

Package ‘GMD’

September 22, 2011

Version 0.2

Date 2011-09-22

Title Generalized Minimum Distance of distributions

Author Xiaobei Zhao <xiaobei@binf.ku.dk>, Albin Sandelin <albin@binf.ku.dk>

Published

Maintainer Xiaobei Zhao <xiaobei@binf.ku.dk>

Description GMD is a package for non-parametric distance measurement between two discrete frequency distributions.

Depends R (>= 2.9.0), tools

License GPL (>=2)

URL

Repository CRAN

Type Package

LazyLoad yes

Collate 'zzz.R' 'GMD-package.R' 'GMD-internal.R' 'gmd.R' 'gmdm.R' 'cage.R'

R topics documented:

GMD-package	2
cage	3
gmd	3
gmdm	5
plot.gmd	6
plot.gmdm	8
print.gmd	9
summary.gmd	10

Index	11
--------------	-----------

Description

Compute Generalized Minimum Distance (GMD) between discrete distributions

Details

Package: GMD
 Type: Package
 Version: 0.2
 Date: Thu Sep 22 2011
 License: GPL (>= 2)

This package contains functions for GMD computation, with GMD algorithm implemented in C to interface with R.

To install from online repositories (e.g. CRAN),

```
install.packages(pkgs="GMD", repos="http://cran.r-project.org")
```

To install from a downloaded source file,

```
install.packages(pkgs="GMD_<current-version>.tar.gz", repos=NULL)
```

For a complete list of functions, use

```
library(GMD); ls("package:GMD")
```

Author(s)

Xiaobei Zhao and Albin Sandelin

Maintainer: Xiaobei Zhao <xiaobei (at) binf.ku.dk>

References

Zhao et al (2011), "Systematic Clustering of Transcription Start Site Landscapes", *PLoS ONE* 6(8): e23409. <http://dx.plos.org/10.1371/journal.pone.0023409>

See `citation("GMD")` for BibTeX entries for LaTeX users.

See Also

[gmd](#), [gmdm](#), [cage](#)

Examples

```
require(GMD) # load GMD
help(GMD) # a help document of GMD
data(package="GMD") # a list of datasets available in GMD
ls("package:GMD") # a list of functions available in GMD
citation("GMD") # for citation
demo("GMD-demo") # run the demo
```

`cage`*CAGE Data*

Description

Transcription Start Site Distributions (TSSDs) by CAGE tags.

Usage

```
cage
cagel
```

Details

`cage` is a list of 8 named TSSDs. `cagel` is a longer version of `cage`, with 50 named TSSDs.

References

Zhao et al (2011), "Systematic Clustering of Transcription Start Site Landscapes", *PLoS ONE* **6**(8): e23409.

<http://dx.plos.org/10.1371/journal.pone.0023409>

See Also

`gmd` and `gm dm`, with examples using `cage`

Examples

```
require(GMD)
data(cage)
print(cage)
## Not run: data(cagel)
## Not run: cagel
```

`gmd`*Generalized Minimum Distance (GMD)*

Description

Generalized Minimum Distance

Usage

```
gmd(v1, v2, labels=c("v1", "v2"), pseudocount=0, sliding=TRUE)
```

Arguments

<code>v1</code>	a numeric vector, giving positional counts as a discrete distribution.
<code>v2</code>	a numeric vector, giving positional counts as a discrete distribution.
<code>labels</code>	a string vector of length 2, giving the names of <code>v1</code> and <code>v2</code> respectively.
<code>pseudocount</code>	a numeric value to be allocated for each position to reduce bias; by default <code>pseudocount = 0</code> .
<code>sliding</code>	logical, indicating whether sliding is allowed or not for an optimal solution; by default <code>sliding = TRUE</code> .

Details

Generalized Minimum Distance

Value

`gmd` returns an object of class `gmd`, a list with components

- `labels`: a string vector, giving the names of distributions
- `v1.ori`: a numeric vector, the first input distribution
- `v2.ori`: a numeric vector, the second input distribution
- `v1`: a numeric vector, the normalized version of the first input distribution
- `v2`: a numeric vector, the normalized version of the second input distribution
- `distance`: numeric, the *GM-Distance (GMD)*
- `sliding`: logical, indicating whether sliding is performed
- `pseudocount`: a numeric value that is allocated at each position in addition to original values
- `gap.pair`: a numeric matrix, giving one gap pair per row: i.e. relative shifts between distributions of one optimal hit
- `n.hit`: numeric, the number of (equally good) optimal hits

References

See `citation("GMD")`

See Also

[print.gmd](#), [summary.gmd](#), [plot.gmd](#), [gmdm](#)

Examples

```
require(GMD)
gmd(c(4,1,1,0,0,0,3,1),c(2,1,1,0,0,0,3,3),sliding=FALSE)
x <- gmd(c(4,1,1,0,0,0,3,1), c(1,1,2,1,1,0,0,3,3,5,5),
pseudocount=1, sliding=TRUE)
print(x)
print(x, "full")
```

gmdm*Generalized Minimum Distance Matrix*

Description

Generalized Minimum Distance Matrix

Usage

```
gmdm(x, labels=names(x), pseudocount=0, sliding=TRUE)
```

Arguments

x	a list of numeric vectors
labels	a character vector of the same length of x, giving the names of the numeric vectors.
pseudocount	a numeric value to be allocated for each position to reduce bias; by default pseudocount = 0.
sliding	logical, indicating whether sliding is allowed or not for an optimal solution; by default sliding = TRUE.

Details

Generalized Minimum Distance Matrix

Value

gmdm returns an object of class gmdm, a list with components

- labels: a string vector, giving the names of distributions
- data.ori: a list of numeric vectors, giving the original input
- data: a list of numeric vectors, giving the normalized version of the original input
- dm: a numeric matrix, the pairwise distance matrix of *GM-Distances*
- gap.pair: a numeric matrix, giving the gap pair of each alignment per row: i.e. relative shifts between distributions of the optimal hit
- sliding: logical, indicating whether sliding is performed
- pseudocount: a numeric value that is allocated at each position in addition to original values

References

See `citation("GMD")`

See Also

[plot.gmdm](#), [gmd](#)

Examples

```
require(GMD)
data(cage)
x <- gmdm(cage)
print(x$labels)
print(x$dm)

## Not run: data(cage1)
x <- gmdm(cage1)
head(x$labels)
head(x$dm)
## End(Not run)
```

plot.gmd

Plot Function for Class gmd

Description

Plot Function for Class gmd

Usage

```
plot.gmd(x, labels=x$labels, colors=c("red", "blue"),
  plot.method=c("separate", "overlay"), plot.type="h", main,
  ylab="Fraction", xlab="Position", ylim, xlim, font.type=1,
  font.family=c("sans", "serif", "mono"), cex.lab=1.2,
  cex.tickmark=1, cex.legend=1.5, lwd.line=1, if.plot.new=TRUE,
  if.text.gmd=TRUE, if.text.gap=FALSE, if.plot.gap=TRUE,
  if.plot.legned=TRUE, x.jitter=ifelse(plot.method == "overlay",
  1/1000, 0), ...)
```

Arguments

<code>x</code>	an object of class <code>gmd</code> .
<code>labels</code>	a string vector of the same length of <code>x\$labels</code> , giving the names of the numeric vectors in <code>x</code> .
<code>colors</code>	the colors of the discrete distributions; by default they are in "red" and "blue".
<code>plot.method</code>	the plot method. This can be specified as a string: "separate": means separated subplots [<i>default</i>]; "overlay": means overlaid subplots.
<code>plot.type</code>	the plot type. See <code>type</code> in <code>plot</code> for possible values; the default <code>plot.type</code> = "h", giving 'histogram' like vertical lines.
<code>main</code>	an overall title for the plot. See <code>help("title", package="graphics")</code> .
<code>ylab</code>	a title for the y axis. See <code>help("title", package="graphics")</code> .
<code>xlab</code>	a title for the x axis. See <code>help("title", package="graphics")</code> .
<code>ylim</code>	range of y values, as in <code>help("plot", package="graphics")</code> .
<code>xlim</code>	range of x values, as in <code>help("plot", package="graphics")</code> .
<code>font.type</code>	the name of a font type for drawing text. See <code>font</code> in <code>par</code> ; the default <code>font.type</code> = 1, corresponding to plain text.

font.family	the name of a font family for drawing text. See family in par; the default font.family = "sans", corresponding to san serif typeface.
cex.lab	a numerical value giving the amount by which xlab and ylab should be magnified relative to the default.
cex.tickmark	a numerical value giving the amount by which tickmarks should be magnified relative to the default.
cex.legend	a numerical value giving the amount by which legends should be magnified relative to the default.
lwd.line	the line width, a <i>positive</i> number, defaulting to 1.
if.plot.new	logical, indicating whether to start a new plot device.
if.text.gmd	logical, indicating whether <i>GM-Distance</i> is reported in the subtitle.
if.text.gap	logical, indicating whether <i>gap</i> is reported in the subtitle.
if.plot.gap	logical, indicating whether <i>gap</i> is plotted.
if.plot.legned	logical, indicating whether <i>legend</i> is plotted.
x.jitter	numeric, indicating how <i>jitter</i> should be added to distinguish subplots; by default x.jitter=ifelse(plot.method=="overlay", 1/1000, 0) giving how jitter should be adjusted according to the <i>x-axis range</i> .
...	arguments to be passed to methods, such as graphical parameters (see par).

Details

Plot Function for Class gmd

References

See help(GMD)

See Also

[gmd](#)

Examples

```
require(GMD)
data(cage)
## Not run: plot(gmd(cage[[1]],cage[[2]],labels=names(cage)[c(1,2)],
pseudocount=1, sliding=TRUE))
## End(Not run)
plot(gmd(cage[[1]],cage[[3]],labels=names(cage)[c(1,3)],
pseudocount=1, sliding=TRUE))
plot(gmd(cage[[1]],cage[[3]],labels=names(cage)[c(1,3)],
pseudocount=1, sliding=TRUE), plot.method="overlay")
```

plot.gmdm

*Plot Function for Class gmdm***Description**

Plot Function for Class gmdm

Usage

```
plot.gmdm(x, labels=x$labels, colors, plot.type="h", main,
          ylab="Fraction", xlab="Position", label.length.max=8,
          label.line.max=3, cex.text=2, cex.tickmark=0.75, if.plot.new=TRUE,
          x.jitter=1/1000, ...)
```

Arguments

<code>x</code>	an object of class <code>gmdm</code> .
<code>labels</code>	a string vector of the same length as <code>x\$data</code> , giving the names of the numeric vectors in <code>x\$data</code> .
<code>colors</code>	the colors of the discrete distributions; the default is <i>"Dark2" colors in ColorBrewer palettes</i> if not specified.
<code>plot.type</code>	the plot type. See <code>type</code> in <code>plot</code> for possible values; the default <code>plot.type = "h"</code> , giving 'histogram' like vertical lines.
<code>main</code>	an overall title for the plot. See <code>help("title", package="graphics")</code> ; the default title is used if not specified.
<code>ylab</code>	a title for the y axis. See <code>help("title", package="graphics")</code> .
<code>xlab</code>	a title for the x axis. See <code>help("title", package="graphics")</code> .
<code>label.length.max</code>	numeric, giving the maximum string width allowed in diagonal labels.
<code>label.line.max</code>	numeric, giving the maximum number of lines allowed in diagonal labels.
<code>cex.text</code>	a numerical value giving the amount by which plot text should be magnified relative to the default.
<code>cex.tickmark</code>	a numerical value giving the amount by which tickmarks should be magnified relative to the default.
<code>if.plot.new</code>	logical, indicating whether to start a new plot device.
<code>x.jitter</code>	numeric, indicating how <i>jitter</i> should be added to distinguish subplots; by default <code>x.jitter=1/1000</code> indicating the jitter is adjusted to 1/1000 of the <i>x-axis range</i> .
<code>...</code>	arguments to be passed to methods, see <code>gmd</code> .

Details

Plot Function for Class gmdm

ReferencesSee `help(GMD)`

See Also

[gmdm](#), [gmd](#)

Examples

```
require(GMD)
data(cage)
plot(gmdm(cage))
```

print.gmd

Print Function for Class gmd

Description

Print Function for Class gmd

Usage

```
print.gmd(x, print.mode=c("brief", "full"), digits=3, ...)
```

Arguments

x	an object of class gmd.
print.mode	a string, indicating whether to print in <i>full</i> mode (<i>default</i>).
digits	integer, indicating the number of decimal places to be printed.
...	arguments to be passed to methods, see <code>print</code> .

Details

Print Function for Class gmd

References

See `help(GMD)`

See Also

[gmd](#)

summary.gmd

Summary Function for Class gmd

Description

Summary Function for Class gmd

Usage

```
summary.gmd(object, ...)
```

Arguments

object	an object of class gmd.
...	arguments to be passed to methods, see <code>summary</code> .

Details

Summary Function for Class gmd

References

See `help(GMD)`

See Also

[gmd](#)

Index

*Topic **classes**

`gmd`, [3](#)

`gmdm`, [5](#)

*Topic **datasets**

`cage`, [3](#)

*Topic **hplot**

`plot.gmd`, [6](#)

`plot.gmdm`, [8](#)

*Topic **methods**

`plot.gmd`, [6](#)

`plot.gmdm`, [8](#)

*Topic **package**

`GMD-package`, [2](#)

`cage`, [2](#), [3](#), [3](#)

`cage1 (cage)`, [3](#)

GMD (GMD-package), [2](#)

`gmd`, [2](#), [3](#), [3](#), [5](#), [7](#), [9](#), [10](#)

`GMD-package`, [2](#)

`gmdm`, [2-4](#), [5](#), [9](#)

`plot.gmd`, [4](#), [6](#)

`plot.gmdm`, [5](#), [8](#)

`print.gmd`, [4](#), [9](#)

`summary.gmd`, [4](#), [10](#)