

Package ‘ClustVarLV’

May 28, 2022

Title Clustering of Variables Around Latent Variables

Version 2.1.1

Author Evelyne Vigneau [aut, cre], Mingkun Chen [ctb], Veronique Cariou [aut]

Maintainer Evelyne Vigneau <evelyne.vigneau@oniris-nantes.fr>

Description Functions for the clustering of variables around Latent Variables, for 2-way or 3-way data. Each cluster of variables, which may be defined as a local or directional cluster, is associated with a latent variable. External variables measured on the same observations or/and additional information on the variables can be taken into account. A “noise” cluster or sparse latent variables can also be defined.

Depends R (>= 4.1.0)

License GPL-3

Encoding UTF-8

LazyData TRUE

Imports Rcpp, doParallel, foreach, parallel, iterators, plyr

LinkingTo Rcpp, RcppEigen

Suggests knitr, rmarkdown

VignetteBuilder knitr

RoxygenNote 7.2.0

NeedsCompilation yes

Repository CRAN

Date/Publication 2022-05-28 15:20:01 UTC

R topics documented:

apples_sh	2
AUPA_pscho	3
authen_NMR	4
block.scale	4
boot_clv	5
ciders	6

CLV	7
CLV3W	9
CLV3W_kmeans	11
CLV_kmeans	13
coffee	15
data_biplot	15
get_comp	16
get_loading	17
get_partition	17
get_weight	18
imput_clv	19
LCLV	20
lm_CLV	21
plot.clv	23
plot.clv3w	24
plot.lclv	24
plot_var	25
plot_var.clv3w	26
predict.lmclv	27
print.clv	28
print.clv3w	28
print.lclv	29
stand_quali	29
summary.clv	30
summary.clv3w	30

Index	32
--------------	-----------

apples_sh	<i>apples from southern hemisphere data set</i>
-----------	---

Description

Sensory characterization and consumers preference for 12 varieties of apples.

Usage

```
data(apples_sh)
```

Format

A data frame with 12 observations and 2 blocks of variables.

senso 43 sensory attributes

pref hedonic scores given by a panel of 60 consumers

References

Dailliant-Spinnler, B, MacFie, H.J.H, Beyts, P.K., Hedderley, D. (1996). Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the southern hemisphere. *Food Quality and Preference*, 7(2), 113-126.

Examples

```
data(apples_sh)
names(apples_sh)
apples_sh$senso
apples_sh$pref
```

AUPA_psych

Psychological eating behavior data set

Description

The psychological behaviour items in this dataset is a part of French Research Project (AUPALE-SENS, 2010-2013) dealing with food behaviour and nutritional status of elderly people. There are 31 psychological items organised into five blocks, each aiming to describe a given behavioural characteristic: emotional eating (E) with six items, external eating (X) with five items, restricted eating (R) with five items, pleasure for food (P) with five items, and self esteem (S) with ten items. Detailed description and analysis of the emotional, external and restricted eating items for this study are available in Bailly, Maitre, Amand, Herve, and Alaphilippe (2012). 559 subjects were considered.

Usage

```
data(AUPA_psych)
```

Format

A data frame with 559 observations, (row names from 1 to 559) and 31 items. The name of the items refers to the corresponding block (E, X, R, P, S).

References

Bailly N, Maitre I, Amand M, Herve C, Alaphilippe D (2012). The Dutch Eating Behaviour Questionnaire(DEBQ). Assessment of eating behaviour in an aging French population. *Appetite*, 59(853-858).

Examples

```
X = data(AUPA_psych)
```

 authen_NMR

Authentication data set/ NMR spectra

Description

Discrimination between authentic and adulterated juices using ^1H NMR spectroscopy. 150 samples were prepared by varying the percentage of co-fruit mixed with the fruit juice of interest. The two first characters in the row names represent this percentage. Authentic juice names begin with "00". Samples prepared with the co-fruit alone are identified by "99" (rather than 100).

Usage

```
data(authen_NMR)
```

Format

150 observations and 2 blocks of variables.

authen_NMR\$Xz1 spectral range from 6 to 9 ppm (300 variables)

authen_NMR\$Xz2 spectral range from 0.5 to 2.3 ppm (180 variables)

References

Vigneau E, Thomas F (2012). Model calibration and feature selection for orange juice authentication by ^1H NMR spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, 117, 22:30.

Examples

```
data(authen_NMR)
xlab=as.numeric(colnames(authen_NMR$Xz2))
plot(xlab, authen_NMR$Xz2[1,], type="l", xlab="ppm", ylab="", ylim=c(14.8,15.8),
xlim=rev(range(xlab)))
for (i in (1:nrow(authen_NMR$Xz2))) lines(xlab,authen_NMR$Xz2[i,])
```

 block.scale

Scaling of a three-way array

Description

Centering and scaling of a three-way array

Usage

```
block.scale(X, xcenter = TRUE, xscale = 0)
```

Arguments

X	: a three-way array
xcenter	: centering of X. By default X will be centered for both mode 2 and 3 (xcenter=TRUE), otherwise xcenter=FALSE
xscale	: scaling parameter applied to X. By default no scaling (xscale=0) 0 : no scaling only centering - the default 1 : scaling with standard deviation of (mode 2 x mode 3) elements 2 : global scaling (each block i.e. each mode 2 slice will have the same inertia) 3 : global scaling (each block i.e. each mode 3 slice will have the same inertia)

Value

Xscaled : the scaled three-way array

 boot_clv

Bootstrapping for assessing the stability of a CLV result

Description

Bootstrapping on the individuals (in row) or the variables (in column) is performed. Choose the "row" option (the default), if the variables are measured on a random sample of individuals, Choose the "column" option, if the variables are taken from a population of variables. The first case is the more usual, but the second may occur, e.g. when variables are consumers assessing specific products. Each bootstrapped data matrix is submitted to CLV in order to get partitions from 1 to nmax clusters. For each number of clusters, K, the Rand Index, the adjusted Rand Index, as well as the cohesion and the isolation of the clusters of the observed partition and the bootstrapped partitions are computed. These criteria are used for assessing the stability of the solution into K clusters. Parallel computing is performed for time saving.

Usage

```
boot_clv(object, case = "row", B = 100, nmax = NULL)
```

Arguments

object	: result of CLV()
case	: "row" or "column" corresponding to the random effect of the design
B	: the number of bootstrap to be run (100 by default)
nmax	: maximal size of the partitions to be considered (if NULL, the value of nmax used for the object is used)

Value

res a list of length 4 for the Rand Index, Adjusted Rand Index, Cohesion and Isolation of the partition results (matrix of size (B x nmax), respectively).

See Also

CLV

`ciders`*ciders data*

Description

Case study pertaining to Quantitative Descriptive Analysis (QDA) applied to ten varieties of cider. The sensory panel consists of seven trained assessors who were asked to rate ten varieties of cider using a list of ten sensory attributes.

Usage`data(ciders)`**Format**

An object of class "array" with 10 ciders (mode 1), 10 sensory attributes (mode 2) and 7 assessors (mode 3):

ciders 1 to 10

sensory attributes sweet, acid, bitter, astringency, odor strength, pungent, alcohol, perfume, intensity, and fruity

Panel Judge.1 to Judge.7

References

Ledauphin, S., Hanafi, M., & Qannari, E. M. (2006). Assessment of the agreement among the subjects in fixed vocabulary profiling. *Food quality and preference*, 17(3-4), 277-280.

Examples

```
data(ciders)
str(ciders)
```

Description

Hierarchical Cluster Analysis of a set of variables with consolidation. Directional or local groups may be defined. Each group of variables is associated with a latent component. Moreover, the latent component may be constrained using external information collected on the observations or on the variables.

Usage

```
CLV(
  X,
  Xu = NULL,
  Xr = NULL,
  method = NULL,
  sX = TRUE,
  sXr = FALSE,
  sXu = FALSE,
  nmax = 20,
  maxiter = 20,
  graph = TRUE
)
```

Arguments

X	: The matrix of variables to be clustered
Xu	: The external variables associated with the columns of X
Xr	: The external variables associated with the rows of X
method	: The criterion to be use in the cluster analysis. 1 or "directional" : the squared covariance is used as a measure of proximity (directional groups). 2 or "local" : the covariance is used as a measure of proximity (local groups)
sX	,TRUE/FALSE : standardization or not of the columns X (TRUE by default) (predefined -> cX = TRUE : column-centering of X)
sXr	,TRUE/FALSE : standardization or not of the columns Xr (FALSE by default) (predefined -> cXr = TRUE : column-centering of Xr)
sXu	,TRUE/FALSE : standardization or not of the columns Xu (FALSE by default) (predefined -> cXu= FALSE : no centering, Xu considered as a weight matrix)
nmax	: maximum number of partitions for which the consolidation will be done (by default nmax=20)
maxiter	: maximum number of iterations allowed for the consolidation/partitioning algorithm (by default maxiter=20)

graph, TRUE/FALSE (by default TRUE) : dendrogram and variation of the optimization criterion.
These plots can also be obtained with "plot"

Details

If external variables are used, define either X_r or X_u , but not both. Use the LCLV function when X_r and X_u are simultaneously provided.

Value

tabres Results of the clustering algorithm. In each line you find the results of one specific step of the hierarchical clustering.

- Columns 1 and 2 : The numbers of the two groups which are merged
- Column 3 : Name of the new cluster
- Column 4 : The value of the aggregation criterion for the Hierarchical Ascendant Clustering (HAC)
- Column 5 : The value of the clustering criterion for the HAC
- Column 6 : The percentage of the explained initial criterion value (method 1 => % var. expl. by the latent comp.)
- Column 7 : The value of the clustering criterion after consolidation
- Column 8 : The percentage of the explained initial criterion value after consolidation
- Column 9 : The number of iterations in the partitioning algorithm.
Remark : A zero in columns 7 to 9 indicates that no consolidation was done

partition K contains a list for each number of clusters of the partition, $K=2$ to n_{max} with

- clusters : in line 1, the groups membership before consolidation; in line 2 the groups membership after consolidation
- comp : The latent components of the clusters (after consolidation)
- loading : if there are external variables X_r or X_u : The loadings of the external variables (after consolidation)

References

Vigneau E., Qannari E.M. (2003). Clustering of variables around latents components. *Comm. Stat*, 32(4), 1131-1150.

Vigneau E., Chen M., Qannari E.M. (2015). ClustVarLV: An R Package for the clustering of Variables around Latent Variables. *The R Journal*, 7(2), 134-148

See Also

CLV_kmeans, LCLV

Examples

```

data(apples_sh)
#directional groups
resclvX <- CLV(X = apples_sh$senso, method = "directional", sX = TRUE)
plot(resclvX,type="dendrogram")
plot(resclvX,type="delta")
#local groups with external variables Xr
resclvYX <- CLV(X = apples_sh$pref, Xr = apples_sh$senso, method = "local", sX = FALSE, sXr = TRUE)

```

CLV3W	<i>Hierarchical clustering of variables (associated with mode 2 three-way array) with consolidation</i>
-------	---

Description

Hierarchical Cluster Analysis of a set of variables (mode 2) given a three-way array with a further consolidation step. Each group of variables is associated with a one-rank PARAFAC model (comp x loading x weight). Moreover, a Non Negativity (NN) constraint may be added to the model, so that the loading coefficients have positive values. Return an object of class `clv3w`.

Usage

```
CLV3W(X,mode.scale=0,NN=FALSE,moddendoinertie=TRUE,gmax=20,graph=TRUE,cp.rand=10)
```

Arguments

<code>X</code>	: a three way array - variables of mode 2 will be clustered
<code>mode.scale</code>	: scaling parameter applied to X, by default centering of X (for mode 2 x mode 3) is done. By default no scaling (<code>mode.scale=0</code>) 0 : no scaling only centering - the default 1 : scaling with standard deviation of (mode 2 x mode 3) elements 2 : global scaling (each block i.e. each mode 2 slice will have the same inertia) 3 : global scaling (each block i.e. each mode 3 slice will have the same inertia)
<code>NN</code>	: non Negativity constraint to be added on the loading coefficients. By default no constraint (<code>NN=FALSE</code>) TRUE : a non negativity constrained is applied on the loading coefficients to set them as positive values FALSE : loading coefficients may be either positive or negative
<code>moddendoinertie</code>	: dendrogram. By default it is based on the delta clustering criterion (<code>moddendoinertie =TRUE</code>) TRUE : dendrogram associated with the clustering criterion delta FALSE : dendrogram associated with the the height (cumulative delta)
<code>gmax</code>	: maximum number of partitions for which the consolidation will be done (default : <code>gmax=11</code>)

- graph : boolean, if TRUE, the graphs associated with the dendrogram and the evolution of the aggregation criterion are displayed (default : graph=TRUE)
- cp.rand : number of random starts associated with the one rank Candecomp/Parafac model (By default cp.rand=10)

Value

- tabres Results of the hierarchical clustering algorithm. In each line you find the results of one specific step of the hierarchical clustering.
- Columns 1 and 2 : the numbers of the two groups which are merged
 - Column 3 : name of the new cluster
 - Column 4 : the value of the aggregation criterion for the Hierarchical Ascendant Clustering (delta) : delta loss
 - Column 5 : the loss value of the clustering criterion for the HAC
 - Column 6 : the percentage of explained inertia of the data array X
 - Column 7 : the loss value of the clustering criterion after consolidation
 - Column 8 : the percentage of explained inertia of the data array X after consolidation
 - Column 9 : number of iterations in the partitioning algorithm.
- Remark : A zero in columns 7 to 9 indicates that no consolidation was done
- hclust contains the results of the HCA
- partition K contains a list for each number of clusters of the partition, K=1 to gmax with
- clusters : in line 1, the groups membership before consolidation; in line 2 the groups membership after consolidation
 - comp : the latent components of the clusters associated with the first mode (after consolidation)
 - loading : the vector of loadings associated with the second mode by cluster (after consolidation)
 - weigth : the vector of weights associated with the third mode by cluster (after consolidation)
 - criterion : vector of loss giving for each cluster the residual amount between the sub-array and its reconstitution associated with the cluster one rank PARAFAC model (after consolidation)
- param contains the clustering parameters
- gmax : maximum number of partitions for which the consolidation has been done
 - X : the scaled three-way array
- call : call of the method

Author(s)

Veronique Cariou, <veronique.cariou@oniris-nantes.fr>

References

Wilderjans, T. F., & Cariou, V. (2016). CLV3W: A clustering around latent variables approach to detect panel disagreement in three-way conventional sensory profiling data. *Food quality and preference*, 47, 45-53.

Cariou, V., & Wilderjans, T. F. (2018). Consumer segmentation in multi-attribute product evaluation by means of non-negatively constrained CLV3W. *Food Quality and Preference*, 67, 18-26.

See Also

CLV3W_kmeans, get_comp, get_loading, get_partition, plot, plot_var.clv3w,

Examples

```
data(ciders)
## Cluster Analysis of cider sensory descriptors with block scaling
## to set the assessors to the same footing
res.cider<-CLV3W(ciders,mode.scale=3,NN=FALSE,moddendoinertie=FALSE,gmax=20,graph=FALSE,cp.rand=5)
plot(res.cider,type="delta")
plot(res.cider,type="dendrogram")
print(res.cider)
summary(res.cider,2)
get_comp(res.cider,2)
get_loading(res.cider,2)
get_weight(res.cider,2)
```

CLV3W_kmeans	<i>Partitioning algorithm of a set of variables (associated with mode 2) oh a three-way array</i>
--------------	---

Description

Each group of variables is associated with a one-rank PARAFAC model (comp x loading x weight). Moreover, a Non Negativity (NN) constraint may be added to the model, so that the loading coefficients have positive values. Return an object of class clv3w.

Usage

```
CLV3W_kmeans(X,K,mode.scale=0,NN=FALSE,init=10,cp.rand=5)
```

Arguments

X : a three way array - variables of mode 2 will be clustered
 K : number of clusters

<code>mode.scale</code>	: scaling parameter applied to X, by default centering of X (for mode 2 x mode 3) is done. By default no scaling (<code>mode.scale=0</code>) 0 : no scaling only centering - the default 1 : scaling with standard deviation of (mode 2 x mode 3) elements 2 : global scaling (each block i.e. each mode 2 slice will have the same inertia) 3 : global scaling (each block i.e. each mode 3 slice will have the same inertia)
<code>NN</code>	: non Negativity constraint to be added on the loading coefficients. By default no constraint (<code>NN=FALSE</code>) <code>TRUE</code> : a non negativity constrained is applied on the loading coefficients to set them as positive values <code>FALSE</code> : loading coefficients may be either positive or negative
<code>init</code>	: either the number of random starts i.e. partitions generated for the initialisation (By default <code>init=10</code>)
<code>cp.rand</code>	: number of random starts associated with the one rank Candecomp/Parafac model (By default <code>cp.rand=10</code>)

Value

<code>results</code>	<ul style="list-style-type: none"> • <code>clusters</code>: in line 1, the groups membership in the initial partition; in line 2 the final groups membership • <code>comp</code>: the latent components of the clusters associated with the first mode • <code>loading</code>: the vector of loadings associated with the second mode by cluster • <code>weight</code>: the vector of weights associated with the third mode by cluster • <code>criterion</code>: vector of loss giving for each cluster the residual amount between the sub-array and its reconstitution associated with the cluster one rank PARAFAC model • <code>niter</code>: number of iterations of the partitioning algorithm
<code>param</code>	contains the clustering parameters <ul style="list-style-type: none"> • <code>X</code>: the scaled three-way array
<code>call</code>	: call of the method

Author(s)

Veronique Cariou, <veronique.cariou@oniris-nantes.fr>

References

- Wilderjans, T. F., & Cariou, V. (2016). CLV3W: A clustering around latent variables approach to detect panel disagreement in three-way conventional sensory profiling data. *Food quality and preference*, 47, 45-53.
- Cariou, V., & Wilderjans, T. F. (2018). Consumer segmentation in multi-attribute product evaluation by means of non-negatively constrained CLV3W. *Food Quality and Preference*, 67, 18-26.

See Also

`summary.clv3W`, `print.clv3W`

Examples

```

data(coffee)
## Cluster Analysis of coffee sensory descriptors with block scaling
## to set the assessors to the same footing
res.coffee <- CLV3W_kmeans(coffee,K=2,NN=TRUE,mode.scale=3,init=1,cp.rand=1)
summary(res.coffee)
get_partition(res.coffee)

```

CLV_kmeans

K-means algorithm for the clustering of variables

Description

K-means algorithm for the clustering of variables. Directional or local groups may be defined. Each group of variables is associated with a latent component. Moreover external information collected on the observations or on the variables may be introduced.

Usage

```

CLV_kmeans(
  X,
  Xu = NULL,
  Xr = NULL,
  method,
  sX = TRUE,
  sXr = FALSE,
  sXu = FALSE,
  clust,
  iter.max = 20,
  nstart = 100,
  strategy = "none",
  rho = 0.3
)

```

Arguments

X	The matrix of the variables to be clustered
Xu	The external variables associated with the columns of X
Xr	The external variables associated with the rows of X
method	The criterion to use in the cluster analysis. 1 or "directional" : the squared covariance is used as a measure of proximity (directional groups). 2 or "local" : the covariance is used as a measure of proximity (local groups)
sX	TRUE/FALSE : standardization or not of the columns X (TRUE by default) (predefined -> cX = TRUE : column-centering of X)

<code>sXr</code>	TRUE/FALSE : standardization or not of the columns Xr (FALSE by default) (predefined -> <code>cXr = TRUE</code> : column-centering of Xr)
<code>sXu</code>	TRUE/FALSE : standardization or not of the columns Xu (FALSE by default) (predefined -> <code>cXu = FALSE</code> : no centering, Xu considered as a weight matrix)
<code>clust</code>	: a number i.e. the size of the partition, K, or a vector of INTEGERS i.e. the group membership of each variable in the initial partition (integer between 1 and K)
<code>iter.max</code>	maximal number of iteration for the consolidation (20 by default)
<code>nstart</code>	nb of random initialisations in the case where init is a number (100 by default)
<code>strategy</code>	"none" (by default), or "kplusone" (an additional cluster for the noise variables), or "sparselv" (zero loadings for the noise variables)
<code>rho</code>	a threshold of correlation between 0 and 1 (0.3 by default)

Details

The initialization can be made at random, repetitively, or can be defined by the user.

The parameter "strategy" makes it possible to choose a strategy for setting aside variables that do not fit into the pattern of any cluster.

Value

<code>tabres</code>	The value of the clustering criterion at convergence. The percentage of the explained initial criterion value. The number of iterations in the partitioning algorithm.
<code>clusters</code>	the group's membership
<code>comp</code>	The latent components of the clusters
<code>loading</code>	if there are external variables Xr or Xu : The loadings of the external variables

References

- Vigneau E., Qannari E.M. (2003). Clustering of variables around latents components. *Comm. Stat.* 32(4), 1131-1150.
- Vigneau E., Chen M., Qannari E.M. (2015). ClustVarLV: An R Package for the clustering of Variables around Latent Variables. *The R Journal*, 7(2), 134-148
- Vigneau E., Chen M. (2016). Dimensionality reduction by clustering of variables while setting aside atypical variables. *Electronic Journal of Applied Statistical Analysis*, 9(1), 134-153

See Also

CLV, LCLV

Examples

```
data(apples_sh)
#local groups with external variables Xr
resclvkmYX <- CLV_kmeans(X = apples_sh$pref, Xr = apples_sh$senso, method = "local",
  sX = FALSE, sXr = TRUE, clust = 2, nstart = 20)
```

`coffee`*coffee data*

Description

Case study pertaining to consumer emotions associations for a variety of 12 coffee aromas. The participants were asked to complete each rating (i.e., rating the odor of 12 aromas on 15 emotion terms) on a 5-point rating scale.

Usage

```
data(coffee)
```

Format

An object of class "array" with 12 odors (mode 1), 84 subjects (mode 2) and 15 emotions (mode 3):

odors Vanilla, B.Rice, Lemon, Coffee.Flower, Cedar, Hazelnut, Coriander.Seed, Honey, Medicine, Apricot, Earth, Hay

subjects persons from Oniris

emotions Amused, Angry, Calm, Disappointed, Disgusted, Energetic, Excited, Free, Happy, Irritated, Nostalgic, Surprised, Unique, Unpleasant and Well

References

Cariou, V., & Wilderjans, T. F. (2018). Consumer segmentation in multi-attribute product evaluation by means of non-negatively constrained CLV3W. *Food Quality and Preference*, 67, 18-26.

Examples

```
data(coffee)
str(coffee)
```

`data_biplot`*biplot for the dataset*

Description

Loading plot of the variables from a Principal Components Analysis. scores of the observations are surimposed

Usage

```
data_biplot(X, sX = TRUE, axeh = 1, axev = 2, cex.lab = 1)
```

Arguments

X	the data matrix
sX	TRUE/FALSE : standardization or not of the columns X (TRUE by default)
axeh	component number for the horizontal axis
axev	component number for the vertical axis
cex.lab	: magnification to be used for labels (1 by default)

get_comp	<i>latent components associated with each cluster</i>
----------	---

Description

To get the latent components associated with each cluster.

Usage

```
get_comp(resclv, K = NULL, graph = FALSE, cex.lab = 1)
```

Arguments

resclv	: result of CLV(), CLV_kmeans(), LCLV(), CLV3W() or CLV3W_km()
K	: the number of clusters chosen (already defined if CLV_kmeans or CLV3W_kmeans is used)
graph	: boolean, if TRUE, the barplot associated with the scores is displayed (default : graph=FALSE)
cex.lab	: magnification to be used for labels (1 by default)

Value

comp	<p>the group latent components (centered)</p> <p>For results of CLV(_kmeans), the latent components returned have their own norm</p> <p>For results of CLV3W(_kmeans), the latent component associated with mode 1 (centered, but not standardized)</p> <p>For results of LCLV, two types of latent components are available :</p> <p>compt : The latent components of the clusters defined according to the Xr variables,</p> <p>compc : The latent components of the clusters defined according to the Xu variables</p>
------	---

Examples

```
data(apples_sh)
resclvX <- CLV(X = apples_sh$senso, method = "directional", sX = TRUE)
comp4G<-get_comp(resclvX, K = 4)
```

get_loading *Loadings of the variables on the latent component, in each cluster.*

Description

To get the variables loadings for the latent component in each cluster.

For CLV(_kmeans), the loadings are of particular interest when method="directional" or when strategy="sparse LV"

For CLV3W(_kmeans), the loadings are given for the variables associated with mode 2 of the 3-way array.

Usage

```
get_loading(resclv, K = NULL, type = "list", graph = FALSE, cex.lab = 1)
```

Arguments

resclv : result of CLV(), CLV_kmeans(), LCLV(), CLV3W() or CLV3W_kmeans()
 K : the number of clusters chosen (already defined if CLV_kmeans or CLV3W_kmeans is used)
 type : outputs in the form of a "list" (one element by cluster, by default) or a "vector" (available only if "sparselv" strategy is used)
 graph : boolean, if TRUE, the barplot associated with the scores is displayed (default : graph=FALSE)
 cex.lab : magnification to be used for labels (1 by default)

Value

loading the loadings of the variables on each cluster's latent component

get_partition *clusters memberships for a partition into K clusters.*

Description

To get the clusters memberships of the variables.

Usage

```
get_partition(resclv, K = NULL, type = "vector")
```

Arguments

resclv : result of CLV(), CLV_kmeans(), LCLV(), CLV3W() or CLV3W_km()
 K : the number of clusters chosen (already defined if CLV_kmeans or CLV3W_kmeans is used)
 type : "vector" (by default) for output given as a vector of integers between 1 and K (with 0 for "kplusone" strategy),
 "matrix", the output given as a binary matrix of size p x n.

Value

partition the group's membership for the variables, in a vector or matrix form.
 For CLV3W object, a vector of memberships with mode 2)

Examples

```
data(apples_sh)
resclvX <- CLV(X = apples_sh$senso, method = "directional", sX = TRUE)
parti4G<-get_partition(resclvX, K = 4)
```

get_weight	<i>Weights of the external variables, or additional mode, on the latent component in each cluster.</i>
------------	--

Description

To get the weights associated with each cluster. For CLV(_kmeans) or LCLV, applies only when external variables (Xr, Xu or both) are involved. For CLV3W(_kmeans), the weights are associated with the third mode of the 3-way array.

Usage

```
get_weight(resclv, K = NULL, graph = FALSE, cex.lab = 1)
```

Arguments

resclv : result of CLV(), CLV_kmeans() or LCLV(), CLV3W() or CLV3W_kmeans()
 K : the number of clusters chosen (already defined if CLV_kmeans or CLV3W_kmeans is used)
 graph : boolean, if TRUE, the barplot associated with the scores is displayed (default : graph=FALSE)
 cex.lab : magnification to be used for labels (1 by default)

Value

weight Weights in each cluster (associated with mode 3 for CLV3W object)
 For each cluster, the vector of weights is set to length 1
 output provided as matrix with K columns (K: number of clusters)
 In the special case of LCLV, two matrices of weights are defined :
 weight_v : weights of the external Xr variables,
 weight_u : weights of the external Xu variables.

 imput_clv

Imputation of a data matrix based on CLV results

Description

For each variable, its missing data will be imputed according to the values of the latent variable of the group in which the variable belong to.

Usage

```
imput_clv(x, X0, K = NULL)
```

Arguments

x : an object of class clv
 X0 : the initial data matrix with missing values (NA)
 K : the number of Latent Variables to be considered, each of them being associated with a group of variables.

Details

It is advised to use a larger number of latent variables, on the basis of which the imputation will be done, than the suspected 'true' number of groups of variables

Value

X0imput : the imputed data matrix, in the original scale

Ximput : the imputed matrix, centered and scaled according to the pretreatment parameters chosen in CLV

LCLV

*L-CLV for L-shaped data***Description**

Define clusters of X-variables around latent components. In each cluster, two latent components are extracted, the first one is a linear combination of the external information collected for the rows of X and the second one is a linear combination of the external information associated with the columns of X.

Usage

```
LCLV(X, Xr, Xu, ccX = FALSE, sX = TRUE, sXr = FALSE, sXu = FALSE, nmax = 20)
```

Arguments

X	The matrix of variables to be clustered
Xr	The external variables associated with the rows of X
Xu	The external variables associated with the columns of X
ccX	TRUE/FALSE : double centering of X (FALSE, by default) If FALSE this implies that cX = TRUE : column-centering of X
sX	TRUE/FALSE : standardization or not of the columns X (TRUE by default)
sXr	TRUE/FALSE : standardization or not of the columns Xr (FALSE by default) (predefined -> cXr = TRUE : column-centering of Xr)
sXu	TRUE/FALSE : standardization or not of the columns Xu (FALSE by default) (predefined -> cXu= FALSE : no centering, Xu considered as a weight matrix)
nmax	maximum number of partitions for which the consolidation will be done (by default nmax=20)

Value

tabres	<p>Results of the clustering algorithm. In each line you find the results of one specific step of the hierarchical clustering.</p> <ul style="list-style-type: none"> • Columns 1 and 2 : The numbers of the two groups which are merged • Column 3 : Name of the new cluster • Column 4 : The value of the aggregation criterion for the Hierarchical Ascendant Clustering (HAC) • Column 5 : The value of the clustering criterion for the HAC • Column 6 : The percentage of the explained initial criterion value • Column 7 : The value of the clustering criterion after consolidation • Column 8 : The percentage of the explained initial criterion value after consolidation • Column 9 : number of iterations in the partitioning algorithm. <p>Remark: A zero in columns 7 to 9 indicates that no consolidation was done</p>
--------	---

- partition K a list for each number of clusters of the partition, $K=2$ to n_{max} with
- clusters : in line 1, the groups membership before consolidation; in line 2 the groups membership after consolidation
 - compt : The latent components of the clusters (after consolidation) defined according to the X_r variables
 - compc : The latent components of the clusters (after consolidation) defined according to the X_u variables
 - loading_v : loadings of the external X_r variables (after consolidation)
 - loading_u : loadings of the external X_u variables (after consolidation)

References

- Vigneau E., Qannari E.M. (2003). Clustering of variables around latents components. *Comm. Stat*, 32(4), 1131-1150.
- Vigneau, E., Charles, M., & Chen, M. (2014). External preference segmentation with additional information on consumers: A case study on apples. *Food Quality and Preference*, 32, 83-92.
- Vigneau E., Chen M., Qannari E.M. (2015). *ClustVarLV: An R Package for the clustering of Variables around Latent Variables*. *The R Journal*, 7(2), 134-148

 lm_CLV

linear model based on CLV

Description

prediction of a response variable, y , based on clusters of predictors variables, X . boosted-like procedure for identifying groups of predictors, and their associated latent component, well correlated with the actual residuals of response variable, y . sparsity is allowed using the strategy options ("sparselv" or "kplusone") and the rho parameter.

Usage

```
lm_CLV(
  X,
  y,
  method = "directional",
  sX = TRUE,
  shrinkp = 0.5,
  strategy = "none",
  rho = 0.3,
  validation = FALSE,
  id.test = NULL,
  maxiter = 100,
  threshold = 1e-05
)
```

Arguments

X	: The matrix of the predictors, the variables to be clustered
y	: The response variable (usually numeric) If y is binary factor, indicator variable (0/1) is generated. A Bayes rule is used to compute class probabilities. Performance criteria is RMSE for numerical variable; RMSE and error rate for binary factor.
method	: The criterion to be use in the cluster analysis. 1 or "directional" : the squared covariance is used as a measure of proximity (directional groups). 2 or "local" : the covariance is used as a measure of proximity (local groups)
sX	: TRUE/FALSE, i.e. standardization or not of the columns X (TRUE by default)
shrinkp	: shrinkage paramater used in the boosting (max : 1, 0.5 by default). If shrinkp is a vector of positive values greater than 0, and lower or equal to 1, the outputs are given for each value.
strategy	: "none" (by default), or "kplusone" (an additional cluster for the unclassifiable variables), or "sparselv" (zero loadings for the unclassifiable variables)
rho	: a threshold of correlation between 0 and 1 (used in "kplusone" or "sparselv" strategy, 0.3 by default)
validation	TRUE/FALSE i.e. using a test set or not. By default no validation
id.test	: if validation==TRUE, the number of the observations used as test set
maxiter	: the maximum number of components extracted (100 by default)
threshold	: used in a stopping rule, when the relative calibration errors sum of squares stabilizes (10e-6 by default)

Value

Group	a list of the groups of variables X in order of the first time extracted.
Comp	a list of the latent components associated with the groups of X variables extracted.
Load	a list for the loadings of the X variables in the latent component.
Alpha	a list of the regression coefficients to be applied to the latent components. The coefficients are aggregated when the same latent component is extracted several times during the iterative steps.
Beta	a list of the beta coefficients to be applied to the pretreated predictors. For a model with the A first latent components, the A first elements of the list must be added together.
GroupImp	Group Importance i.e. the decrease of the residuals' variance provided by the CLV components in the model.
RMSE.cal	the root mean square error for the calibration set, at each step of the procedure.
ERRrate.cal and rocAUC.cal	when y is a binary factor, the classification rate and the AUC for ROC, on the basis of the calibration set, at each step of the procedure.

RMSE.val as RMSE.cal but for the test set, if provided.
 ERRrate.val and rocAUC.val
 as for calibration set but for the test set, if provided.

See Also

CLV, CLV_kmeans

plot.clv *Graphical representation of the CLV clustering stages*

Description

This function plots either the CLV dendrogram or the variations of the consolidated CLV criterion.

Usage

```
## S3 method for class 'clv'
plot(x, type = "dendrogram", cex = 0.8, ...)
```

Arguments

x : an object of class clv
 type : What to plot.
 "dendrogram" : the dendrogram of the hierchical clustering algorithm,
 "delta" : a barplot showing the variation of the clustering criterium after consol-
 idation.
 cex : Character expansion for labels.
 ... further arguments passed to or from other methods

See Also

CLV

`plot.clv3w`*Graphical representation of the CLV3W hierarchical clustering stages*

Description

This function plots either the CLV3W dendrogram or the variations of the consolidated CLV3W criterion.

Usage

```
## S3 method for class 'clv3w'  
plot(x, type = "dendrogram", cex = 0.8, ...)
```

Arguments

`x` : an object of class `clv3w`

`type` : What to plot.
"dendrogram" : the dendrogram of the hierarchical clustering algorithm,
"delta" : a barplot showing the variation of the clustering criterion after consolidation.

`cex` : Character expansion for labels.

... Additional arguments passed on to the real print.

See Also

CLV3W

`plot.lclv`*Graphical representation of the LCLV clustering stages*

Description

This function plots either the CLV dendrogram or the variations of the consolidated CLV criterion.

Usage

```
## S3 method for class 'lclv'  
plot(x, type = "dendrogram", cex = 0.8, ...)
```


Arguments

x : an object of class lclv
 type : What to plot.
 "dendrogram" : the dendrogram of the hierarchical clustering algorithm,
 "delta" : a barplot showing the variation of the clustering criterium after consolidation.
 cex : Character expansion for labels.
 ... further arguments passed to or from other methods

See Also

LCLV

 plot_var

Representation of the variables and their group membership

Description

Loading plot of the variables from a Principal Components Analysis. The group membership of the variables is superimposed.

Usage

```
plot_var(
  resclv,
  K = NULL,
  axeh = 1,
  axev = 2,
  label = FALSE,
  cex.lab = 1,
  v_colors = NULL,
  v_symbol = FALSE,
  beside = FALSE
)
```

Arguments

resclv : results of CLV(), CLV_kmeans() or LCLV()
 K : the number of groups in the partition (already defined if CLV_kmeans is used)
 axeh : component number for the horizontal axis
 axev : component number for the vertical axis
 label : = TRUE :the column names in X are used as labels / = FALSE: no labels (by default)
 cex.lab : magnification to be used for labels (1 by default)

v_colors	default NULL. If missing colors are given, by default
v_symbol	=TRUE : symbols are given instead of colors for the identification of the groups/ =FALSE: no symbol (by default).
beside	=TRUE : a plot per cluster of variables, side-by-side/ =FALSE :an unique plot with all the variables with the identification of their group membership (by default).

Examples

```
data(apples_sh)
resclvX <- CLV(X = apples_sh$senso, method = 1, sX = TRUE)
plot_var(resclvX, K = 4, axeh = 1, axev = 2)
```

plot_var.clv3w	<i>Scores plot from a Candecom Parafac analysis. The group membership of the variables is superimposed.</i>
----------------	---

Description

Scores plot from a Candecom Parafac analysis. The group membership of the variables is superimposed.

Usage

```
plot_var.clv3w(  
  resclv3w,  
  K = NULL,  
  axeh = 1,  
  axev = 2,  
  labels = FALSE,  
  cex.lab = 1,  
  v_colors = NULL,  
  v_symbol = FALSE,  
  beside = FALSE,  
  mode3 = FALSE  
)
```

Arguments

resclv3w	the data matrix
K	the number of groups in the partition (already defined if CLV3W_kmeans is used)
axeh	component number for the horizontal axis
axev	component number for the vertical axis
labels	boolean to add variable' labels (label=TRUE) on the plot or not (label=FALSE). By default label=TRUE

cex.lab	magnification to be used for labels (1 by default)
v_colors	default NULL. If missing colors are given, by default
v_symbol	symbols are given instead of colors for the identification of the groups/ =FALSE: no symbol (by default).
beside	plot per cluster of variables, side-by-side =FALSE : an unique plot with all the variables with the identification of their group membership (by default).
mode3	projection of the mode 3 elements onto the scores plot =FALSE : mode 3 elements are not represented (by default).

predict.lmclv *prediction for lmCLV models.*

Description

To get the predicted response values based on lmCLV model.

Usage

```
## S3 method for class 'lmclv'
predict(object, newdata, shrinkp, ...)
```

Arguments

object	: result of class lmclv
newdata	: Data frame of observations for which to make predictions
shrinkp	: shrinkage parameter.
...	: further arguments passed to or from other methods

Value

a matrix of the predicted values,
each column with an increasing number of CLV component included
the first column being for the null model
if the response is a binary factor, two additional matrices are provided :
the probabilities of belonging to class 1 and the response values (0 or 1).

print.clv *Print the CLV results*

Description

Print the CLV results

Usage

```
## S3 method for class 'clv'  
print(x, ...)
```

Arguments

x an object of class clv
... further arguments passed to or from other methods

See Also

CLV

print.clv3w *Print the CLV3W results*

Description

Print the CLV3W results

Usage

```
## S3 method for class 'clv3w'  
print(x, ...)
```

Arguments

x an object of class clv3w
... Additional arguments passed on to the real print.

See Also

CLV3W, CLV3W_kmeans

print.lclv	<i>Print the LCLV results</i>
------------	-------------------------------

Description

Print the LCLV results

Usage

```
## S3 method for class 'lclv'
print(x, ...)
```

Arguments

x	an object of class lclv
...	further arguments passed to or from other methods

See Also

LCLV

stand_quali	<i>Standardization of the qualitative variables</i>
-------------	---

Description

pretreatment of qualitative variables

Usage

```
stand_quali(X.quali, metric = "chisq")
```

Arguments

X.quali	: a factor or a data frame with several factors
metric	: the metric to be used, i.e. each category is weighted by the inverse of the square-root of its relative frequency

Value

Xdisj.sd : a standardized matrix with as many columns as categories associated with the qualitative variables.

summary.clv

summary and description of the clusters of variables

Description

This function provides the list of the variables within each group and complementary informations. Users will be asked to specify the number of clusters,

Usage

```
## S3 method for class 'clv'
summary(object, K = NULL, ...)
```

Arguments

object : result of CLV() or CLV_kmeans()
K : the number of clusters (unless if CLV_kmeans was used)
... further arguments passed to or from other methods

Details

The outputs include :

- the size of the groups,
- the list of the variables within each group. For each cluster, the correlation of the each variable with its group latent component and the correlation with the next neighbouring group latent component are given.
- the proportion of the variance within each group explained by its latent variable,
- the proportion of the whole dataset account by the group latent variables
- the matrix of correlation between the latent variables.

summary.clv3w

Summary and description of the clusters of (mode 2) variables associated with CLV3W or CLV3W_kmeans

Description

This function provides the list of the variables within each group and complementary informations. Users will be asked to specify the number of clusters,

Usage

```
## S3 method for class 'clv3w'  
summary(object, K = NULL, ...)
```

Arguments

object : result of CLV3W() or CLV3W_kmeans()
K : the number of clusters (unless if CLV3W_kmeans was used)
... Additional arguments passed on to the real summary.

Details

The outputs include :

- the size of the groups,
- the proportion of the variance within each group explained by its latent variable,
- the proportion of the whole dataset accounted by the group latent variables
- the latent components (mode 1) associated to the various groups,
- the weights (mode 3) associated to the various groups,
- the list of the variables within each group. For each cluster, the loading (mode 2) of the variable is given together with the correlation of the block component with its group latent component and the correlation with the next neighbouring group latent component are given.
- the matrix of correlation between the latent variables.

Index

* datasets

- apples_sh, 2
- AUPA_psycho, 3
- authen_NMR, 4
- ciders, 6
- coffee, 15

- apples_sh, 2
- AUPA_psycho, 3
- authen_NMR, 4

- block.scale, 4
- boot_clv, 5

- ciders, 6
- CLV, 7
- CLV3W, 9
- CLV3W_kmeans, 11
- CLV_kmeans, 13
- coffee, 15

- data_biplot, 15

- get_comp, 16
- get_loading, 17
- get_partition, 17
- get_weight, 18

- imput_clv, 19

- LCLV, 20
- lm_CLV, 21

- plot.clv, 23
- plot.clv3w, 24
- plot.lclv, 24
- plot_var, 25
- plot_var.clv3w, 26
- predict.lmclv, 27
- print.clv, 28
- print.clv3w, 28

- print.lclv, 29

- stand_quali, 29
- summary.clv, 30
- summary.clv3w, 30