# Package ‘VDSM’

April 16, 2021

<table>
<thead>
<tr>
<th>Type</th>
<th>Package</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Visualization of Distribution of Selected Model</td>
</tr>
<tr>
<td>Version</td>
<td>0.1.1</td>
</tr>
<tr>
<td>Description</td>
<td>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” <a href="https://homepages.uc.edu/~qinyn/VDSM/VDSM.html">https://homepages.uc.edu/~qinyn/VDSM/VDSM.html</a>.</td>
</tr>
<tr>
<td>License</td>
<td>GPL (&gt;= 2)</td>
</tr>
<tr>
<td>Encoding</td>
<td>UTF-8</td>
</tr>
<tr>
<td>LazyData</td>
<td>true</td>
</tr>
<tr>
<td>Imports</td>
<td>ggplot2, plyr, dplyr, grid, viridis, gridExtra, knitr, stats</td>
</tr>
<tr>
<td>RoxygenNote</td>
<td>7.1.1</td>
</tr>
<tr>
<td>Depends</td>
<td>R (&gt;= 3.5.0)</td>
</tr>
<tr>
<td>Suggests</td>
<td>testthat (&gt;= 3.0.0)</td>
</tr>
<tr>
<td>Config/testthat/edition</td>
<td>3</td>
</tr>
<tr>
<td>NeedsCompilation</td>
<td>no</td>
</tr>
<tr>
<td>Author</td>
<td>Linna Wang [aut, cre], Yichen Qin [aut]</td>
</tr>
<tr>
<td>Maintainer</td>
<td>Linna Wang <a href="mailto:wang219@mail.uc.edu">wang219@mail.uc.edu</a></td>
</tr>
<tr>
<td>Repository</td>
<td>CRAN</td>
</tr>
<tr>
<td>Date/Publication</td>
<td>2021-04-16 09:00:02 UTC</td>
</tr>
</tbody>
</table>

**R topics documented:**

CheckInput ................................. 2
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)

Description

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

Usage

examplef
**Format**

One vector representing the information of f.

---

**Description**

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

**Usage**

```r
exampleX
```

**Format**

One matrix containing the information of X.

---

**Description**

Plotting Gplot.

**Usage**

```r
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)

Description

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

Usage

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

X  A \( m \times p \) 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 \( m \times p \) 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)

**Description**

Plotting the VDSM-heatmap.

**Usage**

```r
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.
VDSM_scatterplot

Description

Plotting the VDSM-Scatterplot.

Usage

VDSM_scatterplot(
  X,
  f,
  p,
  Anchor.estimate,
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

```r
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)
```
Description

Report VDSM-Scatter-heatmap-infoation

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

* datasets
  examplef, 3
  exampleX, 4
CheckInput, 2
DSM_plot, 2
examplef, 3
exampleX, 4
Gplot, 4
Groupinfo, 5
Hplot, 6
VDSM_heatmap, 7
VDSM_scatter_heat, 10
VDSM_scatterplot, 8