Package ‘ClusBoot’

January 20, 2025

Title Bootstrap a Clustering Solution to Establish the Stability of the Clusters

Version 1.2.2

Description Providing a cluster allocation for n samples, either with an $n \times p$ data matrix or an $n \times n$ distance
matrix, a bootstrap procedure is performed. The proportion of bootstrap replicates where a pair of samples
cluster in the same cluster indicates who tightly the samples in a particular cluster clusters together.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.2.3

Depends R (>= 2.10)

LazyData true

Imports graphics, grDevices, stats

Suggests fpc, testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Sugnet Lubbe [aut, cre, cph] (<`https://orcid.org/0000-0003-2762-9944`>)

Maintainer Sugnet Lubbe <`slubbe@sun.ac.za`>

Repository CRAN

Date/Publication 2024-07-08 11:30:02 UTC

Contents

`boot.proportions` *Heatmap of the proportion of bootstrap replicates where objects cluster together*

Description

Heatmap of the proportion of bootstrap replicates where objects cluster together

Usage

```
boot.proportions(
  x,
  col = grDevices::heat.colors(101, rev = TRUE),
  show.vals = F,
  text.col = "black",
  cluster.col = "firebrick",
  ...
)
```

Arguments

`x`	an object of class clusboot
`col`	vector of colours for shading to indicate proportion values
`show.vals`	logical value indicating whether proportion values should be added to individual cells
`text.col`	colour of text for show.vals if `TRUE`
`cluster.col`	colour of lines demarcating cluster membership
`...`	more arguments to be passed to `plot()`

Examples

```
out <- clusboot (scale(case.study.psychiatrist), B=100, k=6, clustering.func=complete.linkage)
boot.proportions(out)
```

`boot.silhouette` *Produces silhouette plots*

Description

Produces silhouette plots

Usage

```
boot.silhouette(clusboot.out, ...)
```

Arguments

`clusboot.out`	an object of class clusboot
`...`	more arguments to be passed to `barplot()`

Value

list of silhouette widths

Examples

```
out <- clusboot (scale(case.study.psychiatrist), B=100, k=6, clustering.func=complete.linkage)
boot.silhouette(out)
```

`calc.silhouette` *Computes the silhouette vales based on proportion of times items cluster together*

Description

Computes the silhouette vales based on proportion of times items cluster together

Usage

```
calc.silhouette(clusboot.out)
```

Arguments

`clusboot.out`	an object of class clusboot

Value

an object of class clusboot

`case.study.psychiatrist`

Patient by psychiatric symptom data

Description

Presence/absence ratings of 24 psychiatric symptoms in 30 psychiatric inpatients made by an individual psychiatrist. The data have been collected in a case study of an individual psychiatrist to identify his implicit taxonomy.

Usage

```
case.study.psychiatrist
```

Format

`case.study.psychiatrist`**:**

A data frame with 30 observations on the following 28 variables:

`V1` inappropriate affect, appearance or behavior; binary vector

`V2` interview belligerence - negativism; binary vector

`V3` agitation - excitement; binary vector

`V4` retardation; binary vector

`V5` lack of emotions; binary vector

`V6` speech disorganization; binary vector

`V7` grandiosity; binary vector

`V8` suspicion - ideas of persecution; binary vector

`V9` hallucinations - delusions; binary vector

`V10` overt anger; binary vector

`V11` depression; binary vector

`V12` anxiety; binary vector

`V13` obsession - compulsion; binary vector

`V14` suicide; binary vector

`V15` self injury; binary vector

`V16` somatic concerns; binary vector

`V17` social isolation; binary vector

`V18` daily routine impairment; binary vector

`V19` leisure time impairment; binary vector

`V20` antisocial impulses or acts; binary vector

`V21` alcohol abuse; binary vector

`V22` drug abuse; binary vector

`V23` disorientation; binary vector

`V24` memory impairment; binary vector

`V25`	rating on Global Assessment Scale, a 101-point scale for overall severity of psychiatric disturbance; a numeric vector
`V26`	Affective (Affective Disorder or Anxiety Disorder); binary vector
`V27`	Psychotic (Schizophrenic Disorder or Paranoid Disorder); binary vector
`V28`	Substance abuse (Substance Use Disorder or Substance-Induced Disorder); binary vector

Details

The data set forms part of the International Federation of Classification Societies Cluster Benchmark Data Repository

Source

Van Mechelen, I., & De Boeck, P. (1989). Implicit taxonomy in psychiatric diagnosis: A case study. Journal of Social and Clinical Psychology, 8, 276-287. `https://ifcs.boku.ac.at/repository/data/case_study_psychiatrist/index.html`

`clusboot` *Performs bootstrap on a cluster analysis output*

Description

Performs bootstrap on a cluster analysis output

Usage

```
clusboot(datmat, B = 1000, clustering.func = complete.linkage, ...)
```

Arguments

`datmat`	a data matrix or distance object which will be the input to the clustering function
`B`	number of bootstrap replicates
`clustering.func`	the function which will perform the clustering and output a vector of cluster memberships
`...`	more arguments to be passed to the clustering function

Details

Any R function performing cluster analysis can be specified in `clustering.func` although a wrapper function is typically needed to isolate only the vector output of cluster memberships. See `?complete.linkage` as an example. Should users perfer to use alternative resamling schemes, other than the bootstrap, Hennig (2007) discuss a variety of options which could be accessed by specifying `clustering.func` = `fpc.clusterboot`. In addition, the sampling method is specified in the argument `bootmethod` and additional arguments for the function `clusterboot` in the package `fpc` must be given. Note that only the resampling facilities of `clusterboot` is utilised while the computation of proportions and silhouette widths remain unchanged. The output object of class `clusboot` will remain unchanged as only the resampling section of `clusterboot` is used.

Value

an object of class `clusboot` which is a list with the following components:

`proportions`	matrix of size nxn with cell ij containing the proportion of bootstrap replicates in which object i and object j clustered together.
`clustering`	a vector of length n containing the cluster membership of the n input objects.
`sil`	a vector of length the number of clusters containing the bootstrap-silhouette values for the clusters.
`indv.sil`	a vector of length n containing the bootstrap-silhouette values for the individual objects.
`sil.order`	a vector of length n containing the ordering of the n objects used by the functions `boot.silhouette` and `boot.proportions` to order objects in the same cluster adjacent and clusters in decreasing order of cluster tightness.
`ave.sil.width`	the overall stability of the clustering solution, obtained by averaging over the individual object bootstrap-silhouette values.

References

Hennig, C., 2007. Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis, 52(1), pp.258-271. Lubbe, S., 2024. Bootstrapping Cluster Analysis Solutions with the R Package ClusBoot. Austrian Journal of Statistics, 53(3), pp.1-19.

Examples

```
clusboot (scale(case.study.psychiatrist), B=100, k=6, clustering.func=complete.linkage)
library(fpc)
clusboot (scale(case.study.psychiatrist), B=100, k=6, clustering.func=fpc.clusterboot,
          clustermethod=hclustCBI, method="complete", bootmethod="subset", subtuning=10)
```

`complete.linkage` *Wrapper function for performing complete linkage clustering*

Description

Wrapper function for performing complete linkage clustering

Usage

```
complete.linkage(X, k)
```

Arguments

`X`	samples x variables data matrix
`k`	number of clusters

Value

vector of cluster memberships

Examples

```
complete.linkage(scale(case.study.psychiatrist), k=6)
```

`fpc.clusterboot` *Resampling according to the methods discussed in Hennig (2007)*

Description

Resampling according to the methods discussed in Hennig (2007)

Usage

```
fpc.clusterboot(
  data,
  B,
  distances = (inherits(data, "dist")),
  bootmethod = "boot",
  bscompare = TRUE,
  multipleboot = FALSE,
  jittertuning = 0.05,
  noisetuning = c(0.05, 4),
  subtuning = floor(nrow(data)/2),
  clustermethod,
  noisemethod = FALSE,
  count = TRUE,
  seed = NULL,
  datatomatrix = TRUE,
  ...
)
```

Arguments

`data`	a data matrix or distance object which will be the input to the clustering function
`B`	number of bootstrap replicates
`distances`	see `?fpc::clusterboot`
`bootmethod`	see `?fpc::clusterboot`
`bscompare`	see `?fpc::clusterboot`
`multipleboot`	see `?fpc::clusterboot`
`jittertuning`	see `?fpc::clusterboot`
`noisetuning`	see `?fpc::clusterboot`

`subtuning`	see `?fpc::clusterboot`
`clustermethod`	see `?fpc::clusterboot`
`noisemethod`	see `?fpc::clusterboot`
`count`	see `?fpc::clusterboot`
`seed`	see `?fpc::clusterboot`
`datatomatrix`	see `?fpc::clusterboot`
`...`	additional arguments to be sent to the function specified in clustermethod

Value

a list with two components; boot.out contains the computations for clusboot and out contains the clustering solution of the original data set

---

`plot.clusboot` *MDS plot of similarities given by the proportion of bootstrap replicates where objects cluster together*

Description

MDS plot of similarities given by the proportion of bootstrap replicates where objects cluster together

Usage

```
## S3 method for class 'clusboot'
plot(x, col, show.silhouette = TRUE, ...)
```

Arguments

`x`	an object of class clusboot
`col`	single colour or a vector specifying a colour for each object
`show.silhouette`	logical indicating whether plotting character size should represent the inidivual silhouette values
`...`	more arguments to be passed to `plot()`

Value

matrix of similarities (proportions)

Examples

```
out <- clusboot (scale(case.study.psychiatrist), B=100, k=6, clustering.func=complete.linkage)
plot(out)
```

Index