Package: VDSM (via r-universe)

September 16, 2024

Type Package

Title Visualization of Distribution of Selected Model

Version 0.1.1

Description Although model selection is ubiquitous in scientific
discovery, the stability and uncertainty of the selected model
is often hard to evaluate. How to characterize the random
behavior of the model selection procedure is the key to
understand and quantify the model selection uncertainty. This R
package offers several graphical tools to visualize the
distribution of the selected model. For example, Gplot(),
Hplot(), VDSM_scatterplot() and VDSM_heatmap(). To the best of
our knowledge, this is the first attempt to visualize such a
distribution. About what distribution of selected model is and
how it work please see Qin,Y.and Wang,L. (2021) ``Visualization
of Model Selection Uncertainty''
<https://homepages.uc.edu/~qinyn/VDSM/VDSM.html>.

License GPL (>= 2)

Encoding UTF-8

LazyData true

Imports ggplot2, plyr, dplyr, grid, viridis, gridExtra, knitr, stats

RoxygenNote 7.1.1

Depends R (>= 3.5.0)

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Linna Wang [aut, cre], Yichen Qin [aut]

Maintainer Linna Wang <wang2l9@mail.uc.edu>

Repository CRAN

Date/Publication 2021-04-16 09:00:02 UTC

Contents

CheckInput   *Check if the input is valid or not*

Description

Input a valid matrix

Usage

```
CheckInput(X, f, p)
```

Arguments

X	A m*p matrix which each row represents one unique model with the elements either 0 or 1.
f	A vector with m elements contain each model's frequency in X.
p	The number of variate in the model

Value

The standardized matrix

DSM_plot	*DSM_plot plot the naive visualization of the distribution of selected model*

Description

DSM_plot plot the naive visualization of the distribution of selected model

Usage

```
DSM_plot(
  X,
  f,
  p,
  Anchor.model = NULL,
  circlesize = NULL,
  linewidth = NULL,
  fontsize = NULL
)
```

Arguments

X	A m*p matrix which contains m different p-dimensional models. All the elements are either 0 or 1.
f	A vector with m elements which represent each model's frequency in X.
p	The number of variate in the model
Anchor.model	A vector containing p elements with either 1 or 0 value and must be found in X. Default is the model with the highest frequency.
circlesize	customize the size of the circle in the plot, default is 10.
linewidth	Customize the width of the line in the plot, default is 1.
fontsize	Customize the size of the font in the circles, default is 1.5.

Value

A summarized information of the grouped models.

Examples

```
data(exampleX)
X=exampleX
data(examplef)
f=examplef
p=8
DSM_example1 = DSM_plot(X,f,p)
```

---

examplef   *examplef*

Description

This small data set contains the frequencies of thoes m=30 models in exampleX data set.

Usage

examplef

Format

One vector representing the information of f.

---

exampleX   *exampleX*

Description

This small data set contains m=30 unique models and p=8 variates.

Usage

exampleX

Format

One matrix containing the information of X.

Gplot   *Gplot.*

Description

Plotting Gplot.

Usage

```
Gplot(
  X,
  f,
  p,
  Anchor.model = NULL,
  xlim = NULL,
  ylim = NULL,
  circlesize = NULL,
  linewidth = NULL,
  fontsize = NULL
)
```

Arguments

X	A m*p matrix which contains m different p-dimensional models. All the elements are either 0 or 1.
f	A vector with m elements which represent each model's frequency in X.
p	The number of variate in the model.
Anchor.model	A vector containing p elements with either 1 or 0 value and must be found in X. Default is the model with the highest frequency.
xlim	A vector with two elements which determine the range of x-axis in the plot.
ylim	A vector with two elements which determine the range of y-axis in the plot.
circlesize	customize the size of the circle in the plot, default is 10.
linewidth	Customize the width of the line in the plot, default is 1.
fontsize	Customize the size of the font in the circles, default is 1.5.

Value

A list with components

Gplot.info	The table includes all the information about each group, i.e., the total possible number of models in the group and the actual existing number of model in the group.
MC.histogram	The frequency of model complexity.
HD.histogram	The frequency of Hamming distance.

Examples

```
data(exampleX)
X=exampleX
data(examplef)
f=examplef
p=8
G_example1 = Gplot(X,f,p)
G_example2 = Gplot(X,f,p,xlim=c(0,7),ylim=c(3,8))
G_example3 = Gplot(X,f,p,xlim=c(0,7),ylim=c(3,8),circlesize=15,linewidth=2,fontsize=3)
```

Groupinfo	*Group the models according to their Hamming distance and Model complexity to the anchor model*

Description

Group the given models

Usage

```
Groupinfo(X, f, p, Anchor.model = NULL)
```

Arguments

X	A m*p matrix which contains m different p-dimensional models. All the elements are either 0 or 1.
f	A vector with m elements which represent each model's frequency in X.
p	The number of variate in the model
Anchor.model	A vector containing p elements with either 1 or 0 value and must be found in X. Default is the model with the highest frequency.

Value

A summarized information of the grouped models.

Hplot	*Hplot.*

Description

Plotting Hplot.

Usage

```
Hplot(
  X,
  f,
  p,
  Anchor.model = NULL,
  xlim = NULL,
  ylim = NULL,
  circlesize = NULL,
  linewidth = NULL,
  fontsize = NULL
)
```

Arguments

X	A m*p matrix which contains m different p-dimensional models. All the elements are either 0 or 1.
f	A vector with m elements which represent each model's frequency in X.
p	The number of variate in the model.
Anchor.model	A vector containing p elements with either 1 or 0 value and must be found in X. Default is the model with the highest frequency.
xlim	A vector with two elements which determine the range of x-axis in the plot.
ylim	A vector with two elements which determine the range of y-axis in the plot.
circlesize	customize the size of the circle in the plot, default is 10.
linewidth	Customize the width of the line in the plot, default is 1.
fontsize	Customize the size of the font in the circles, default is 1.5.

Value

A list with components

Hplot.info	The table includes all the information about each group, i.e., the total possible number of models in the group and the actual existing number of model in the group.
Hplus.histogram	
	The frequency of Hamming distance plus.
Hminus.histogram	
	The frequency of Hamming distance minus.

Examples

```
data(exampleX)
X=exampleX
data(examplef)
f=examplef
p=8
H_example1 = Hplot(X,f,p)
H_example2 = Hplot(X,f,p,xlim=c(0,4),ylim=c(0,2))
H_example3 = Hplot(X,f,p,xlim=c(0,4),ylim=c(0,2),circlesize=15,linewidth=2,fontsize=3)
```

VDSM_heatmap	*VDSM-heatmap.*

Description

Plotting the VDSM-heatmap.

Usage

```
VDSM_heatmap(
  X,
  f,
  p,
  Anchor.estimate,
  xlim = NULL,
  ylim = NULL,
  Anchor.model = NULL,
  fontsize = NULL
)
```

Arguments

X	A m*p matrix which contains m different p-dimensional models. All the elements are either 0 or 1.
f	A vector with m elements which represent each model's frequency in X.
p	The number of variate in the model.
Anchor.estimate	
	An estimation for the anchor model.
xlim	A vector with two elements which determine the range of x-axis in the plot.
ylim	A vector with two elements which determine the range of y-axis in the plot.
Anchor.model	A vector containing p elements with either 1 or 0 value and must be found in X. Default is the model with the highest frequency.
fontsize	Customize the size of the font in the circles, default is 1.5.

Value

A list with components

Heatmap.info	The table includes all the information about each group, i.e., the total possible number of models in the group and the actual existing number of model in the group.
Hplus.histogram	
	The frequency of Hamming distance plus.
Hminus.weighted.histogram	
	The frequency of Hamming distance minus-weighted.

Examples

```
data(exampleX)
X=exampleX
data(examplef)
f=examplef
p=8
Anchor.estimate=c(3,2.5,2,1.5,1,0,0,0)
Heatmap_example1 = VDSM_heatmap(X,f,p,Anchor.estimate)
```

```
Heatmap_example2 = VDSM_heatmap(X,f,p,Anchor.estimate,fontsize=3)
Heatmap_example3 = VDSM_heatmap(X,f,p,Anchor.estimate,xlim=c(0,5),ylim=c(0,5),fontsize=3)
```

VDSM_scatterplot   *VDSM-Scatterplot.*

Description

Plotting the VDSM-Scatterplot.

Usage

```
VDSM_scatterplot(
  X,
  f,
  p,
  Anchor.estimate,
  xlim = NULL,
  ylim = NULL,
  Anchor.model = NULL,
  circlesize = NULL,
  fontsize = NULL
)
```

Arguments

X	A m*p matrix which contains m different p-dimensional models. All the elements are either 0 or 1.
f	A vector with m elements which represent each model's frequency in X.
p	The number of variate in the model.
Anchor.estimate	
	An estimation for the anchor model.
xlim	A vector with two elements which determine the range of x-axis in the plot.
ylim	A vector with two elements which determine the range of y-axis in the plot.
Anchor.model	A vector containing p elements with either 1 or 0 value and must be found in X. Default is the model with the highest frequency.
circlesize	customize the size of the circle in the plot, default is 10.
fontsize	Customize the size of the font in the circles, default is 1.5.

Value

A list with components

Scatterplot.info

> The table includes all the information about each group, i.e., the total possible number of models in the group and the actual existing number of model in the group.

```
Hplus.histogram
                    The frequency of Hamming distance plus.
Hminus.weighted.histogram
                    The frequency of Hamming distance minus-weighted.
```

Examples

```
data(exampleX)
X=exampleX
data(examplef)
f=examplef
p=8
Anchor.estimate=c(3,2.5,2,1.5,1,0,0,0)
Scatter_example1 = VDSM_scatterplot(X,f,p,Anchor.estimate)
Scatter_example2 = VDSM_scatterplot(X,f,p,Anchor.estimate,xlim=c(0,5),
ylim=c(0,8),circlesize=15,fontsize=2)
```

VDSM_scatter_heat   *VDSM-Scatter-heatmap-info*

Description

Report VDSM-Scatter-heatmap-infomation

Usage

```
VDSM_scatter_heat(X, f, p, Anchor.estimate, Anchor.model = NULL)
```

Arguments

X	A m*p matrix which contains m different p-dimensional models. All the elements are either 0 or 1.
f	A vector with m elements which represent each model's frequency in X.
p	The number of variate in the model
Anchor.estimate	
	An estimation for the anchor model
Anchor.model	A vector containing p elements with either 1 or 0 value and must be found in X. Default is the model with the highest frequency.

Value

A list of information which helps to plot VDSM-Scatter-heatmap.

Index