

Package ‘SparseBiplots’

July 2, 2019

Type Package

Title 'HJ Biplot' using Different Ways of Penalization

Version 3.5.0

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

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

Description Contains a set of functions that allow to represent multivariate on a subspace of low dimension, in such a way that most of the variability of the information is captured. This representation is carried out through the 'HJ Biplot' methodology. A first method performs the 'HJ Biplot'. Then, the package implements three new techniques and constructs in each case the 'HJ Biplot', adapting restrictions to contract and / or produce zero charges in the main components, using three methods of regularization: Ridge, LASSO and Elastic Net.

License GPL (>= 3)

Encoding UTF-8

LazyData true

Imports sparsepca

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-07-02 16:30:13 UTC

R topics documented:

| | |
|-------------------------------|---|
| ElasticNet_HJBiplot | 2 |
| HJBiplot | 3 |
| LASSO_HJBiplot | 5 |
| Ridge_HJBiplot | 6 |

| | |
|--------------|----------|
| Index | 9 |
|--------------|----------|

ElasticNet_HJBiplot *Elastic Net HJ Biplot*

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',
                    ind_name = FALSE, vec_name = TRUE)
```

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" or "scale". For default is "scale". |
| ind_name | bool; 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. |
| vec_name | bool; 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. |

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, 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.

Examples

```
data(mtcars)
ElasticNet_HJBiplot(mtcars, 0.2, 0.1, transform_data = 'scale', ind_name=TRUE)
```

HJBiplot

HJ Biplot

Description

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

Usage

```
HJBiplot (X, transform_data = 'scale', ind_name=FALSE,
          vec_name = TRUE)
```

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. For default is "scale". |
| ind_name | bool; If it is TRUE it prints the name for each row of X. If it is FALSE (default) does not print the names. |
| vec_name | bool; If it is TRUE (default) it prints the name for each column of X. If it FALSE does not print the names. |

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:

| | |
|-------------|--|
| 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. |
| 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. |

Author(s)

Mitzi Cubilla-Montilla, Carlos Torres, 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.

Examples

```
data(mtcars)
HJBiplot(mtcars, transform_data = 'scale', ind_name = TRUE)
```

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',
               ind_name=FALSE, vec_name = TRUE)
```

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. For default is "scale". |
| operator | character; The operator used to solve the norm L1. |
| ind_name | bool; If it is TRUE it prints the name for each row of X. If it is FALSE (default) does not print the names. |
| vec_name | bool; If it is TRUE (default) it prints the name for each column of X. If it is FALSE does not print the names. |

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, 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.

Examples

```
data(mtcars)
LASSO_HJBiplot(mtcars, 0.2, transform_data = 'scale', operator = 'Hard-Thresholding',
ind_name = TRUE)
```

Ridge_HJBiplot

Ridge HJ Biplot

Description

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

Usage

```
Ridge_HJBiplot (X, lambda, transform_data = 'scale', ind_name=FALSE,
               vec_name = TRUE)
```

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. For default it is "scale". |
| ind_name | bool; If it is TRUE it prints the name for each row of X. If it is FALSE (default) does not print the names. |
| vec_name | bool; If it is TRUE (default) it prints the name for each column of X. If it FALSE does not print the names. |

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:

| | |
|-------------|---|
| 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. |
| 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, Ana Belen Nieto Librero and Purificacion Galindo Villardon

References

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

Examples

```
data(mtcars)
Ridge_HJBiplot(mtcars, 0.2, transform_data = 'scale', ind_name = TRUE)
```


Index

ElasticNet_HJBiplot, [2](#)

HJBiplot, [3](#)

LASSO_HJBiplot, [5](#)

Ridge_HJBiplot, [6](#)