

Example Workflows

Hierarchical Clustering

The workflow clusters the data items in Iris data set by first estimating the distances between data instances. Distance matrix is passed to hierarchical clustering, which renders the dendrogram. Select different parts of the dendrogram to further analyze the corresponding data.

Notice that hierarchical clustering can only handle small data sets, that is, those that contain only a small number of data instances. For larger data sets the distance matrix may get too big and may not fit in the memory. An alternative method that can consider such data sets is k-means clustering.

