Package ‘LDABiplots’

January 20, 2025

Type Package

Title Biplot Graphical Interface for LDA Models

Version 0.1.2

Maintainer Luis Pilacuan-Bonete <`luis.pilacuanb@ug.edu.ec`>

Description Contains the development of a tool that provides a web-based graphical user interface (GUI) to perform Biplots representations from a scraping of news from digital newspapers under the Bayesian approach of Latent Dirichlet Assignment (LDA) and machine learning algorithms. Contains LDA methods described by Blei , David M., Andrew Y. Ng and Michael I. Jordan (2003) <`https://jmlr.org/papers/volume3/blei03a/blei03a.pdf`>, and Biplot methods described by Gabriel K.R(1971) <doi:10.1093/biomet/58.3.453> and Galindo-Villardon P(1986) <`https://diarium.usal.es/pgalindo/files/2012/07/Questiio.pdf`>.

License GPL-3

Encoding UTF-8

Imports shiny, shinyBS, shinydashboard, shinyWidgets, shinyalert, shinybusy, shinyjs, shinycssloaders, dplyr, ggplot2, rvest, DT, highcharter, tidyr, SnowballC, ldatuning, topicmodels, textmineR, chinese.misc, stringr, htmlwidgets, ggrepel, textplot, glasso, qgraph, Matrix, utils, factoextra, quanteda

RoxygenNote 7.2.0

Suggests rmarkdown, knitr, beepr, readxl

VignetteBuilder knitr

NeedsCompilation no

Author Luis Pilacuan-Bonete [cre, aut] (<`https://orcid.org/0000-0002-6625-0905`>), Purificacion Galindo-Villardón [aut] (<`https://orcid.org/0000-0001-6977-7545`>), Javier De La Hoz Maestre [aut] (<`https://orcid.org/0000-0001-7779-0803`>), Francisco Javier Delgado-Álvarez [aut] (<`https://orcid.org/0000-0002-2626-2709`>)

Repository CRAN

Date/Publication 2022-07-18 12:00:08 UTC

Contents

`dtmcorr` *Pearson Correlation for Sparse Matrices*

Description

Pearson Correlation for Sparse Matrices. More memory and time-efficient than `cor(as.matrix(x))`.

Usage

```
dtmcorr(x)
```

Arguments

`x`	A matrix, potentially a sparse matrix such as a "dgCMatrix" object

Value

a correlation matrix

`dtmremovetfidf` *Remove terms from a Document-Term-Matrix and documents with no terms based on the term frequency inverse document frequency*

Description

Remove terms from a Document-Term-Matrix and documents with no terms based on the term frequency inverse document frequency. Either giving in the maximum number of terms (argument `top`), the tfidf cutoff (argument `cutoff`) or a quantile (argument `prob`)

Usage

```
dtmremovetfidf(dtm, top, cutoff, prob, remove_emptydocs = TRUE)
```

Arguments

`dtm`	an object class "dgCMatrix"
`top`	integer with the number of terms which should be kept as defined by the highest mean tfidf
`cutoff`	numeric cutoff value to keep only terms in `dtm` where the tfidf obtained by `dtmtfidf` is higher than this value
`prob`	numeric quantile indicating to keep only terms in `dtm` where the tfidf obtained by `dtmtfidf` is higher than the `prob` percent quantile
`remove_emptydocs`	logical indicating to remove documents containing no more terms after the term removal is executed. Defaults to `TRUE`.

Value

a sparse Matrix as returned by `sparseMatrix` where terms with high tfidf are kept and documents without any remaining terms are removed

---

`dtmtfidf` *Term Frequency - Inverse Document Frequency calculation*

Description

Term Frequency - Inverse Document Frequency calculation. Averaged by each term.

Usage

```
dtmtfidf(dtm)
```

Arguments

`dtm`	an object class "dgCMatrix"

Value

a vector with tfidf values, one for each term in the `dtm` matrix

---

`GHBiplot` *GHBiplot*

Description

This function performs the representation of GHBiplot (Gabriel,1971).

Usage

```
GHBiplot (X, Transform.Data = 'scale')
```

Arguments

`X`	array_like; A data frame which provides the data to be analyzed. All the variables must be numeric.
`Transform.Data`	character; A value indicating whether the columns of X (variables) should be centered or scaled. The options are: "center" if center is TRUE, centering is done by subtracting the column means (omitting NA) of x from their corresponding columns, and if center is FALSE, centering is not done. "scale" the value of scale determines how column scaling is performed (after centering). If scale is a numeric-alike vector with length equal to the number of columns of x, then each column of x is divided by the corresponding value from scale. If scale is TRUE then scaling is done by dividing the (centered) columns of x by their standard deviations if center is TRUE, and the root mean square otherwise. If scale is FALSE, no scaling is done. To scale by standard deviations without centering, use scale(x,center=FALSE,scale=apply(x,2,sd,na.rm=TRUE)),"center_scale" center=TRUE and scale=TRUE,"none" neither center nor scale is done. The default value is "scale".

Details

Algorithm used to construct the GH Biplot. The Biplot is obtained as result of the configuration of markers for individuals and markers for variables in a reference system defined by the factorial axes resulting from the Decomposition in Singular Values (DVS).

Value

`GHBiplot` returns a list containing the following components:

`eigenvalues`	array_like; vector with the eigenvalues.
`explvar`	array_like; an vector containing the proportion of variance explained by the first 1, 2,.,k principal components obtained.
`loadings`	array_like; the loadings of the principal components.
`coord_ind`	array_like; matrix with the coordinates of individuals.
`coord_var`	array_like; matrix with the coordinates of variables.

References

- Gabriel, K. R. (1971). The Biplot graphic display of matrices with applications to principal components analysis. Biometrika, 58(3), 453-467.

Examples

```
GHBiplot(mtcars)
```

`HJBiplot` *HJBiplot*

Description

This function performs the representation of HJ Biplot (Galindo, 1986).

Usage

```
HJBiplot (X, Transform.Data = 'scale')
```

Arguments

`X`	array_like; A data frame which provides the data to be analyzed. All the variables must be numeric.
`Transform.Data`	character; A value indicating whether the columns of X (variables) should be centered or scaled. The options are: "center" if center is TRUE, centering is done by subtracting the column means (omitting NA) of x from their corresponding columns, and if center is FALSE, centering is not done. "scale" the value of scale determines how column scaling is performed (after centering). If scale is

a numeric-alike vector with length equal to the number of columns of x, then each column of x is divided by the corresponding value from scale. If scale is TRUE then scaling is done by dividing the (centered) columns of x by their standard deviations if center is TRUE, and the root mean square otherwise. If scale is FALSE, no scaling is done. To scale by standard deviations without centering, use scale(x,center=FALSE,scale=apply(x,2,sd,na.rm=TRUE)),"center_scale" center=TRUE and scale=TRUE,"none" neither center nor scale is done. The default value is "scale".

Details

Algorithm used to construct the HJ Biplot. The Biplot is obtained as result of the configuration of markers for individuals and markers for variables in a reference system defined by the factorial axes resulting from the Decomposition in Singular Values (DVS).

Value

`HJBiplot` returns a list containing the following components:

`eigenvalues`	array_like; vector with the eigenvalues.
`explvar`	array_like; an vector containing the proportion of variance explained by the first 1, 2,.,k principal components obtained.
`loadings`	array_like; the loadings of the principal components.
`coord_ind`	array_like; matrix with the coordinates of individuals.
`coord_var`	array_like; matrix with the coordinates of variables.

References

- Gabriel, K. R. (1971). The Biplot graphic display of matrices with applications to principal components analysis. Biometrika, 58(3), 453-467.
- Galindo-Villardon, P. (1986). Una alternativa de representacion simultanea: HJ-Biplot (An alternative of simultaneous representation: HJ-Biplot). Questiio, 10, 13-23.

Examples

```
HJBiplot(mtcars)
```

`JKBiplot` *JKBiplot*

Description

This function performs the representation of JK Biplot (Gabriel,1971).

Usage

```
JKBiplot (X, Transform.Data = 'scale')
```

Arguments

`X`	array_like; A data frame which provides the data to be analyzed. All the variables must be numeric.
`Transform.Data`	character; A value indicating whether the columns of X (variables) should be centered or scaled. The options are: "center" if center is TRUE, centering is done by subtracting the column means (omitting NA) of x from their corresponding columns, and if center is FALSE, centering is not done. "scale" the value of scale determines how column scaling is performed (after centering). If scale is a numeric-alike vector with length equal to the number of columns of x, then each column of x is divided by the corresponding value from scale. If scale is TRUE then scaling is done by dividing the (centered) columns of x by their standard deviations if center is TRUE, and the root mean square otherwise. If scale is FALSE, no scaling is done. To scale by standard deviations without centering, use scale(x,center=FALSE,scale=apply(x,2,sd,na.rm=TRUE)),"center_scale" center=TRUE and scale=TRUE,"none" neither center nor scale is done. The default value is "scale".

Details

Algorithm used to construct the JK Biplot. The Biplot is obtained as result of the configuration of markers for individuals and markers for variables in a reference system defined by the factorial axes resulting from the Decomposition in Singular Values (DVS).

Value

`JKBiplot` returns a list containing the following components:

`eigenvalues`	array_like; vector with the eigenvalues.
`explvar`	array_like; an vector containing the proportion of variance explained by the first 1, 2,.,k principal components obtained.

`loadings`	array_like; the loadings of the principal components.
`coord_ind`	array_like; matrix with the coordinates of individuals.
`coord_var`	array_like; matrix with the coordinates of variables.

References

- Gabriel, K. R. (1971). The Biplot graphic display of matrices with applications to principal components analysis. Biometrika, 58(3), 453-467.

Examples

```
JKBiplot(mtcars)
```

`Plot_Biplot` *Plotting Biplot*

Description

`Plot_Biplot` initializes a ggplot2-based visualization of the caracteristics presented in the data analized by the Biplot selected.

Usage

```
Plot_Biplot(X, axis = c(1,2), hide = "none",
 labels = "auto", ind.shape = 19,
 ind.color = "red", ind.size = 2,
 ind.label = FALSE, ind.label.size = 4,
 var.color = "black", var.size = 0.5,
 var.label = TRUE, var.label.size = 4, var.label.angle = FALSE)
```

Arguments

`X`	List containing the output of one of the functions of the package.
`axis`	Vector with lenght 2 which contains the axis ploted in x and y axis.
`hide`	Vector specifying the elements to be hidden on the plot. Default value is “none”. Other allowed values are “ind” and “var”.
`labels`	It indicates the label for points. If it is "auto" the labels are the row names of the coordinates of individuals. If it isn’t auto it would be a vector containing the labels.
`ind.shape`	Points shape. It can be a number to indicate the shape of all the points or a factor to indicate different shapes.
`ind.color`	Points colors. It can be a character indicating the color of all the points or a factor to use different colors.

`ind.size`	Size of points.
`ind.label`	Logical value, if it is TRUE it prints the name for each row of X. If it is FALSE (default) does not print the names.
`ind.label.size`	Numeric value indicating the size of the labels of points.
`var.color`	Character indicating the color of the arrows.
`var.size`	Size of arrow.
`var.label`	Logical value, if it is TRUE (default) it prints the name for each column of X. If it is FALSE does not print the names.
`var.label.size`	Numeric value indicating the size of the labels of variables.
`var.label.angle`	Logical value, if it it TRUE (default) it print the vector names with orentation of the angle of the vector. If it is FALSE the angle of all tags is 0.

Value

Return a `ggplot2` object.

See Also

`HJBiplot`

Examples

```
hj.biplot <- HJBiplot(mtcars)
Plot_Biplot(hj.biplot, ind.label = TRUE)
```

`runLDABiplots` *Shiny UI for LDABiplots package*

Description

Shiny UI for LDABiplots package

Usage

```
runLDABiplots(host = "127.0.0.1", port = NULL, launch.browser = TRUE)
```

Arguments

`host`	The IPv4 address that the application should listen on. Defaults to the shiny.host option, if set, or "127.0.0.1" if not.
`port`	is the TCP port that the application should listen on. If the port is not specified, and the shiny.port option is set (with options(shiny.port = XX)), then that port will be used. Otherwise, use a random port.
`launch.browser`	If true, the system's default web browser will be launched automatically after the app is started. Defaults to true in interactive sessions only. This value of this parameter can also be a function to call with the application's URL.

Value

No return value

Examples

```
if(interactive()){
runLDABiplots()
}
```

Index