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 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_psycho ............................................................ 3
authen_NMR .............................................................. 4
block.scale ............................................................... 4
boot_clv ................................................................. 5
ciders ................................................................. 6
apples_sh

apples from southern hemisphere data set

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

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

---

AUPA_psycho  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_psycho)

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

Examples
X = data(AUPA_psycho)
**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**

```r
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**


**Examples**

```r
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**

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

Arguments

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

Value

Xscaled : the scaled three-way array

boot_clv

Boostrapping 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

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>object</td>
<td>result of CLV()</td>
</tr>
<tr>
<td>case</td>
<td>&quot;row&quot; or &quot;column&quot; corresponding to the random effect of the design</td>
</tr>
<tr>
<td>B</td>
<td>the number of bootstrap to be run (100 by default)</td>
</tr>
<tr>
<td>nmax</td>
<td>maximal size of the partitions to be considered (if NULL, the value of nmax used for the object is used)</td>
</tr>
</tbody>
</table>

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


Examples

- data(ciders)
- str(ciders)
**Hierarchical clustering of variables with consolidation**

**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**

```r
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 used 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)
TRUE/FALSE (by default TRUE) : dendrogram and variation of the optimization criterion.
Theese plots can also be obtained with "plot"

Details

If external variables are used, define either Xr or Xu, but not both. Use the LCLV function when Xr
and Xu 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 As-
    cendant 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 nmax 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 Xr or Xu : The loadings of the
    external variables (after consolidation)

References

32(4), 1131-1150.


See Also

CLV_kmeans, LCLV
**Examples**

```r
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)
```

---

**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**

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

**Arguments**

- **X**: a three way array - variables of mode 2 will be clustered
- **mode.scale**: scaling parameter applied to X, by default centering of X (for mode 2 x mode 3) is done. By default no scaling (mode.scale=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)
- **NN**: non Negativity constraint to be added on the loading coefficients. By default no constraint (NN=FALSE)
  - 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
- **moddendoinertie**: dendrogram. By default it is based on the delta clustering criterion (moddendoinertie=TRUE)
  - TRUE: dendrogram associated with the clustering criterion delta
  - FALSE: dendrogram associated with the the height (cumulative delta)
- **gmax**: maximum number of partitions for which the consolidation will be done (default: gmax=11)
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

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

Value hclust contains the results of the HCA

Value 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)
• weight : 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)

Value param contains the clustering parameters
• gmax : maximum number of partitions for which the consolidation has been done
• X : the scaled three-way array

Value call : call of the method

Author(s)

Veronique Cariou, <veronique.cariou@oniris-nantes.fr>
CLV3W_kmeans

References


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

Partitioning algorithm of a set of variables (associated with mode 2) oh a three-way array

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
mode.scale : scaling parameter applied to X, by default centering of X (for mode 2 x mode 3) is done. By default no scaling (mode.scale=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 )

NN : non Negativity constraint to be added on the loading coefficients. By default no constraint (NN=FALSE)
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

init : either the number of random starts i.e. partitions generated for the initialisation (By default init=10)

cp.rand : number of random starts associated with the one rank Candecomp/Parafac model (By default cp.rand=10)

Value

results • clusters: in line 1, the groups membership in the initial partition; in line 2 the final groups membership
• comp: the latent components of the clusters associated with the first mode
• loading: the vector of loadings associated with the second mode by cluster
• weigth: the vector of weights associated with the third mode by cluster
• 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
• niter: number of iterations of the partitioning algorithm

param contains the clustering parameters
• X: the scaled three-way array

call : call of the method

Author(s)
Veronique Cariou, <veronique.cariou@oniris-nantes.fr>

References

See Also
summary.clv3W, print.clv3W
**Examples**

```r
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)
```

---

**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**

```r
CLV_kmeans(
    X,  # The matrix of the variables to be clustered
    Xu = NULL,  # The external variables associated with the columns of X
    Xr = NULL,  # The external variables associated with the rows of X
    method,  # The criterion to use in the cluster analysis.
    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)
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)

clust       : 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)

iter.max    maximal number of iteration for the consolidation (20 by default)

nstart      nb of random initialisations in the case where init is a number (100 by default)

strategy    "none" (by default), or "kplusone" (an additional cluster for the noise variables),  
or "sparselv" (zero loadings for the noise variables)

rho         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

tabres      The value of the clustering criterion at convergence.
            The percentage of the explained initial criterion value.
            The number of iterations in the partitioning algorithm.

clusters    the group's membership

comp        The latent components of the clusters

loading     if there are external variables Xr or Xu : The loadings of the external variables

References

32(4), 1131-1150.

around Latent Variables. The R Journal, 7(2), 134-148

Vigneau E., Chen M. (2016). Dimensionality reduction by clustering of variables while setting  

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

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


Examples

data(coffee)
str(coffee)

data_biplot

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)
get_comp

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>the data matrix</td>
</tr>
<tr>
<td>sX</td>
<td>TRUE/FALSE : standardization or not of the columns X (TRUE by default)</td>
</tr>
<tr>
<td>axeh</td>
<td>component number for the horizontal axis</td>
</tr>
<tr>
<td>axev</td>
<td>component number for the vertical axis</td>
</tr>
<tr>
<td>cex.lab</td>
<td>: magnification to be used for labels (1 by default)</td>
</tr>
</tbody>
</table>

Description

To get the latent components associated with each cluster.

Usage

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

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>resclv</td>
<td>: result of CLV(), CLV_kmeans(), LCLV(), CLV3W() or CLV3W_km()</td>
</tr>
<tr>
<td>K</td>
<td>: the number of clusters chosen (already defined if CLV_kmeans or CLV3W_km is used)</td>
</tr>
<tr>
<td>graph</td>
<td>: boolean, if TRUE, the barplot associated with the scores is displayed (default : graph=FALSE)</td>
</tr>
<tr>
<td>cex.lab</td>
<td>: magnification to be used for labels (1 by default)</td>
</tr>
</tbody>
</table>

Value

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>comp</td>
<td>the group latent components (centered)</td>
</tr>
</tbody>
</table>

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")`
get_weight

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 : "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

Weights of the external variables, or additional mode, on the latent
component in each cluster.

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

geweight(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.

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 pretratment 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**

\[
\text{LCLV}(X, X_r, X_u, \text{ccX} = \text{FALSE}, \text{sX} = \text{TRUE}, \text{sXr} = \text{FALSE}, \text{sXu} = \text{FALSE}, n_{\text{max}} = 20)
\]

**Arguments**

- **X**
  - The matrix of variables to be clustered
- **X_r**
  - The external variables associated with the rows of X
- **X_u**
  - 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)
- **sX_r**
  - TRUE/FALSE : standardization or not of the columns X_r (FALSE by default) (predefined -> cX_r = TRUE : column-centering of X_r)
- **sX_u**
  - TRUE/FALSE : standardization or not of the columns X_u (FALSE by default) (predefined -> cX_u = FALSE : no centering, X_u considered as a weight matrix)
- **n_{\text{max}}**
  - maximum number of partitions for which the consolidation will be done (by default n_{\text{max}} = 20)

**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
    - 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.
  - Remark: A zero in columns 7 to 9 indicates that no consolidation was done
partition K a list for each number of clusters of the partition, K=2 to nmax 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 Xr variables
  • compc : The latent components of the clusters (after consolidation) defined
    according to the Xu variables
  • loading_v : loadings of the external Xr variables (after consolidation)
  • loading_u : loadings of the external Xu variables (after consolidation)

References


lm_CLV  

linear model based on CLV

Description

prediction of a response variable, y, based on clusters of predictors variables, X. boosted-liked 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

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 consolidation.
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**

```r
## 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**

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

**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**

```r
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 def-
ault).

Examples

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

Description

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

Usage

```r
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 i 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
predict.lmclv

**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**

```r
## 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 if 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
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

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**

```r
## 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 each variable with its group latent component and the correlation with the next neighboring 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,
summary.clv3w

Usage

```r
## 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