jHc: A Java desktop application for agglomerative hierarchical clustering

Manual

jHc is a standalone Java desktop application that allows the user to subject data to different approaches of agglomerative hierarchical clustering.

Data Input

Copy and paste the relevant data into the spreadsheet in the tab Data Input. To paste the clipboard contents, simply click on cell A1 and then use Ctrl-v to paste the data.

Any rows with non-numerical data (i.e. data not to be analysed) need to be marked accordingly. This can be done by clicking on the spreadsheet row number and then using a mouse right-click to display a pop-up menu from which the option Ignore row needs to be selected. Similarly, any columns that contain non-numerical data need to be marked. This is done by clicking on the spreadsheet column name and performing a mouse right-click; from the pop-up menu, select Ignore column. For the column that contains the labels of the individual data sets choose the option Mark as label.

Analysis

Hierarchical clustering

Data can be normalised prior to calculation of initial distances. The available normalisation options are:

- no normalisation
- MinMax
- Z-Score

The linkage approaches avilable are:

- Average linkage
- Complete linkage
- Single linkage
- Ward's method
- Weighted linkage

Distances can be calculated as

- Euclidean
- Manhattan
- Supremum

When choosing Ward's method, distances are automatically calculated as Euclidean distances.

Dendrogram

The resultant dendrogram is automatically displayed in the panel Dendrogram. If an Output directory is specified in the section Output, an image file with the dendrogram will be written based on the specifications in this section.

Output

Data and log files are automatically written to the user-specified Output directory. If no Output directory is given, no output files will be generated. You can also specify an Output prefix to be added in front of all output files.

Dendrogram

Shows the resultant dendrogram after performing a hierarchical clustering analysis.

Log

Shows the log messages of the hierarchical clustering analysis.