R documentation

of 'DensityResponse.Rd' etc.

March 30, 2014

DensityResponse *Plot Distribution of the Response Variable*

Description

This function creates plots the distribution of the response variable.

Usage

```
DensityResponse(Data, xlab = "", ylab = "", main = "", alpha = 0.2, binwidth = NULL, histFill =
```

Arguments

Data	A numeric vector.
xlab	Title of the x axis.
ylab	Title of the y axis.
main	Title of the plot.
alpha	Alpha for the fill color of the distribution. Default value 0.2.
binwidth	Width of the histogram bins. Default value NULL.
histFill	Fill color of the histogram bars. Default value 'white'.
histCol	Color of the histogram lines. Default value 'black'.
densityFill	Fill color of the distribution. Default value "#FF6666".
TitleSize	Title font size. Default value 15.
TextSize	Text size. Default value 15.
XAxisSize	Size of the text on the X axis. Default value 15.
YAxisSize	Size of the text on the Y axis. Default value 15.
AngleLab	Angle of the labels in the X axis. Default value 30.
LegendPosition	Position of the legend. Default value 'right'.
TitleAxesSize	Font size of the axes lables. Default value 15.
tmar	Top margin size. Default values is 1.
bmar	Bottom margin size. Default values is 1.
rmar	Right margin size. Default values is 1.
lmar	Left margin size. Default values is 1.

Details

Additional ggplot2 layers can be added with "+".

Value

Returns a ggplot object.

Author(s)

Isidro Cortes and Daniel Murrell.

Examples

```
## Example:

DensityResponse(rnorm(100),xlab = "Random values", ylab = "Frequency", main = "Exampl DensityResponse funct
    binwidth = NULL, histFill = "white", histCol = "black", densityFill = "#FF6666",
TitleSize = 15, TextSize = 15, XAxisSize = 15, YAxisSize = 15,
AngleLab = 30, LegendPosition = "right", TitleAxesSize = 15,
tmar = 1, bmar = 1, rmar = 1, lmar = 1
```

GetPropertiesSDF *Retrieval of Properties from .sdf file*

Description

The function retrieves the values of all fields in a .sdf file for a user-defined number of molecules.

Usage

```
GetPropertiesSDF(structures.file, number_processed = -1, type = 1)
```

Arguments

structures.file	File in .sdf format with the compounds.
number_processed	Number of molecules from which the properties are to be extracted. The function starts at the beginning of the .sdf file till the molecule which ordinal position is equal to the argument. The default value is -1. In that case, the properties are extracted for all molecules in the .sdf file.
type	1 corresponds to .sdf format, which is the only format currentyl supported.

Value

A data.frame with all properties which rows are indexed by the compounds in the .sdf file and columns by the properties.

Author(s)

Isidro Cortes and Daniel Murrell.

See Also

GetPropertySDF

`GetPropertySDF` *Retrieval of a given Property from a .sdf File.*

Description

The function extracts the values of one .sdf field for a user-defined number of molecules.

Usage

```
GetPropertySDF(structures.file, property = "", number_processed = -1, type = 1)
```

Arguments

`structures.file`	The .sdf file containing the molecules.
`property`	The property which values are to be extracted.
`number_processed`	Number of molecules for which the value are to be extracted. If this argument is set to -1 (default) the property will be extracted for all compounds.
`type`	-1 corresponds to .sdf file format, which is the only format currently supported.

Value

A one-field data.frame with the values of the property.

Author(s)

Isidro Cortes and Daniel Murrell.

See Also

GetPropertiesSDF

`MaxPerf` *Distribution of Maximum Theoretical Values Achievable given the Dataset and its Uncertainty*

Description

Calculates the ditribution of model validation metrics that are achievable given the size of the dataset, the uncertainty in the response variable, and the distribution of the responsible variable quantified by its mean and standard deviation. Therefore, these distributions help to assess models overfitting; e.g. a model trained on a dataset with high uncertainty exhibiting high correlation values might be overoptimistic.

Usage

```
MaxPerf(meanNoise = 0, sdNoise, meanResp, sdResp, lenPred, iters = 1000, filename = NULL, pdfW =
```

Arguments

meanNoise	Mean value of the noise in the data. Default value 0.
sdNoise	Standard deviation of the noise in the data. See the work by Kramer et al. about uncertainty in public bioactivity databases.
meanResp	Mean of the response variable.
sdResp	Standard deviation of the response variable.
lenPred	Number of datapoints of the external (hold-out) set.
iters	Number of iterations. Default value 1000.
filename	If not NULL, file where the plot will be saved. Default value NULL.
pdfW	Width of the .pdf file, in centimeters, where the plot will be saved. Default value 10.
pdfH	Height of the .pdf file, in centimeters, where the plot will be saved. Default value 10.
TextSize	Fontsize of the text in the plot. Default value 15.
TitleSize	Fontsize of the title. Default value 15.
XAxisSize	Fontsize of the X axis. Default value 15.
YAxisSize	Fontsize of the Y axis. Default value 15.
TitleAxesSize	Fontsize of the axes titles. Default value 15.
tmar	Top margin size. Default value 1.
bmar	Bottom margin size. Default value 1.
rmar	Right margin size. Default value 1.
lmar	Left margin size. Default value 1.
AngleLab	Angle of the labels of the X axis. Default value 30.
LegendPosition	Position of the legend. Default value 'right'.

Author(s)

Isidro Cortes and Daniel Murrell.

References

Cortes-Ciriano et al. 'Proteochemometric Modeling in a Bayesian Framework'. 2014

PairwiseDist *Pairwise Distance (Similarity) Matrix*

Description

The function is based on the vegdist function from the vegan package. It calculated the pairwise distance similarity matrix for all vectors input in a matrix or data.frame. The functions operates on a row basis.

Usage

```
PairwiseDist(Data, method = "jaccard", ..)
```

Arguments

Data	A numeric data.frame or matrix containing compound, protein or amino acid descriptors (or any combination thereof).
method	Available distance metrics are: "manhattan", "euclidean", "canberra", "bray", "kulczynski", "jaccard", "gower", "altGower", "morisita", "horn", "mountford", "raup", "binomial", "chao", "cao". See the documentation of the R package vegan for details.
..	

Details

For further detials see the documentation in the R package vegan.

Value

A data.frame with the all pairwise distances.

See Also

PairwiseDistPlot

Examples

```
m = matrix(abs(rnorm(20)),4,4)
mDist = PairwiseDist(m)
head(mDist)
```

PairwiseDistPlot *Distribution of Pairwise Similarities*

Description

The function depicts the distribution of pairwise similarities.

Usage

```
PairwiseDistPlot(Data, xlab = "", ylab = "", main = "", TextSize = 15, TitleSize = 15, XAxisSize
```

Arguments

Data	A data.frame with a single column named 'Distance'. This is the default output of PairwiseDist.
xlab	Label of the X axis.
ylab	Label of the Y axis.
main	Plot title.
TextSize	Fontsize of the text in the plot. Default value 15.
TitleSize	Fontsize of the title. Default value 15.
XAxisSize	Fontsize of the X axis. Default value 15.
YAxisSize	Fontsize of the Y axis. Default value 15.
TitleAxesSize	Fontsize of both the X and Y axes. Default value 15.
tmar	Top margin size. Default value 1.
bmar	Bottom margin size. Default value 1.
rmar	Right margin size. Default value 1.
lmar	Left margin size. Default value 1.
AngleLab	Angle of the labels in the X axis. Default value 30.
binwidth	Width of the bins of the hitogram. Default NULL, which corresponds to 1/30 of the range of the data (see ??stat_bin of ggplot2).
fillCol	Fill color of the histogram. Default value 'white'.
Colour	Line color of the histogram. Default value 'black'.
DensityFill	Fill color of the distribution. Default value "#FF6666".
DensityAlpha	Alpha for the fill color of the distribution. Default value 0.2.

Value

A ggplot2 object with the pairwise distance distribution.

Author(s)

Isidro Cortes and Daniel Murrell.

References

Package ggplot2.

See Also

PairwiseDist

Examples

```
m <- matrix(abs(rnorm(1600)),40,40)
mDist <- PairwiseDist(m)
head(mDist)
mDistPlot <- PairwiseDistPlot(mDist,xlab = "", ylab = "", main = "", TextSize = 15, TitleSize = 15, XAxisSi
```

`PlotMolecules` *Plot Compounds from a .sdf File.*

Description

The function plots the chemical structures provided in a .sdf file. The plots can also be sent to a 2x2 grid in a .pdf file.

Usage

```
PlotMolecules(sdf.file, IDs, pdf.file = NULL, PDFMain = NULL, useNameAsTitle = TRUE)
```

Arguments

`sdf.file`	The .sdf file with the molecules.
`IDs`	The IDs of the molecules to be depicted (the ordinal position of the molecules in the .sdf file). Currently, a maximum of four IDs is supported.
`pdf.file`	If not NULL, the .pdf where the molecules will be depicted.
`PDFMain`	If not NULL, the title of the molecule depiction in the .pdf file.
`useNameAsTitle`	If TRUE, the names of the molecules as especified in the .sdf file are used as molecules names in the depiction.

Value

A list with the plots of the molecules.

Author(s)

Isidro Cortes and Daniel Murrell.

SeqDescs *Whole Protein Sequence Descriptors*

Description

Calculation of the following 12 whole sequence protein descriptors:

- Amino Acid Composition ('AAC') - Dipeptide Composition ('DC') - Tripeptide Composition ('TC') - Normalized Moreau-Broto Autocorrelation ('MoreauBroto') - Moran Autocorrelation ('Moran') - Geary Autocorrelation ('Geary') - CTD (Composition/Transition/Distribution) ('CTD') - Conjoint Traid ('CTriad') - Sequence Order Coupling Number ('SOCN') - Quasi-sequence Order Descriptors ('QSO') - Pseudo Amino Acid Composition ('PACC') - Amphiphilic Pseudo Amino Acid Composition ('APAAC')

Usage

```
SeqDescs(Data, UniProtID = TRUE, type = "AAC", ..)
```

Arguments

Data	The function takes one or several protein sequences, or one or several UniProt IDs as argument.
UniProtID	If TRUE the argument calculates the descriptors for the proteins which UniProt IDs have been indicated in the argument 'Data'.
type	The type of protein descriptors to be calculated (see above). Any combination thereof is valid. A vector containing the abbreviation of the desired descriptors is taken as argument. Default value 'AAC'.
..	

Value

A numeric matrix which rows are indexed by proteins and the columns by descriptors. If multiple descriptors are chosen, the function returns a matrix where descriptors are concatenated per row for the ease of modeling.

Author(s)

Isidro Cortes and Daniel Murrell.

References

R package protr.

Examples

```
COX1_human <- SeqDescs("P23219",UniProtID=TRUE,type = c("AAC","Moran"))
COX_1_and_2_human = SeqDescs(c("P23219","P35354"),UniProtID=TRUE)
```

camb-package *camb - Chemistry Aware Model Builder ~~ camb ~~*

Description

A group of utilities to rapidly train and analyse models built to predict physicochemical properties and bioactivities of small molecules. TBD add more content here.

Details

Package:	camb
Type:	Package
Version:	0.1
Date:	2013-02-10
License:	GPL
LazyLoad:	yes

Author(s)

Daniel Murrell <dsmurrell@gmail.com> Isidro Cortes <isidrolauscher@gmail.com>

References

TBD: insert reference to published paper

expGrid *Exponential Grid Definition*

Description

The function defines an exponential series, which can be used, e.g. when defining the parameter space when training some models such as Support Vector Machines or Gaussian Processes.

Usage

```
expGrid(power.from, power.to, power.by, base)
```

Arguments

power.from	The starting exponential of the series.
power.to	The latest exponential of the series.
power.by	The exponential step of the series.
base	The base of the exponential series.

Value

A vector with the exponential series.

Author(s)

Isidro Cortes and Daniel Murrell.

Examples

```
expGrid(power.from=-10,power.to=10,power.by=2,base=10)
```

Index

∗Topic **\textasciitildeDensityResponse**
DensityResponse, 1
∗Topic **\textasciitildeGetPropertiesSDF**
GetPropertiesSDF, 2
∗Topic **\textasciitildeGetPropertySDF**
GetPropertySDF, 3
∗Topic **\textasciitildeMaxPerf**
MaxPerf, 3
∗Topic **\textasciitildeMaximumPerformance**
MaxPerf, 3
∗Topic **\textasciitildePairwiseDist**
PairwiseDist, 5
∗Topic **\textasciitildePlotMolecules**
PlotMolecules, 7
∗Topic **\textasciitildeSeqDescs**
SeqDescs, 8
∗Topic **\textasciitildeSequenceDescriptors**
SeqDescs, 8
∗Topic **\textasciitildeexpGrid**
expGrid, 9
∗Topic **\textasciitildekwd1**
PairwiseDistPlot, 6
∗Topic **\textasciitildekwd2**
PairwiseDistPlot, 6
PlotMolecules, 7
∗Topic **chemistry aware model building**
camb-package, 9

camb *(*camb-package*)*, 9
camb-package, 9

DensityResponse, 1

expGrid, 9

GetPropertiesSDF, 2
GetPropertySDF, 3

MaxPerf, 3

PairwiseDist, 5
PairwiseDistPlot, 6
PlotMolecules, 7

SeqDescs, 8