Package 'phiDelta'

October 14, 2022

Type Package

Title Tool for Phi Delta Analysis of Features

Version 1.0.1

Author Nikolas Rothe and Ursula Neumann

Maintainer Ursula Neumann <ursula.neumann@uni-marburg.de>

Description Analysis of features by phi delta diagrams. In particular, functions for reading data and calculating phi and delta as well as the functionality to plot it. Moreover it is possible to do further analysis on the data by generating rankings. For more information on phi delta diagrams, see also Giuliano Armano (2015) <doi:10.1016/j.ins.2015.07.028>.

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1.9000

NeedsCompilation no

Repository CRAN

Date/Publication 2018-05-08 08:27:22 UTC

R topics documented:

`borders` *borders of the phi delta space*

Description

calculates the corners of the phi delta space

Usage

```
borders(ratio)
```

Arguments

`ratio`	is the ratio of positive and negative of the data. The default is 1

Value

a matrix. Each row represents a corner in the following order: top, right, bottom, left

Author(s)

rothe

Examples

```
borders(1.0)
borders(0.5)
borders(2)
```

calculate_delta *calculate delta*

Description

calculates delta out of specificity and sensitivity depending on the ratio

Usage

```
calculate_delta(spec, sens, ratio = 1)
```

Arguments

spec	is the specificity, the true negative rate
sens	is the sensitivity, the true positive rate
ratio	is the ratio of positive and negative of the data. The default is 1

Value

delta

Author(s)

rothe

Examples

```
calculate_delta(1,0)
calculate_delta(0.5,0.3)
```

calculate_entropy *calculate entropy*

Description

calculates the entropy of a specificity and sensitivity tuple considering the ratio

Usage

```
calculate_entropy(spec, sens, ratio = 1)
```

Arguments

spec	numeric, is the specificity, the true negative rate
sens	numeric, is the sensitivity, the true positive rate
ratio	numeric, is the ratio of positive and negative of the data

Value

entropy of the tuple

Author(s)

rothe

Examples

```
calculate_entropy(1,0)
calculate_entropy(0.5,0.6,0.7)
```

`calculate_phi` *calculate phi*

Description

calculates phi out of specificity and sensitivity depending on the ratio

Usage

```
calculate_phi(spec, sens, ratio = 1)
```

Arguments

`spec`	is the specificity, the true negative rate
`sens`	is the sensitivity, the true positive rate
`ratio`	is the ratio of positive and negative of the data. The default is 1

Value

phi

Author(s)

rothe

Examples

```
calculate_phi(1,0)
calculate_phi(0.5,0.3)
```

`calculate_ratio` *calculate ratio*

Description

calculates the ratio between positive and negative samples

Usage

```
calculate_ratio(stats)
```

Arguments

`stats`	c_statistics

Value

ratio

Author(s)

rothe

Examples

```
x <- c_statistics(climate_data)
ratio <- calculate_ratio(x)
```

`climate_data` *Meteorological data for feature selection analysis*

Description

A dataset with meteorological data from a weather station in Frankfurt (Oder), Germany from february 2016

Usage

```
climate_data
```

Format

a data frame with 29 entries and following 7 variables

`RainBool` classification variable: if it has not rained: 0, if it has rained: 1

`date` index variable from 1 to 29

`Tmin` temperature minimum of the day

`Tmax` temperature maximum of the day

`SunAvg` sunshine duration of the day

`RelHumAvg` average relative humidity of the day

`WindForceAvg` average wind force of the day

References

modified data from http://wetterstationen.meteomedia.de/

`crossings` *Diagram crossings*

Description

adds crossings to the plot depending on the ratio

Usage

```
crossings(ratio, col = "darkblue", ...)
```

Arguments

`ratio`	is the ratio of positive and negative of the data
`col`	the color of the lines. Default is darkblue
`...`	further graphical parameters, see par

Author(s)

Neumann

Examples

```
x <- c_statistics(climate_data)
ratio <- calculate_ratio(x)
phiDelta_plot_from_data(x, crossing = FALSE)
crossings(ratio, col = "green")
```

c_matrices *confusion matrices*

Description

calculates the confusion matrices from the c_statistics

Usage

```
c_matrices(stats)
```

Arguments

stats	c_statistics

Value

a matrix. Each column represents a feature. Each row describes in this order: true negative, FALSE negative, true positive, FALSE negative

Author(s)

rothe

Examples

```
x <- c_statistics(climate_data)
cmat <- c_matrices(x)
```

c_statistics *Raw Confusion Statistics*

Description

reformarts the raw file data to c_statistics data so it can be used for most of the functions in this package. it can be used directly after loading data from a file like .csv

Usage

```
c_statistics(file)
```

Arguments

file	raw data from a file, for example the output of read.csv. the file must be formarted as follows: The first column contains tho output of the classifier. It should only be 1 or 0 The other columns represent the features. The names of the columns 2.. are considered as the names of the features

Value

dataframe, first column are the labels, 0 is a negative sample, 1 a positve the other columns contain the

Author(s)

rothe

Examples

```
data("climate_data")
x <- c_statistics(climate_data)
```

`dist_to_middle` *distance to the middle of the space*

Description

calculates the euclidic distance of a phi delta tuple to the middle of the phi delta space. This could be used for a rating of the features

Usage

```
dist_to_middle(phi, delta, ratio)
```

Arguments

`phi`	numeric value or vector of phi
`delta`	numeric value or vector of delta
`ratio`	is the ratio of positive and negative of the data. The default is 1

Value

the euclidic distance of the tuple to the middle

Author(s)

rothe

Examples

```
dist_to_middle(1,0,1)
dist_to_middle(0.5,0.3,1)
```

dist_to_top *distance to top or bottom*

Description

calculates the distance of the tuple to the closer corner of top and bottom of the phi delta space with ratio 1. This can be used for a ranking of the features

Usage

```
dist_to_top(phi, delta)
```

Arguments

phi	numeric value or vector of phi
delta	numeric value or vector of delta

Value

distance to the top or the bottom corner

Author(s)

rothe

Examples

```
dist_to_top(1,0)
dist_to_top(0.5,0.3)
```

iso_accuracy *isometric accuracy lines*

Description

adds isometric lines for the accuracy to the plot depending on the ratio

Usage

```
iso_accuracy(ratio = 1, granularity = 0.25, lty = "longdash",
  col = "blue", ...)
```

Arguments

ratio	numeric value for the ratio of positive and negative of the data
granularity	numeric value between 0 and 1 for the granularity of the lines. It is a value for the distance between 2 lines
lty	the type of line, see par
col	the color of the lines
...	further graphical parameters, see par

Author(s)

rothe

Examples

```
x <- c_statistics(climate_data)
ratio <- calculate_ratio(x)
phiDelta_plot_from_data(x)
iso_accuracy(ratio, col = "green")
```

---

iso_entropy_curve *isometric entropy*

Description

draws isometric curves for the entropy by calculating the entropy for all points in a grid and connecting those within a epsilon enviroment of the value

Usage

```
iso_entropy_curve(x, ratio = 1, eps = 0.001, grid_granularity = 0.01)
```

Arguments

x	numeric, is the offset for the points
ratio	numeric, is the ratio
eps	numeric, the epsilon for entropies to be selected
grid_granularity	numeric between 0 and 1, defines the granularity of the grid

Author(s)

Neumann

iso_negative_predictive_value *isometric negative predictive value lines*

Description

adds isometric lines for the negative predictive value to the plot depending on the ratio

Usage

```
iso_negative_predictive_value(ratio = 1, granularity = 0.25,
  lty = "longdash", col = "blue", ...)
```

Arguments

ratio	numeric value for the ratio of positive and negative of the data
granularity	numeric value between 0 and 1 for the granularity of the lines. It is a value for the distance between 2 lines
lty	the type of line, see par
col	the color of the lines
...	further graphical parameters, see par

Author(s)

rothe

Examples

```
x <- c_statistics(climate_data)
ratio <- calculate_ratio(x)
phiDelta_plot_from_data(x)
iso_negative_predictive_value(ratio, col = "green")
```

iso_precision *isometric precision lines*

Description

adds isometric lines for the precision to the plot depending on the ratio

Usage

```
iso_precision(ratio = 1, granularity = 0.25, lty = "longdash",
  col = "blue", ...)
```

Arguments

ratio	numeric value for the ratio of positive and negative of the data
granularity	numeric value between 0 and 1 for the granularity of the lines. It is a value for the distance between 2 lines
lty	the type of line, see par
col	the color of the lines
...	further graphical parameters, see par

Author(s)

rothe

Examples

```
x <- c_statistics(climate_data)
ratio <- calculate_ratio(x)
phiDelta_plot_from_data(x)
iso_precision(ratio, col = "green")
```

---

iso_sensitivity *isometric sensitivity lines*

Description

adds isometric lines for the sensitivity to the plot depending on the ratio

Usage

```
iso_sensitivity(ratio = 1, granularity = 0.25, col = "blue",
  lty = "longdash", ...)
```

Arguments

ratio	numeric value for the ratio of positive and negative of the data
granularity	numeric value between 0 and 1 for the granularity of the lines. It is a value for the distance between 2 lines
col	the color of the lines
lty	the type of line, see par
...	further graphical parameters, see par

Author(s)

Neumann

Examples

```
x <- c_statistics(climate_data)
ratio <- calculate_ratio(x)
phiDelta_plot_from_data(x)
iso_sensitivity(ratio, col = "green")
```

`iso_specificity` *isometric specificity lines*

Description

adds isometric lines for the specificity to the plot depending on the ratio

Usage

```
iso_specificity(ratio = 1, granularity = 0.25, col = "blue",
  lty = "longdash", ...)
```

Arguments

`ratio`	numeric value for the ratio of positive and negative of the data
`granularity`	numeric value between 0 and 1 for the granularity of the lines. It is a value for the distance between 2 lines
`col`	the color of the lines
`lty`	the type of line, see par
`...`	further graphical parameters, see par

Author(s)

rothe

Examples

```
x <- c_statistics(climate_data)
ratio <- calculate_ratio(x)
phiDelta_plot_from_data(x)
iso_specificity(ratio, col = "green")
```

n_matrices *normalized confusion matrices*

Description

normalizes the confusion matrices

Usage

```
n_matrices(c_matrices)
```

Arguments

c_matrices	confusion matrices

Value

a matrix. Each column represents a feature. Each row describes in this order: true negative rate, FALSE negative rate, true positive rate, FALSE negative rate

Author(s)

rothe

Examples

```
x <- c_statistics(climate_data)
cmat <- c_matrices(x)
nmat <- n_matrices(cmat)
```

phiDelta.convert *Convertion of specificity and sensitivity to phi and delta*

Description

converts specificity and sensitivity to phi and delta depending on the ratio

Usage

```
phiDelta.convert(spec, sens, ratio = 1)
```

Arguments

spec	is the specificity, the true negative rate
sens	is the sensitivity, the true positive rate
ratio	is the ratio of positive and negative of the data. The default is 1

Value

List with phi and delta vectors

Author(s)

neumann

Examples

```
phiDelta.convert(1,0)
phiDelta.convert(0.5,0.3, ratio = 0.8)
```

`phiDelta.plot` *Plot of phi delta diagram*

Description

Plots delta against phi within the phi delta diagram shape

Usage

```
phiDelta.plot(phi, delta, ratio = 1, names = NULL, border = "red",
  filling = "grey", crossing = TRUE, iso_specificity = FALSE,
  iso_sensitivity = FALSE, iso_neg_predictive_value = FALSE,
  iso_precision = FALSE, iso_accuracy = FALSE, highlighted = NULL)
```

Arguments

`phi`	numeric value or vector of phi
`delta`	numeric value or vector of delta
`ratio`	numeric, is the ratio of positive and negative of the data
`names`	string with feature names
`border`	the color of the border of the shape. NA for no border
`filling`	the color to fill the shape with
`crossing`	logical, if the crossing should be drawn
`iso_specificity`	logical, if isometric lines of the specificity should be drawn
`iso_sensitivity`	logical, if isometric lines of the sensitivity should be drawn
`iso_neg_predictive_value`	logical, if isometric lines of the negative predictive value should be drawn
`iso_precision`	logical, if isometric lines of the precision should be drawn
`iso_accuracy`	logical, if isometric lines of the accuracy should be drawn
`highlighted`	numeric vector, indices of the points to higlight highlighted points will be orange

Author(s)

rothe

Examples

```
x <- climate_data
phiDelta <- phiDelta.stats(x[,-1],x[,1])
phiDelta.plot(phiDelta$phi, phiDelta$delta)
phiDelta.plot(phiDelta$phi, phiDelta$delta,
  ratio = phiDelta$ratio,
  border = "green",
  iso_neg_predictive_value = TRUE,
  crossing = FALSE)
```

---

phiDelta.stats *Phi delta statistics from dataframe*

Description

calculates phi, delta and the ratio directly from the dataframe with provided information and generates a list with the names of the features, their phi and delta value and the ratio

Usage

```
phiDelta.stats(data, labels, ratio_corrected = TRUE)
```

Arguments

`data`	dataframe without labels
`labels`	vector of labels
`ratio_corrected`	locigal, if true phi and delta will be calculated in respect to the ratio of positive and negative samples

Value

dataframe, first column are the names of the features second column the phi values third column the delta values

Author(s)

rothe

Examples

```
x <- climate_data
phiDelta <- phiDelta.stats(x[,-1],x[,1], ratio_corrected = FALSE)
with_ratio <- phiDelta.stats(x[,-1],x[,1])
```

phiDelta_from_data *phi delta matrix*

Description

calculates phi and delta directly from the stats and generates a matrix with the names of the features, their phi and their delta value

Usage

```
phiDelta_from_data(stats, ratio_corrected = TRUE)
```

Arguments

stats	c_statistics
ratio_corrected	locigal, if true phi and delta will be calculated in respect to the ratio of positive and negative samples

Value

dataframe, first column are the names of the features second column the phi values third column the delta values

Author(s)

rothe

Examples

```
x <- c_statistics(climate_data)
phiDelta <- phiDelta_from_data(x, ratio_corrected = FALSE)
with_ratio <- phiDelta_from_data(x)
```

phiDelta_plot_from_data *phi delta plot of raw statistic data*

Description

this will create a basic plot directly out of the statistic data (c_statistics)

Usage

```
phiDelta_plot_from_data(stats, names = NULL, ratio_corrected = TRUE, ...)
```

Arguments

stats	matrix of the statistic data of the features and the classifier
names	vector with feature names
ratio_corrected	logical, if true the plot will concider the ratio of the positive and negative data samples
...	further parameters for the diagram see phiDelta.plot

Author(s)

rothe

Examples

```
x <- c_statistics(climate_data)
phiDelta_plot_from_data(x)
phiDelta_plot_from_data(x, ratio_corrected = FALSE, iso_spec = TRUE, iso_sens = TRUE)
```

---

rank_stats *ranking of the features*

Description

this function puts together a number of rankings of the features

Usage

```
rank_stats(stats, ratio_corrected = FALSE, delta_dist = 1)
```

Arguments

stats	c_statistics, the data input
ratio_corrected	logical, true if ratio shoud be considerd
delta_dist,	numeric, the delta value of the anchor for the geometrical ranking see symmetric_distance

Author(s)

rothe

`symmetric_distance` *X symmetric distance of a point*

Description

calculates the Distance from the positive anchor and the negative anchor to the point and returns the smaller one. That means, if y is positive the distance to the positive anchor will be return, if it is negative, the negative anchor distance will be calculated

Usage

```
symmetric_distance(x, y, anchor)
```

Arguments

`x, y`	numerical, in this case phi and delta but in general the input coordinates
`anchor`	vector (x,y) the anchor for the calculation of the distance

Value

the smaller distance of (x,y) to eather the positive or negative anchor

Examples

```
symmetric_distance(0.5,0.5,c(0,0))
```

Index