

$$d(K, L) = \frac{|I| \cdot d(I, L) + |J| \cdot d(J, L)}{|I| + |J|}$$

linkMethod = weighted

There is no general expression for the dissimilarity between clusters as it depends on the order in which the clusters are merged. At each step, new dissimilarities are computed as:

$$d(K, L) = \frac{d(I, L) + d(J, L)}{2}$$

linkMethod = centroid

This method is suitable only for use with the Euclidean dissimilarity metric.

$$d(A, B) = \|\vec{c}_A - \vec{c}_B\|$$

where \vec{c} indicates the centroid of a cluster.

$$d(K, L) = \sqrt{\frac{|I| \cdot d(I, L)^2 + |J| \cdot d(J, L)^2}{|I| + |J|} - \frac{|I| \cdot |J| \cdot d(I, J)^2}{(|I| + |J|)^2}}$$

linkMethod = median

This method is suitable only for use with the Euclidean dissimilarity metric.

assigns $d(K, L)$ like the centroid method. When two clusters I and J are combined into a new cluster K , the average of centroids i and j give the new centroid k . Or,

$$d(K, L) = \sqrt{\frac{d(I, L)^2}{2} + \frac{d(J, L)^2}{2} - \frac{d(I, J)^2}{4}}$$

linkMethod = ward

This method is suitable only for use with the Euclidean dissimilarity metric.

$$d(A, B) = \sqrt{\frac{2|A||B|}{|A| + |B|}} \cdot \|\vec{c}_A - \vec{c}_B\|$$

The Ward variance minimization algorithm

$$d(K, L) = \sqrt{\frac{(|I| + |L|) \cdot d(I, L)^2 + (|J| + |L|) \cdot d(J, L)^2 - |L| \cdot d(I, J)^2}{|I| + |J| + |L|}}$$

Dendrogram Wave Format

The HCluster operation optionally produces a dendrogram output wave that can be used to create a dendrogram plot. This section describes the format of the dendrogram output wave.

The dendrogram wave contains all the information about the node pairs that are combined by HCluster, and the dissimilarity between the nodes. It also has a list of the indices into the original data in the order in which, for instance, labels should be drawn to avoid node connector lines that cross. The wave has four columns and N rows, where N is the number of original data vectors.