however it rarely provides the best solution. One interesting application of this technique however is tracking viruses through Phylogenetic trees. Viruses with high mutation rates can be tracked through this method, and allows for doctors to keep up with treatments.



Block. T. What is hierarchical clustering?

Reference

Block, T. ‘What is hierarchical clustering?’ *DisplayR*. Available at: https://www.displayr.com/what-is-hierarchical-clustering/ (Accessed 11/01/2021)