

Package ‘SparseBiplots’

June 28, 2020

Type Package

Title 'HJ-Biplot' using Different Ways of Penalization Plotting with 'ggplot2'

Version 4.0.0

Author Mitzi Isabel Cubilla-Montilla <mitzi@usal.es>,
Carlos Alfredo Torres-Cubilla <carlos_t22@usal.es>,
Purificacion Galindo Villardon <pgalindo@usal.es> and
Ana Belen Nieto-Librero <ananieto@usal.es>

Maintainer Mitzi Isabel Cubilla-Montilla <mitzi@usal.es>

Description 'HJ-Biplot' is a multivariate method that allow represent multivariate data on a subspace of low dimension, in such a way that most of the variability of the information is captured in a few dimensions. This package implements three new techniques and constructs in each case the 'HJ-Biplot', adapting restrictions to reduce weights and / or produce zero weights in the dimensions, based on the regularization theories. It implements three methods of regularization: Ridge, LASSO and Elastic Net.

License GPL (>= 3)

Encoding UTF-8

LazyData true

Depends R (>= 3.2), ggplot2

Imports ggrepel, gtable, rlang, stats, sparsepca, testthat

URL <https://github.com/mitzicubillamontilla/SparseBiplots>

BugReports <https://github.com/mitzicubillamontilla/SparseBiplots/issues>

RoxygenNote 7.1.0

NeedsCompilation no

Repository CRAN

Date/Publication 2020-06-28 00:00:12 UTC

R topics documented:

| | |
|-------------------------------|---|
| ElasticNet_HJBiplot | 2 |
| HJBiplot | 4 |
| LASSO_HJBiplot | 5 |
| Plot_Biplot | 7 |
| Ridge_HJBiplot | 8 |

| | |
|--------------|-----------|
| Index | 10 |
|--------------|-----------|

| | |
|---------------------|------------------------------|
| ElasticNet_HJBiplot | <i>Elastic Net HJ Biplot</i> |
|---------------------|------------------------------|

Description

This function is a generalization of the Ridge regularization method and the LASSO penalty. Realizes the representation of the SPARSE HJ Biplot through a combination of LASSO and Ridge, on the data matrix. This means that with this function you can eliminate weak variables completely as with the LASSO regularization or contract them to zero as in Ridge.

Usage

```
ElasticNet_HJBiplot(X, Lambda = 1e-04, Alpha = 1e-04, Transform.Data = 'scale')
```

Arguments

| | |
|----------------|---|
| X | array_like; A data frame with the information to be analyzed |
| Lambda | float; Tuning parameter of the LASSO penalty. Higher values lead to sparser components. |
| Alpha | float; Tuning parameter of the Ridge shrinkage |
| Transform.Data | character; A value indicating whether the columns of X (variables) should be centered or scaled. Options are: "center" that removes the columns means and "scale" that removes the columns means and divide by its standard deviation. Default is "scale". |

Details

Algorithm used to perform automatic selection of variables and continuous contraction simultaneously. With this method, the model obtained is simpler and more interpretable. It is a particularly useful method when the number of variables is much greater than the number of observations.

Value

ElasticNet_HJBiplot returns a list containing the following components:

| | |
|-------------|---|
| loadings | array_like; penalized loadings, the loadings of the sparse principal components. |
| n_ceros | array_like; number of loadings equal to zero in each component. |
| coord_ind | array_like; matrix with the coordinates of individuals. |
| coord_var | array_like; matrix with the coordinates of variables. |
| eigenvalues | array_like; vector with the eigenvalues penalized. |
| explvar | array_like; an vector containing the proportion of variance explained by the first 1, 2,..,k sparse principal components obtained. |

Author(s)

Mitzi Cubilla-Montilla, Carlos Torres-Cubilla, Ana Belen Nieto Librero and Purificacion Galindo Villardon

References

- Galindo, M. P. (1986). Una alternativa de representacion simultanea: HJ-Biplot. *Questiio*, 10(1), 13-23.
- Erichson, N. B., Zheng, P., Manohar, K., Brunton, S. L., Kutz, J. N., & Aravkin, A. Y. (2018). Sparse principal component analysis via variable projection. *arXiv preprint arXiv:1804.00341*.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2), 301-320.

See Also

[spca](#), [Plot_Biplot](#)

Examples

```
ElasticNet_HJBiplot(mtcars, Lambda = 0.2, Alpha = 0.1)
```

HJBiplot

HJ Biplot

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. Options are: "center" that removes the columns means and "scale" that removes the columns means and divide by its standard deviation. Default 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. |

Author(s)

Mitzi Cubilla-Montilla, Carlos Torres-Cubilla, Ana Belen Nieto Librero and Purificacion Galindo Villardon

References

- Gabriel, K. R. (1971). The Biplot graphic display of matrices with applications to principal components analysis. *Biometrika*, 58(3), 453-467.
- Galindo, M. P. (1986). Una alternativa de representacion simultanea: HJ-Biplot. *Questiio*, 10(1), 13-23.

See Also

[Plot_Biplot](#)

Examples

```
HJBiplot(mtcars)
```

LASSO_HJBiplot

LASSO HJ Biplot

Description

This function performs the representation of the SPARSE HJ Biplot applying the LASSO regularization, on the original data matrix, implementing the norm L1.

Usage

```
LASSO_HJBiplot(X, Lambda, Transform.Data = 'scale', Operator = 'Hard-Thresholding')
```

Arguments

| | |
|----------------|---|
| X | array_like; A data frame which provides the data to be analyzed. All the variables must be numeric. |
| Lambda | float; Tuning parameter for the LASSO penalty |
| Transform.Data | character; A value indicating whether the columns of X (variables) should be centered or scaled. Options are: "center" that removes the columns means and "scale" that removes the columns means and divide by its standard deviation. Default is "scale". |
| Operator | character; The operator used to solve the norm L1. Allowed values are "Soft-Thresholding" and "Hard-Thresholding". |

Details

Algorithm that performs a procedure of contraction and selection of variables. LASSO imposes a penalty that causes the charges of some components to be reduced to zero. By producing zero loadings for some components and not zero for others, the Lasso technique performs selection of variables. As the value of the penalty approaches one, the loadings approach zero.

Value

LASSO_HJBiplot returns a list containing the following components:

| | |
|-------------|---|
| loadings | array_like; penalized loadings, the loadings of the sparse principal components. |
| n_ceros | array_like; number of loadings equal to zero in each component. |
| coord_ind | array_like; matrix with the coordinates of individuals. |
| coord_var | array_like; matrix with the coordinates of variables. |
| eigenvalues | array_like; vector with the eigenvalues penalized. |
| explvar | array_like; an vector containing the proportion of variance explained by the first 1, 2,..,k sparse principal components obtained. |

Author(s)

Mitzi Cubilla-Montilla, Carlos Torres-Cubilla, Ana Belen Nieto Librero and Purificacion Galindo Villardon

References

- Galindo, M. P. (1986). Una alternativa de representacion simultanea: HJ-Biplot. *Questiio*, 10(1), 13-23.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3), 273-282.

See Also

[Plot_Biplot](#)

Examples

```
LASSO_HJBiplot(mtcars, Lambda = 0.2, Operator = 'Hard-Thresholding')
```

Plot_Biplot

Plotting Biplot

Description

Plot_Biplot initializes a ggplot2-based visualization of the characteristics presented in the data analyzed 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 length 2 which contains the axis plotted 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 is TRUE (default) it prints the vector names with orientation of the angle of the vector. If it is FALSE the angle of all tags is 0. |

Value

Return a `ggplot2` object.

Author(s)

Mitzi Cubilla-Montilla, Carlos Torres-Cubilla, Ana Belen Nieto Librero and Purificacion Galindo Villardon

See Also

[HJBiplot](#), [Ridge_HJBiplot](#), [ElasticNet_HJBiplot](#)

Examples

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

| | |
|----------------|------------------------|
| Ridge_HJBiplot | <i>Ridge HJ Biplot</i> |
|----------------|------------------------|

Description

This function performs the representation of the HJ Biplot applying the Ridge regularization, on the original data matrix, implementing the norm L2.

Usage

```
Ridge_HJBiplot (X, Lambda, Transform.Data = 'scale')
```

Arguments

| | |
|----------------|---|
| X | array_like; A data frame which provides the data to be analyzed. All the variables must be numeric. |
| Lambda | float; Tuning parameter for the Ridge penalty |
| Transform.Data | character; A value indicating whether the columns of X (variables) should be centered or scaled. Options are: "center" that removes the columns means and "scale" that removes the columns means and divide by its standard deviation. Default is "scale". |

Details

Algorithm used to contract the loads of the main components towards zero, but without achieving the nullity of any. If the penalty parameter is less than or equal to $1e-4$ the result is like Galindo's HJ Biplot (1986).

Value

Ridge_HJBiplot returns a list containing the following components:

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

Author(s)

Mitzi Cubilla-Montilla, Carlos Torres-Cubilla, Ana Belen Nieto Librero and Purificacion Galindo Villardon

References

- Galindo, M. P. (1986). Una alternativa de representacion simultanea: HJ-Biplot. *Questiio*, 10(1), 13-23.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2), 265-286.

See Also

[Plot_Biplot](#)

Examples

```
Ridge_HJBiplot(mtcars, Lambda = 0.2)
```

Index

ElasticNet_HJBiplot, [2](#), [8](#)

ggplot2, [8](#)

HJBiplot, [4](#), [8](#)

LASSO_HJBiplot, [5](#)

Plot_Biplot, [3](#), [5](#), [6](#), [7](#), [9](#)

Ridge_HJBiplot, [8](#), [8](#)

spca, [3](#)