Agglomerative Clustering

Definition and purpose

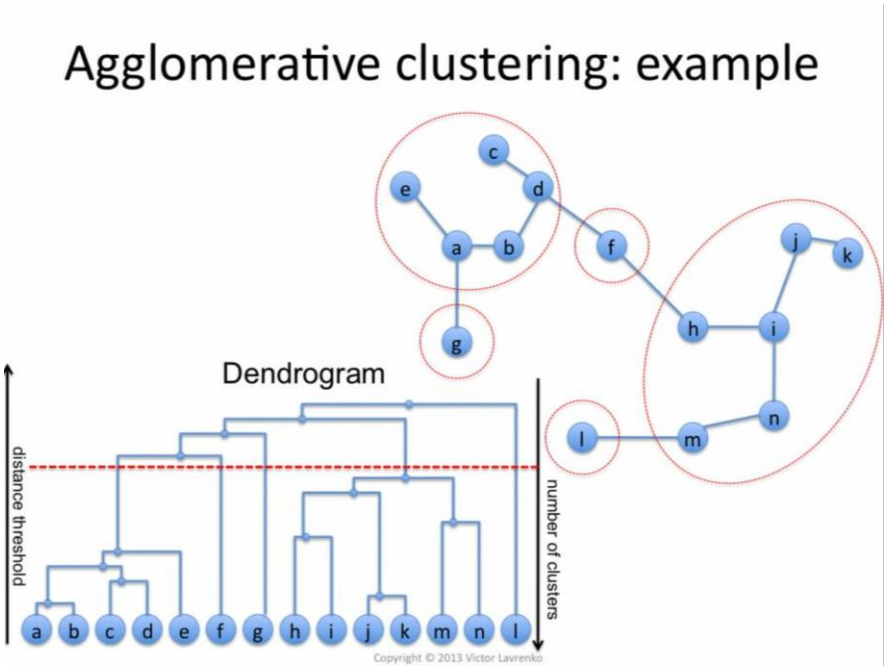
The agglomerative clustering is to ensure that nearby points end up in the same cluster, **is a "bottom-up" hierarchical clustering** approach (Divisive is a "top-down" approach).

Algorithm

1. Start with a collection C of singleton clusters, each cluster contains one data point: $c_i = \{x_i\}$
2. Repeat until only one cluster is left
    a. Find a pair of clusters that closest: $\min D(c_i, c_j)$
    b. Merge the cluster c_i, c_j into a new cluster c_{i+j}
    c. Remove c_i, c_j from the collection C, add c_{i+j}
3. For measuring the distance among clusters, we could use the following linkages

Single Linkage	This is the distance between the closest members of the two clusters.
Complete Linkage	This is the distance between the members that are farthest apart.
Average Linkage	This method involves looking at the distances between all pairs and averages all of these distances. This is also called Unweighted Pair Group Mean Averaging.

4. **Produces a dendrogram**: hierarchical tree of clusters (Leaf – individuals, Root-clusters)



Properties

1. Slow: $O(n^2 d + n^3)$
2. Complete linkage is sensitive to outliers.