

Package ‘VisuClust’

February 17, 2016

Type Package

Title Visualisation of Clusters in Multivariate Data

Version 1.2

Date 2016-2-15

Author

Michael Sieger [cre, aut], Georg Ohmayer [aut, ths], Hochschule Weihenstephan-Triesdorf [cph]

Suggests MASS, cluster

Depends aplpack

Maintainer Michael Sieger <sieger-michael@web.de>

Description Displays multivariate data, based on Sammon's nonlinear mapping.

License LGPL (>= 3)

NeedsCompilation no

Repository CRAN

Date/Publication 2016-02-17 11:19:16

R topics documented:

FuzzyPlot	2
LinkageMap	3
MilchBig	5
MilchSmall	5
VisuClust	6

Index	7
--------------	----------

FuzzyPlot

*Displays a Fuzzy Plot***Description**

A Fuzzy Plot is a 2D Scatter Plot that displays the result of Sammon's Nonlinear Mapping together with a fuzzy clustering. The values $u[i,k]$ ($i=1,\dots,nPoints$; $k=1,\dots,nPoints$), which states the membership of the i .th observation for the cluster k , can be computed for example with [fanny](#) from the package `cluster`.

If the slider is on the last position, all cluster are displayed together. The color and the symbol indicates the nearest crisp clustering then ($\max u[i,k]$, $k=1,\dots,nClusters$). The color intensity displays the probability of membership to the nearest cluster.

A single cluster can be selected too by setting the slider in position 1 to $nClusters$. The color intensity and the label size (if labels are set) shows the probability of membership to the selected cluster.

Usage

```
FuzzyPlot(xSammon, probs, clusterColors=rainbow(dim(probs)[2]),
clusterSymbols=rep(21,dim(probs)[2]), labels=NULL,
labelSize=c(0.6, 1.0), xlab="", ylab="", main="",
enableLegend=TRUE, cex=c(0.7, 1.4))
```

Arguments

<code>xSammon</code>	A matrix with size $(nPoints,2)$ that contains the projected points from Sammon's Nonlinear Mapping. See sammon (package <code>MASS</code>).
<code>probs</code>	A matrix with size $(nPoints, nClusters)$ that contains the probabilities of membership for each cluster. See the membership argument of fanny (package <code>cluster</code>).
<code>clusterColors</code>	A vector of size $nClusters$ that contains the colors for the clusters. One should only choose colors with a very high intensity. See colors .
<code>clusterSymbols</code>	A vector of size $nClusters$ that contains the symbols for the clusters. Not all symbols can be filled. If this is wanted, one should use the symbols 15-20. See points .
<code>labels</code>	A string-array with labels for the points (size: $nPoints$).
<code>labelSize</code>	The size of the labels in the form $c(\min, \max)$.
<code>xlab</code>	A title for the x axis (as described in plot).
<code>ylab</code>	A title for the y axis (as described in plot).
<code>main</code>	An overall title for the plot (as described in plot).
<code>enableLegend</code>	Switch legend on/off.
<code>cex</code>	Two numerical values in the form $c(\min, \max)$ giving the amount by which plotting text and symbols should be magnified relative to the default (as described in par).

Author(s)

Michael Sieger <sieger-michael@web.de>
 Georg Ohmayer <georg.ohmayer@hswt.de>

Examples

```
library(cluster)
library(MASS)

# The example data
data("MilchSmall")

M <- scale(MilchSmall[3:6])
D <- dist(M)
S <- sammon(D)
F <- fanny(D, 6)

FuzzyPlot(S$points, F$membership, labels=MilchSmall[,2],
clusterColors=c("red", "green", "magenta", "yellow", "blue", "black"))
```

 LinkageMap

Displays a Linkage Map

Description

A Linkage Map is a 2D Scatter Plot that displays the result of Sammon's Nonlinear Mapping. A number of ranges (≥ 1) for the distance values $d[i,j]$ can be defined. Each point pair whose distance is inside of the defined range is connected with a line.

The ranges are defined by thresholds $t[k]$ ($1 \leq k \leq 3$) in the following way range[1]: $d[i,j] \leq t[1]$ range[k]: $t[k-1] < d[i,j] \leq t[k]$ The ranges can be changed dynamically at runtime by changing the thresholds $t[k]$ with sliders.

A estimation of the density function for the distances $d[i,j]$ (for example with kernel density estimation, see function [density](#)) is helpful for finding the ideal thresholds $t[k]$ (=local minima in the density function).

Usage

```
LinkageMap(xSammon, dist, lineTypes=c("solid","dotted", "dashed"),
lineColors=c("red","green","blue"), lineWidths=c(1,1,1),
labels = NULL, cluster = NULL, maxValue=0.5, legendDigits = 2,
xlab = "", ylab = "", main = "")
```

Arguments

<code>xSammon</code>	A matrix with size (nPoints,2) that contains the projected points from Sammon's Nonlinear Mapping.
<code>dist</code>	A distance matrix for the given data (to compute with <code>dist</code>).
<code>lineTypes</code>	An array of line types. The size must match with line colors and line widths. The various line types are described in <code>par</code> .
<code>lineColors</code>	An array of colors. The size must match with line types and line widths. see <code>colors</code>
<code>lineWidths</code>	An array of line widths. The size must match with line colors and line types. see <code>par</code>
<code>labels</code>	A string-array with labels for the points (size: nPoints).
<code>cluster</code>	A array containing cluster memberships of the points. The point membership will be indicated with different colors. The array size must match with nPoints.
<code>maxValue</code>	<code>maxValue*max(d[i,j])</code> is the maximum value that can be adjusted with the sliders.
<code>legendDigits</code>	The number of fractional digits to be displayed in the legend.
<code>xlab</code>	A title for the x axis (as described in plot).
<code>ylab</code>	A title for the y axis (as described in plot).
<code>main</code>	An overall title for the plot (as described in plot).

Author(s)

Michael Sieger <sieger-michael@web.de>
 Georg Ohmayer <georg.ohmayer@hswt.de>

Examples

```
library(MASS)

data("MilchSmall")
M <- scale(MilchSmall[3:6])

D <- dist(M)
S <- sammon(D)
K <- kmeans(D,center=5)
LinkageMap(S$points, D, cluster=K$cluster, labels=MilchSmall[,2])
```

MilchBig

Milk components of 50 mammals

Description

This dataset contains the fat, protein, lactose and ash of different milks.

Usage

MilchBig

Author(s)

Georg Ohmayer <georg.ohmayer@hswt.de>,
Herbert Seiler

References

Ohmayer G., Seiler H. (1985): Numerische Gruppierung und graphische Darstellung von Daten: Ein Methodenvergleich; EDV in Medizin und Biologie, Vol. 2, S. 65-73

MilchSmall

Milk components of mammals

Description

This dataset contains the fat, protein, lactose and ash of different milks. This is a reduced version (20 instead of 50 observations).

Usage

MilchSmall

Author(s)

Georg Ohmayer <georg.ohmayer@hswt.de>,
Herbert Seiler

References

Ohmayer G., Seiler H. (1985): Numerische Gruppierung und graphische Darstellung von Daten: Ein Methodenvergleich; EDV in Medizin und Biologie, Vol. 2, S. 65-73

Description

The VisuClust package provides functions to display the result of clusterings, if they are either disjoint (every observation belongs to one cluster) or fuzzy (for each observation i and cluster k there is a value $u[i,k]$ ($i=1,\dots,nPoints$; $k=1,\dots,nClusters$)), which states the membership of the i .th observation for the cluster k , where $\sum(u[i,k],k=1,\dots,nClusters) = 1$). A disjoint clustering can be created with known hierarchical methods, if the desired number of clusters is provided, or by using k -means-clustering (for example [hclust](#), [kmeans](#) from the package [cluster](#)). A fuzzy clustering can be created for example with [fanny](#) from the package [cluster](#).

The basis of the visualization is Sammon's method Nonlinear Mapping (NLM), which displays all observations in a plane so that the distances in the multivariate space are approximated best (see [sammon](#) from the package [MASS](#)). The clusters in the NLM diagram can be indicated by coloring, drawing lines between similar point pairs, and other features.

Author(s)

Michael Sieger <sieger-michael@web.de>
Georg Ohmayer <georg.ohmayer@hswt.de>

References

Will be published in the GIL conference transcript 2012: Visualisierung von Clustern in multivariaten Daten unter Einsatz von R Georg Ohmayer, Michael Sieger

Index

*Topic **datasets**

MilchBig, [5](#)

MilchSmall, [5](#)

cluster, [2, 6](#)

colors, [2, 4](#)

density, [3](#)

dist, [4](#)

fanny, [2](#)

FuzzyPlot, [2](#)

hclust, [6](#)

kmeans, [6](#)

LinkageMap, [3](#)

MilchBig, [5](#)

MilchSmall, [5](#)

par, [2, 4](#)

plot, [2](#)

points, [2](#)

sammon, [2, 6](#)

VisuClust, [6](#)