Package ‘cutpointr’

April 14, 2020

Type Package
Title Determine and Evaluate Optimal Cutpoints in Binary Classification Tasks
Version 1.0.2
Date 2020-04-06
Description Estimate cutpoints that optimize a specified metric in binary classification tasks and validate performance using bootstrapping. Some methods for more robust cutpoint estimation and various plotting functions are included.
License GPL-3
URL https://github.com/thie1e/cutpointr
BugReports https://github.com/thie1e/cutpointr/issues
Encoding UTF-8
LazyData true
Depends R (>= 3.5.0)
Importing grideExtra (>= 2.2.1), foreach (>= 1.4.3), dplyr (>= 0.8.0),
tidyR (>= 1.0.0), purrr (>= 0.3.0), tibble (>= 3.0.0), ggplot2
(>= 3.0.0), Rcpp (>= 0.12.12), stats, utils, rlang (>= 0.4.0)
RoxygenNote 7.1.0
Suggests KernSmooth (>= 2.23-15), fANCOVA (>= 0.5.1), testthat (>= 1.0.2), doRNG (>= 1.6), doParallel (>= 1.0.11), knitr,
rmardown, mgcv (>= 1.8), crayon (>= 1.3.4), registry (>= 0.5-1), pkgmaker(>= 0.31.1),
VignetteBuilder knitr
NeedsCompilation yes
Author Christian Thiele [aut, cre]
Maintainer Christian Thiele <c.thiele@gmx-topmail.de>
Repository CRAN
Date/Publication 2020-04-14 08:50:10 UTC
R topics documented:

abs_d_ppv_npv ........................................... 3
abs_d_sens_spec ........................................... 4
accuracy ..................................................... 5
add_metric .................................................. 6
auc ........................................................... 6
boot_ci ....................................................... 7
boot_test ..................................................... 8
cohens_kappa .............................................. 10
cutpoint ..................................................... 11
cutpointr .................................................... 11
cutpoint_knots ............................................. 17
F1_score ..................................................... 20
false omission_rate ....................................... 21
Jaccard ....................................................... 22
maximize_boot_metric ..................................... 23
maximize_gam_metric ..................................... 25
maximize_loess_metric ................................... 27
maximize_metric ........................................... 30
maximize_spline_metric .................................. 32
metric_constrain .......................................... 34
misclassification_cost .................................... 36
multi_cutpointr .......................................... 37
npv .......................................................... 38
oc_manual .................................................. 39
oc_mean ..................................................... 39
oc_median ................................................... 40
oc_youden_kernel ........................................ 41
oc_youden_normal ........................................ 42
odds_ratio .................................................. 43
plot.cutpointr ............................................. 44
plot.multi_cutpointr ..................................... 45
plot.roc_cutpointr ....................................... 45
plot_cutpoint ............................................... 46
plot_cut_boot .............................................. 47
plot_metric .................................................. 48
plot_metric_boot .......................................... 49
plot_precision_recall .................................... 49
plot_roc ..................................................... 50
plot_sensitivity_specificity ......................... 51
plot_x ....................................................... 52
plr .......................................................... 53
ppv .......................................................... 54
precision .................................................... 55
predict.cutpointr ........................................ 56
print.cutpointr ............................................ 56
Calculate the absolute difference of positive and negative predictive value

Description

Calculate the absolute difference of positive predictive value (PPV) and negative predictive value (NPV) from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
ppv = \frac{tp}{tp + fp} \\
npv = \frac{tn}{tn + fn} \\
abs_d_ppv_npv = |ppv - npv|
\]

Usage

abs_d_ppv_npv(tp, fp, tn, fn, ...)

Arguments

- **tp** (numeric) number of true positives.
- **fp** (numeric) number of false positives.
- **tn** (numeric) number of true negatives.
- **fn** (numeric) number of false negatives.
- ... for capturing additional arguments passed by method.
See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()`

Examples

```r
abs_d_ppv_npv(10, 5, 20, 10)
abs_d_ppv_npv(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
```

abs_d_sens_spec

*Calculate the absolute difference of sensitivity and specificity*

Description

Calculate the absolute difference of sensitivity and specificity from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

```
sensitivity = tp / (tp + fn)
specificity = tn / (tn + fp)
abs_d_sens_spec = |sensitivity - specificity|
```

Usage

```
abs_d_sens_spec(tp, fp, tn, fn, ...)
```

Arguments

- `tp` (numeric) number of true positives.
- `fp` (numeric) number of false positives.
- `tn` (numeric) number of true negatives.
- `fn` (numeric) number of false negatives.
- `...` for capturing additional arguments passed by method.

See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()`
accuracy

Examples

abs_d_sens_spec(10, 5, 20, 10)
abs_d_sens_spec(c(10, 8), c(5, 7), c(20, 12), c(10, 18))

accuracy

Calculate accuracy

Description

Calculate accuracy from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

accuracy = (tp + tn) / (tp + fp + tn + fn)

Usage

accuracy(tp, fp, tn, fn, ...)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tp</td>
<td>(numeric) number of true positives.</td>
</tr>
<tr>
<td>fp</td>
<td>(numeric) number of false positives.</td>
</tr>
<tr>
<td>tn</td>
<td>(numeric) number of true negatives.</td>
</tr>
<tr>
<td>fn</td>
<td>(numeric) number of false negatives.</td>
</tr>
<tr>
<td>...</td>
<td>for capturing additional arguments passed by method.</td>
</tr>
</tbody>
</table>

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), cohens_kappa(), cutpoint(), false_omission_rate(), metric_constrain(), misclassification_cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_ppv_npv(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()

Examples

accuracy(10, 5, 20, 10)
accuracy(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
add_metric

Add metrics to a cutpoint or roc_cutpoint object

Description
By default, the output of cutpoint includes the optimized metric and several other metrics. This function adds further metrics. Suitable metric functions are all metric functions that are included in the package or that comply with those standards.

Usage
add_metric(object, metric)

Arguments

object
A cutpoint or roc_cutpoint object.

metric
(list) A list of metric functions to be added.

Value
A cutpoint or roc_cutpoint object (a data.frame) with one or more added columns.

See Also

Other main cutpoint functions: boot_ci(), boot_test(), cutpoint(), multi_cutpoint(), predict.cutpoint(), roc()

Examples

library(dplyr)
library(cutpoint)
cutpoint(suicide, dsi, suicide, gender) %>%
  add_metric(list(ppv, npv)) %>%
  select(optimal_cutpoint, subgroup, AUC, sum_sens_spec, ppv, npv)

auc
Calculate AUC from a roc_cutpoint or cutpoint object

Description
Calculate the area under the ROC curve using the trapezoidal rule.
Usage

auc(x)

## S3 method for class 'roc_cutpointtr'

auc(x)

## S3 method for class 'cutpointtr'

auc(x)

Arguments

x Data frame resulting from the roc() or cutpointtr() function.

Value

Numeric vector of AUC values

Source

Forked from the AUC package

Description

Given a cutpointtr object that includes bootstrap results this function calculates a bootstrap confidence interval for a selected variable. Missing values are removed before calculating the quantiles. Values of the selected variable are returned for the percentiles alpha / 2 and 1 - alpha / 2. The metrics in the bootstrap data frames of cutpointtr are suffixed with _b and _oob to indicate in-bag and out-of-bag, respectively. For example, to calculate quantiles of the in-bag AUC variable = AUC_b should be set.

Usage

boot_ci(x, variable, in_bag = TRUE, alpha = 0.05)

Arguments

x (character) The numeric independent (predictor) variable.
variable Variable to calculate CI for
in_bag Whether the in-bag or out-of-bag results should be used for testing
alpha Alpha level. Quantiles of the bootstrapped values are returned for (alpha / 2) and 1 - (alpha / 2).
Value

A data frame with the columns quantile and value

See Also

Other main cutpointr functions: add_metric(), boot_test(), cutpointr(), multi_cutpointr(), predict.cutpointr(), roc()

Examples

## Not run:
opt_cut <- cutpointr(suicide, dsi, suicide, gender,
                     metric = youden, boot_runs = 1000)
boot_ci(opt_cut, optimal_cutpoint, in_bag = FALSE, alpha = 0.05)
boot_ci(opt_cut, acc, in_bag = FALSE, alpha = 0.05)
boot_ci(opt_cut, cohens_kappa, in_bag = FALSE, alpha = 0.05)
boot_ci(opt_cut, AUC, in_bag = TRUE, alpha = 0.05)

## End(Not run)

boot_test Test for equivalence of a metric

Description

This function performs a significance test based on the bootstrap results of cutpointr to test whether
a chosen metric is equal between subgroups or between two cutpointr objects. The test statistic is
calculated as the standardized difference of the metric between groups. If x contains subgroups, the
test is run on all possible pairings of subgroups. An additional adjusted p-value is returned in that
case.

Usage

boot_test(x, y = NULL, variable = "AUC", in_bag = TRUE, correction = "holm")

Arguments

x A cutpointr object with bootstrap results
y If x does not contain subgroups another cutpointr object
variable The variable for testing
in_bag Whether the in-bag or out-of-bag results should be used for testing
correction The type of correction for multiple testing. Possible values are as in p.adjust.methods
Details

The variable name is looked up in the columns of the bootstrap results where the suffixes _b and _oob indicate in-bag and out-of-bag estimates, respectively (controlled via the in_bag argument). Possible values are optimal_cutpoint, AUC, acc, sensitivity, specificity, and the metric that was selected in cutpointr. Note that there is no "out-of-bag optimal cutpoint", so when selecting variable = optimal_cutpoint the test will be based on the in-bag data.

The test statistic is calculated as z = (t1 - t2) / sd(t1 - t2) where t1 and t2 are the metric values on the full sample and sd(t1 - t2) is the standard deviation of the differences of the metric values per bootstrap repetition. The test is two-sided.

If two cutpointr objects are compared and the numbers of bootstrap repetitions differ, the smaller number will be used.

Value

A data.frame (a tibble) with the columns test_var, p, d, sd_d, z and in_bag. If a grouped cutpointr object was tested, the additional columns subgroup1, subgroup2 and p_adj are returned.

Source


See Also

Other main cutpointr functions: add_metric(), boot_ci(), cutpointr(), multi_cutpointr(), predict.cutpointr(), roc()

Examples

## Not run:
library(cutpointr)
library(dplyr)
set.seed(734)
cp_f <- cutpointr(suicide %>% filter(gender == "female"), dsi, suicide,
  boot_runs = 1000, boot_stratify = TRUE)
set.seed(928)
cp_m <- cutpointr(suicide %>% filter(gender == "male"), dsi, suicide,
  boot_runs = 1000, boot_stratify = TRUE)
# No significant differences:
boot_test(cp_f, cp_m, AUC, in_bag = TRUE)
boot_test(cp_f, cp_m, sum_sens_spec, in_bag = FALSE)
set.seed(135)
cp <- cutpointr(suicide, dsi, suicide, gender, boot_runs = 1000,
  boot_stratify = TRUE)
# Roughly same result as above:
boot_test(cp, variable = AUC, in_bag = TRUE)
boot_test(cp, variable = sum_sens_spec, in_bag = FALSE)
cohens_kappa

Calculate Cohen's Kappa

Description

Calculate the Kappa metric from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
\begin{align*}
\text{mrg}_a &= ((tp + fn) \times (tp + fp)) / (tp + fn + fp + tn) \\
\text{mrg}_b &= ((fp + tn) \times (fn + tn)) / (tp + fn + fp + tn) \\
\text{expec_agree} &= (\text{mrg}_a + \text{mrg}_b) / (tp + fn + fp + tn) \\
\text{obs_agree} &= (tp + tn) / (tp + fn + fp + tn) \\
\text{cohens_kappa} &= (\text{obs_agree} - \text{expec_agree}) / (1 - \text{expec_agree})
\end{align*}
\]

Usage

cohens_kappa(tp, fp, tn, fn, ...)

Arguments

tp (numeric) number of true positives.
fp (numeric) number of false positives.
tn (numeric) number of true negatives.
fn (numeric) number of false negatives.
... for capturing additional arguments passed by method.

Value

A numeric matrix with the column name "cohens_kappa".

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cutpoint(), false_omission_rate(), metric_constrain(), misclassification_cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_ppv_npv(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()

Examples

cohens_kappa(10, 5, 20, 10)
cohens_kappa(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
cutpoint  

Extract the cutpoints from a ROC curve generated by cutpointr

Description

This is a utility function for extracting the cutpoints from a roc_cutpointr object. Mainly useful in conjunction with the plot_cutpointr function if cutpoints are to be plotted on the x-axis.

Usage

cutpoint(x, ...)

cutpoints(x, ...)

Arguments

x
  A roc_cutpointr object.

...  Further arguments.

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cohens_kappa(), false omission rate(), metric_constrain(), misclassification_cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_ppv_npv(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()

Examples

oc <- cutpointr(suicide, dsi, suicide, gender)
plot_cutpointr(oc, cutpoint, accuracy)

---

cutpointr  

Determine and evaluate optimal cutpoints

Description

Using predictions (or e.g. biological marker values) and binary class labels, this function will determine "optimal" cutpoints using various selectable methods. The methods for cutpoint determination can be evaluated using bootstrapping. An estimate of the cutpoint variability and the out-of-sample performance can then be returned with summary or plot. For an introduction to the package please see vignette("cutpointr", package = "cutpointr")
Usage

cutpointr(...)

## Default S3 method:
cutpointr(
  data,
  x,
  class,
  subgroup = NULL,
  method = maximize_metric,
  metric = sum_sens_spec,
  pos_class = NULL,
  neg_class = NULL,
  direction = NULL,
  boot_runs = 0,
  boot_stratify = FALSE,
  use_midpoints = FALSE,
  break_ties = median,
  na.rm = FALSE,
  allowParallel = FALSE,
  silent = FALSE,
  tol_metric = 1e-06,
  ...
)

## S3 method for class 'numeric'
cutpointr(
  x,
  class,
  subgroup = NULL,
  method = maximize_metric,
  metric = sum_sens_spec,
  pos_class = NULL,
  neg_class = NULL,
  direction = NULL,
  boot_runs = 0,
  boot_stratify = FALSE,
  use_midpoints = FALSE,
  break_ties = median,
  na.rm = FALSE,
  allowParallel = FALSE,
  silent = FALSE,
  tol_metric = 1e-06,
  ...
)
**Arguments**

... Further optional arguments that will be passed to method. minimize_metric and maximize_metric pass ... to metric.

data A data.frame with the data needed for x, class and optionally subgroup.
x The variable name to be used for classification, e.g. predictions. The raw vector of values if the data argument is unused.
class The variable name indicating class membership. If the data argument is unused, the vector of raw numeric values.
subgroup An additional covariate that identifies subgroups or the raw data if data = NULL. Separate optimal cutpoints will be determined per group. Numeric, character and factor are allowed.
method (function) A function for determining cutpoints. Can be user supplied or use some of the built in methods. See details.
metric (function) The function for computing a metric when using minimize_metric or maximize_metric as method and and for the out-of-bag values during bootstrapping. A way of internally validating the performance. User defined functions can be supplied, see details.
pos_class (optional) The value of class that indicates the positive class.
neg_class (optional) The value of class that indicates the negative class.
direction (character, optional) Use ">=" or "\<=" to indicate whether x is supposed to be larger or smaller for the positive class.
boot_runs (numerical) If positive, this number of bootstrap samples will be used to assess the variability and the out-of-sample performance.
boot_stratify (logical) If the bootstrap is stratified, bootstrap samples are drawn in both classes and then combined, keeping the number of positives and negatives constant in every resample.
use_midpoints (logical) If TRUE (default FALSE) the returned optimal cutpoint will be the mean of the optimal cutpoint and the next highest observation (for direction = ">=") or the next lowest observation (for direction = "\<=") which avoids biasing the optimal cutpoint.
break_ties If multiple cutpoints are found, they can be summarized using this function, e.g. mean or median. To return all cutpoints use c as the function.
nna.rm (logical) Set to TRUE (default FALSE) to keep only complete cases of x, class and subgroup (if specified). Missing values with na.rm = FALSE will raise an error.
allowParallel (logical) If TRUE, the bootstrapping will be parallelized using foreach. A local cluster, for example, should be started manually beforehand.
silent (logical) If TRUE suppresses all messages.
tol_metric All cutpoints will be returned that lead to a metric value in the interval [m\_max - tol\_metric, m\_max + tol\_metric] where m\_max is the maximum achievable metric value. This can be used to return multiple decent cutpoints and to avoid floating-point problems. Not supported by all method functions, see details.
Details

If direction and/or pos_class and neg_class are not given, the function will assume that higher values indicate the positive class and use the class with a higher median as the positive class.

This function uses tidyeval to support unquoted arguments. For programming with cutpointr the operator !! can be used to unquote an argument, see the examples.

Different methods can be selected for determining the optimal cutpoint via the method argument. The package includes the following method functions:

- maximize_metric: Maximize the metric function
- minimize_metric: Minimize the metric function
- maximize_loess_metric: Maximize the metric function after LOESS smoothing
- minimize_loess_metric: Minimize the metric function after LOESS smoothing
- maximize_spline_metric: Maximize the metric function after spline smoothing
- minimize_spline_metric: Minimize the metric function after spline smoothing
- maximize_boot_metric: Maximize the metric function as a summary of the optimal cutpoints in bootstrapped samples
- minimize_boot_metric: Minimize the metric function as a summary of the optimal cutpoints in bootstrapped samples
- oc_youden_kernel: Maximize the Youden-Index after kernel smoothing the distributions of the two classes
- oc_youden_normal: Maximize the Youden-Index parametrically assuming normally distributed data in both classes
- oc_manual: Specify the cutpoint manually

User-defined functions can be supplied to method, too. As a reference, the code of all included method functions can be accessed by simply typing their name. To define a new method function, create a function that may take as input(s):

- data: A data.frame or tbl_df
- x: (character) The name of the predictor or independent variable
- class: (character) The name of the class or dependent variable
- metric_func: A function for calculating a metric, e.g. accuracy
- pos_class: The positive class
- neg_class: The negative class
- direction: "&gt;" if the positive class has higher x values, "&lt;=" otherwise
- tol_metric: (numeric) In the built-in methods a tolerance around the optimal metric value
- use_midpoints: (logical) In the built-in methods whether to use midpoints instead of exact optimal cutpoints
- ... Further arguments

The ... argument can be used to avoid an error if not all of the above arguments are needed or in order to pass additional arguments to method. The function should return a data.frame or tbl_df with one row, the column "optimal_cutpoint", and an optional column with an arbitrary name with the metric value at the optimal cutpoint.

Built-in metric functions include:
• accuracy: Fraction correctly classified
• youden: Youden- or J-Index = sensitivity + specificity - 1
• sum_sens_spec: sensitivity + specificity
• sum_ppv_npv: The sum of positive predictive value (PPV) and negative predictive value (NPV)
• prod_sens_spec: sensitivity * specificity
• prod_ppv_npv: The product of positive predictive value (PPV) and negative predictive value (NPV)
• cohens_kappa: Cohen’s Kappa
• abs_d_sens_spec: The absolute difference between sensitivity and specificity
• roc01: Distance to the point (0,1) on ROC space
• abs_d_ppv_npv: The absolute difference between positive predictive value (PPV) and negative predictive value (NPV)
• p_chisquared: The p-value of a chi-squared test on the confusion matrix of predictions and observations
• odds_ratio: The odds ratio calculated as (TP / FP) / (FN / TN)
• risk_ratio: The risk ratio (relative risk) calculated as (TP / (TP + FN)) / (FP / (FP + TN))
• positive and negative likelihood ratio calculated as \( plr = \frac{TP}{FP} \) / (false positive rate)
• nlr = false negative rate / true negative rate
• misclassification_cost: The sum of the misclassification cost of false positives and false negatives \( fp * cost_fp + fn * cost_fn \). Additional arguments to cutpointr: cost_fp, cost_fn
• total_utility: The total utility of true / false positives / negatives calculated as utility_tp * TP + utility_tn * TN - cost_fp * FP - cost_fn * FN. Additional arguments to cutpointr: utility_tp, utility_tn, cost_fp, cost_fn
• F1_score: The F1-score \( (2 * TP) / (2 * TP + FP + FN) \)
• sens_constrain: Maximize sensitivity given a minimal value of specificity
• spec_constrain: Maximize specificity given a minimal value of sensitivity
• metric_constrain: Maximize a selected metric given a minimal value of another selected metric

Furthermore, the following functions are included which can be used as metric functions but are more useful for plotting purposes, for example in plot_cutpointr, or for defining new metric functions: \( tp, fp, tn, fn, tpr, fpr, tnr, fnr, false omission_rate, false discovery rate, ppv, npv, precision, recall, sensitivity, and specificity. \]

User defined metric functions can be created as well which can accept the following inputs as vectors:

• \( tp \): Vector of true positives
• \( fp \): Vector of false positives
• \( tn \): Vector of true negatives
• \( fn \): Vector of false negatives
If the metric function is used in conjunction with any of the maximize / minimize methods, further arguments can be passed.

The function should return a numeric vector or a matrix or a data.frame with one column. If the column is named, the name will be included in the output and plots. Avoid using names that are identical to the column names that are by default returned by `cutpointtr`.

If `boot_runs` is positive, that number of bootstrap samples will be drawn and the optimal cutpoint using method will be determined. Additionally, as a way of internal validation, the function in metric will be used to score the out-of-bag predictions using the cutpoints determined by method. Various default metrics are always included in the bootstrap results.

If multiple optimal cutpoints are found, the column `optimal_cutpoint` becomes a list that contains the vector(s) of the optimal cutpoints.

If `use_midpoints = TRUE` the mean of the optimal cutpoint and the next highest or lowest possible cutpoint is returned, depending on direction.

The `tol_metric` argument can be used to avoid floating-point problems that may lead to exclusion of cutpoints that achieve the optimally achievable metric value. Additionally, by selecting a large tolerance multiple cutpoints can be returned that lead to decent metric values in the vicinity of the optimal metric value. `tol_metric` is passed to metric and is only supported by the maximization and minimization functions, i.e. `maximize_metric`, `minimize_metric`, `maximize_loess_metric`, `minimize_loess_metric`, `maximize_spline_metric`, and `minimize_spline_metric`. In `maximize_boot_metric` and `minimize_boot_metric` multiple optimal cutpoints will be passed to the `summary_func` of these two functions.

**Value**

A `cutpointtr` object which is also a data.frame and tbl_df.

**See Also**

Other main `cutpointtr` functions: `add_metric()`, `boot_ci()`, `boot_test()`, `multi_cutpointtr()`, `predict.cutpointtr()`, `roc()`

**Examples**

```r
library(cutpointtr)

## Optimal cutpoint for dsi
data(suicide)
opt_cut <- cutpointtr(suicide, dsi, suicide)
opt_cut
s_opt_cut <- summary(opt_cut)
plot(opt_cut)

## Not run:
## Predict class for new observations
predict(opt_cut, newdata = data.frame(dsi = 0:5))

## Supplying raw data, same result
cutpointtr(x = suicide$dsi, class = suicide$suicide)
```
# direction, class labels, method and metric can be defined manually
# Again, same result

cutpointr(suicide, dsi, suicide, direction = ">=",
    pos_class = "yes",
    method = maximize_metric, metric = youden)

## Optimal cutpoint for dsi, as before, but for the separate subgroups

```r
opt_cut <- cutpointr(suicide, dsi, suicide, gender)
```

```r
(s_opt_cut <- summary(opt_cut))
tibble:::print.tbl(s_opt_cut)
```

## Bootstrapping also works on individual subgroups

```r
set.seed(30)
opt_cut <- cutpointr(suicide, dsi, suicide, gender, boot_runs = 1000,
    boot_stratify = TRUE)
opt_cut
```

```r
summary(opt_cut)
```

```r
plot(opt_cut)
```

## Parallelized bootstrapping

```r
library(doParallel)
library(doRNG)
c1 <- makeCluster(2) # 2 cores
registerDoParallel(c1)
registerDoRNG(12) # Reproducible parallel loops using doRNG
```

```r
opt_cut <- cutpointr(suicide, dsi, suicide, gender,
    boot_runs = 1000, allowParallel = TRUE)
```

```r
stopCluster(c1)
opt_cut
```

```r
plot(opt_cut)
```

## Robust cutpoint method using kernel smoothing for optimizing Youden-Index

```r
opt_cut <- cutpointr(suicide, dsi, suicide, gender,
    method = oc_youden_kernel)
opt_cut
```

## End(Not run)

---

**cutpointr_**

*The standard evaluation version of cutpointr (deprecated)*

---

**Description**

This function is equivalent to `cutpointr` but takes only quoted arguments for `x`, `class` and `subgroup`. This was useful before `cutpointr` supported `tidyeval`. 
Usage

cutpointr_(
data, 
x, 
class, 
subgroup = NULL, 
method = maximize_metric, 
metric = sum_sens_spec, 
pos_class = NULL, 
neg_class = NULL, 
direction = NULL, 
boot_runs = 0, 
boot_stratify = FALSE, 
use_midpoints = FALSE, 
break_ties = median, 
na.rm = FALSE, 
allowParallel = FALSE, 
silent = FALSE, 
tol_metric = 1e-06, 
...
)

Arguments

data A data.frame with the data needed for x, class and optionally subgroup.
x (character) The variable name to be used for classification, e.g. predictions or test values.
class (character) The variable name indicating class membership.
subgroup (character) The variable name of an additional covariate that identifies subgroups. Separate optimal cutpoints will be determined per group.
method (function) A function for determining cutpoints. Can be user supplied or use some of the built in methods. See details.
metric (function) The function for computing a metric when using maximize_metric or minimize_metric as method and and for the out-of-bag values during bootstrapping. A way of internally validating the performance. User defined functions can be supplied, see details.
pos_class (optional) The value of class that indicates the positive class.
neg_class (optional) The value of class that indicates the negative class.
direction (character, optional) Use ">=" or "<=" to indicate whether x is supposed to be larger or smaller for the positive class.
boot_runs (numerical) If positive, this number of bootstrap samples will be used to assess the variability and the out-of-sample performance.
boot_stratify (logical) If the bootstrap is stratified, bootstrap samples are drawn in both classes and then combined, keeping the number of positives and negatives constant in every resample.
cutpoint_knots (logical) If TRUE (default FALSE) the returned optimal cutpoint will be the mean of the optimal cutpoint and the next highest observation (for direction = "\geq") or the next lowest observation (for direction = "\leq") which avoids biasing the optimal cutpoint.

use_midpoints (logical) If TRUE (default FALSE) the returned optimal cutpoint will be the mean of the optimal cutpoint and the next highest observation (for direction = "\geq") or the next lowest observation (for direction = "\leq") which avoids biasing the optimal cutpoint.

break_ties If multiple cutpoints are found, they can be summarized using this function, e.g. mean or median. To return all cutpoints use c as the function.

na.rm (logical) Set to TRUE (default FALSE) to keep only complete cases of x, class and subgroup (if specified). Missing values with na.rm = FALSE will raise an error.

allowParallel (logical) If TRUE, the bootstrapping will be parallelized using foreach. A local cluster, for example, should be started manually beforehand.

silent (logical) If TRUE suppresses all messages.

tol_metric All cutpoints will be returned that lead to a metric value in the interval \[m_{\text{max}} - \text{tol\_metric}, m_{\text{max}} + \text{tol\_metric}\] where \(m_{\text{max}}\) is the maximum achievable metric value. This can be used to return multiple decent cutpoints and to avoid floating-point problems. Not supported by all method functions, see details.

Examples

library(cutpointr)

## Optimal cutpoint for dsi
data(suicide)
opt_cut <- cutpointr_(suicide, "dsi", "suicide")
summary(opt_cut)
plot(opt_cut)
predict(opt_cut, newdata = data.frame(dsi = 0:5))

---

**cutpoint_knots**  
*Calculate number of knots to use in spline smoothing*

**Description**

This function calculates the number of knots when using smoothing splines for smoothing a function of metric values per cutpoint value. The function for calculating the number of knots is equal to stats::n knots_smspl but uses the number of unique cutpoints in the data as n.

**Usage**

cutpoint_knots(data, x)
Arguments

data  A data frame
x     (character) The name of the predictor variable

Examples

cutpoint_knots(suicide, "dsi")

F1_score  Calculate the F1-score

Description

Calculate the F1-score from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
F1\_score = \frac{(2 \times tp)}{(2 \times tp + fp + fn)}
\]

Usage

F1_score(tp, fp, tn, fn, ...)

Arguments

tp   (numeric) number of true positives.
fp   (numeric) number of false positives.
tn   (numeric) number of true negatives.
fn   (numeric) number of false negatives.
...  for capturing additional arguments passed by method.

See Also

Other metric functions: Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cohens_kappa(), cutpoint(), false_omission_rate(), metric_constrain(), misclassification_cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_ppv_npv(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()

Examples

F1_score(10, 5, 20, 10)
F1_score(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
false omission rate

Description

Calculate the false omission rate or false discovery rate from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
\text{false omission rate} = \frac{fn}{(tn + fn)} = 1 - \text{npv} \\
\text{false discovery rate} = \frac{fp}{(tp + fp)} = 1 - \text{ppv}
\]

Usage

false omission rate(tp, fp, tn, fn, ...)
false discovery rate(tp, fp, tn, fn, ...)

Arguments

| tp        | (numeric) number of true positives. |
| fp        | (numeric) number of false positives. |
| tn        | (numeric) number of true negatives. |
| fn        | (numeric) number of false negatives. |
| ...       | for capturing additional arguments passed by method. |

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cohens_kappa(), cutpoint(), metric_constrain(), misclassification_cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_ppv_npv(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()

Examples

false omission rate(10, 5, 20, 10)
false omission rate(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
Calculate the Jaccard Index

Description

Calculate the Jaccard Index from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
J \text{accard} = \frac{tp}{tp + fp + fn}
\]

Usage

\[
\text{Jaccard}(tp, fp, tn, fn, \ldots)
\]

Arguments

- \(tp\) (numeric) number of true positives.
- \(fp\) (numeric) number of false positives.
- \(tn\) (numeric) number of true negatives.
- \(fn\) (numeric) number of false negatives.
- \(\ldots\) for capturing additional arguments passed by method.

See Also

Other metric functions: `F1_score()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()`

Examples

\[
\text{Jaccard}(10, 5, 20, 10)
\]
\[
\text{Jaccard}(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
\]
maximize_boot_metric

Optimize a metric function in binary classification after bootstrapping

Description

Given a function for computing a metric in `metric_func`, these functions bootstrap the data `boot_cut` times and maximize or minimize the metric by selecting an optimal cutpoint. The returned optimal cutpoint is the result of applying `summary_func`, e.g. the mean, to all optimal cutpoints that were determined in the bootstrap samples. The metric function should accept the following inputs:

- `tp`: vector of number of true positives
- `fp`: vector of number of false positives
- `tn`: vector of number of true negatives
- `fn`: vector of number of false negatives

Usage

```r
maximize_boot_metric(
  data,
  x,
  class,
  metric_func = youden,
  pos_class = NULL,
  neg_class = NULL,
  direction,
  summary_func = mean,
  boot_cut = 50,
  inf_rm = TRUE,
  tol_metric,
  use_midpoints,
  ...
)
```

```r
minimize_boot_metric(
  data,
  x,
  class,
  metric_func = youden,
  pos_class = NULL,
  neg_class = NULL,
  direction,
  summary_func = mean,
  boot_cut = 50,
  inf_rm = TRUE,
  tol_metric,
  use_midpoints,
  ...
)
```
maximize_boot_metric

Arguments

data A data frame or tibble in which the columns that are given in x and class can be found.
x (character) The variable name to be used for classification, e.g. predictions or test values.
class (character) The variable name indicating class membership.
metric_func (function) A function that computes a single number metric to be maximized. See description.
pos_class The value of class that indicates the positive class.
neg_class The value of class that indicates the negative class.
direction (character) Use ">=" or "<=" to select whether an x value >= or <= the cutoff predicts the positive class.
summary_func (function) After obtaining the bootstrapped optimal cutpoints this function, e.g. mean or median, is applied to arrive at a single cutpoint.
boot_cut (numeric) Number of bootstrap repetitions over which the mean optimal cutpoint is calculated.
infin_rm (logical) whether to remove infinite cutpoints before calculating the summary.
tol_metric All cutpoints will be passed to summary_func that lead to a metric value in the interval \([m_{max} - tol_{metric}, m_{max} + tol_{metric}]\) where \(m_{max}\) is the maximum achievable metric value. This can be used to return multiple decent cutpoints and to avoid floating-point problems.
use_midpoints (logical) If TRUE (default FALSE) the returned optimal cutpoint will be the mean of the optimal cutpoint and the next highest observation (for direction = ">") or the next lowest observation (for direction = "<") which avoids biasing the optimal cutpoint.

Details

The above inputs are arrived at by using all unique values in x, Inf, and -Inf as possible cutpoints for classifying the variable in class. The reported metric represents the usual in-sample performance of the determined cutpoint.

Value

A tibble with the column optimal_cutpoint
maximize_gam_metric

See Also

Other method functions: maximize_gam_metric(), maximize_loess_metric(), maximize_metric(), maximize_spline_metric(), oc_manual(), oc_mean(), oc_median(), oc_youden_kernel(), oc_youden_normal()

Examples

set.seed(100)
cutpointr(suicide, dsi, suicide, method = maximize_boot_metric, metric = accuracy, boot_cut = 30)
set.seed(100)
cutpointr(suicide, dsi, suicide, method = minimize_boot_metric, metric = abs_d_sens_spec, boot_cut = 30)

maximize_gam_metric  Optimize a metric function in binary classification after smoothing via generalized additive models

Description

Given a function for computing a metric in metric_func, these functions smooth the function of metric value per cutpoint using generalized additive models (as implemented in mgcv), then maximize or minimize the metric by selecting an optimal cutpoint. For further details on the GAM smoothing see ?mgcv::gam. The metric function should accept the following inputs:

• tp: vector of number of true positives
• fp: vector of number of false positives
• tn: vector of number of true negatives
• fn: vector of number of false negatives

Usage

maximize_gam_metric(
  data,
  x,
  class,
  metric_func = youden,
  pos_class = NULL,
  neg_class = NULL,
  direction,
  formula = m ~ s(x.sorted),
  optimizer = c("outer", "newton"),
  tol_metric,
  use_midpoints,
  ...
)
minimize_gam_metric(
  data,
  x,
  class,
  metric_func = youden,
  pos_class = NULL,
  neg_class = NULL,
  direction,
  formula = m ~ s(x.sorted),
  optimizer = c("outer", "newton"),
  tol_metric,
  use_midpoints,
  ...
)

Arguments

data  A data frame or tibble in which the columns that are given in x and class can be found.
x    (character) The variable name to be used for classification, e.g. predictions or test values.
class (character) The variable name indicating class membership.
metric_func (function) A function that computes a metric to be maximized. See description.
pos_class The value of class that indicates the positive class.
neg_class The value of class that indicates the negative class.
direction (character) Use ">=" or "<<" to select whether an x value >= or <= the cutoff predicts the positive class.
formula A GAM formula. See help("gam",package = "mgcv") for details.
optimizer An array specifying the numerical optimization method to use to optimize the smoothing parameter estimation criterion (given by method). See help("gam",package = "mgcv") for details.
tol_metric All cutpoints will be returned that lead to a metric value in the interval [m_max - tol_metric, m_max + tol_metric] where m_max is the maximum achievable metric value. This can be used to return multiple decent cutpoints and to avoid floating-point problems.
use_midpoints (logical) If TRUE (default FALSE) the returned optimal cutpoint will be the mean of the optimal cutpoint and the next highest observation (for direction = ">") or the next lowest observation (for direction = "<") which avoids biasing the optimal cutpoint.
...
Further arguments that will be passed to metric_func or the GAM smoother.

Details

The above inputs are arrived at by using all unique values in x, Inf, and -Inf as possible cutpoints for classifying the variable in class.
maximize_loess_metric

Value

A tibble with the columns optimal_cutpoint, the corresponding metric value and roc_curve, a nested tibble that includes all possible cutoffs and the corresponding numbers of true and false positives / negatives and all corresponding metric values.

See Also

Other method functions: maximize_boot_metric(), maximize_loess_metric(), maximize_metric(), maximize_spline_metric(), oc_manual(), oc_mean(), oc_median(), oc_youden_kernel(), oc_youden_normal()

Examples

oc <- cutpointr(suicide, dsi, suicide, gender, method = maximize_gam_metric, metric = accuracy)
plot_metric(oc)
oc <- cutpointr(suicide, dsi, suicide, gender, method = minimize_gam_metric, metric = abs_d_sens_spec)
plot_metric(oc)

maximize_loess_metric  Optimize a metric function in binary classification after LOESS smoothing

Description

Given a function for computing a metric in metric_func, these functions smooth the function of metric value per cutpoint using LOESS, then maximize or minimize the metric by selecting an optimal cutpoint. For further details on the LOESS smoothing see ?fANCOVA::loess.as. The metric function should accept the following inputs:

- tp: vector of number of true positives
- fp: vector of number of false positives
- tn: vector of number of true negatives
- fn: vector of number of false negatives

Usage

maximize_loess_metric(  
data,  
x,  
class,  
metric_func = youden,  
pos_class = NULL,  
neg_class = NULL,  
direction,  
criterion = "aicc",)
maximize_loess_metric

```r
degree = 1,
family = "symmetric",
user.span = NULL,
tol_metric,
use_midpoints,
... )
```

```r
minimize_loess_metric(
  data,
x,
class,
metric_func = youden,
pos_class = NULL,
neg_class = NULL,
direction,
criterion = "aicc",
degree = 1,
family = "symmetric",
user.span = NULL,
tol_metric,
use_midpoints,
... )
```

**Arguments**

- **data**
  A data frame or tibble in which the columns that are given in `x` and `class` can be found.

- **x**
  (character) The variable name to be used for classification, e.g. predictions or test values.

- **class**
  (character) The variable name indicating class membership.

- **metric_func**
  (function) A function that computes a metric to be maximized. See description.

- **pos_class**
  The value of `class` that indicates the positive class.

- **neg_class**
  The value of `class` that indicates the negative class.

- **direction**
  (character) Use ">=" or "+=" to select whether an `x` value >= or <= the cutoff predicts the positive class.

- **criterion**
  the criterion for automatic smoothing parameter selection: "aicc" denotes bias-corrected AIC criterion, "gcv" denotes generalized cross-validation.

- **degree**
  the degree of the local polynomials to be used. It can be 0, 1 or 2.

- **family**
  if "gaussian" fitting is by least-squares, and if "symmetric" a re-descending M estimator is used with Tukey's biweight function.

- **user.span**
  The user-defined parameter which controls the degree of smoothing

- **tol_metric**
  All cutpoints will be returned that lead to a metric value in the interval \([m_{\text{max}} - \text{tol}\text{metric}, m_{\text{max}} + \text{tol}\text{metric}]\) where \(m_{\text{max}}\) is the maximum achievable
maximize_loess_metric

metric value. This can be used to return multiple decent cutpoints and to avoid floating-point problems.

use_midpoints (logical) If TRUE (default FALSE) the returned optimal cutpoint will be the mean of the optimal cutpoint and the next highest observation (for direction = ">") or the next lowest observation (for direction = "<") which avoids biasing the optimal cutpoint.

... Further arguments that will be passed to metric_func or the loess smoother.

Details

The above inputs are arrived at by using all unique values in x, Inf, and -Inf as possible cutpoints for classifying the variable in class.

Value

A tibble with the columns optimal_cutpoint, the corresponding metric value and roc_curve, a nested tibble that includes all possible cutoffs and the corresponding numbers of true and false positives / negatives and all corresponding metric values.

Source


See Also

Other method functions: maximize_boot_metric(), maximize_gam_metric(), maximize_metric(), maximize_spline_metric(), oc_manual(), oc_mean(), oc_median(), oc_youden_kernel(), oc_youden_normal()

Examples

oc <- cutpointr(suicide, dsi, suicide, gender, method = maximize_loess_metric, criterion = "aicc", family = "symmetric", degree = 2, user.span = 0.7, metric = accuracy)
plot_metric(oc)
oc <- cutpointr(suicide, dsi, suicide, gender, method = minimize_loess_metric, criterion = "aicc", family = "symmetric", degree = 2, user.span = 0.7, metric = misclassification_cost, cost_fp = 1, cost_fn = 10)
plot_metric(oc)
maximize_metric  

**Optimize a metric function in binary classification**

**Description**

Given a function for computing a metric in `metric_func`, these functions maximize or minimize that metric by selecting an optimal cutpoint. The metric function should accept the following inputs:

- `tp`: vector of number of true positives
- `fp`: vector of number of false positives
- `tn`: vector of number of true negatives
- `fn`: vector of number of false negatives

**Usage**

```r
maximize_metric(
  data,  
  x,  
  class,  
  metric_func = youden,  
  pos_class = NULL,  
  neg_class = NULL,  
  direction,  
  tol_metric,  
  use_midpoints,  
  ...  
)

minimize_metric(
  data,  
  x,  
  class,  
  metric_func = youden,  
  pos_class = NULL,  
  neg_class = NULL,  
  direction,  
  tol_metric,  
  use_midpoints,  
  ...  
)
```

**Arguments**

- `data`  
  A data frame or tibble in which the columns that are given in `x` and `class` can be found.
maximize_metric

x (character) The variable name to be used for classification, e.g. predictions or test values.

class (character) The variable name indicating class membership.

metric_func (function) A function that computes a metric to be maximized. See description.

pos_class The value of class that indicates the positive class.

neg_class The value of class that indicates the negative class.

direction (character) Use ">=" or "<=" to select whether an x value >= or <= the cutoff predicts the positive class.

tol_metric All cutpoints will be returned that lead to a metric value in the interval \([m_{max} - tol_{metric}, m_{max} + tol_{metric}]\) where \(m_{max}\) is the maximum achievable metric value. This can be used to return multiple decent cutpoints and to avoid floating-point problems.

use_midpoints (logical) If TRUE (default FALSE) the returned optimal cutpoint will be the mean of the optimal cutpoint and the next highest observation (for direction = ">") or the next lowest observation (for direction = "<") which avoids biasing the optimal cutpoint.

... Further arguments that will be passed to metric_func.

Details

The above inputs are arrived at by using all unique values in x, Inf, or -Inf as possible cutpoints for classifying the variable in class.

Value

A tibble with the columns optimal_cutpoint, the corresponding metric value and roc_curve, a nested tibble that includes all possible cutoffs and the corresponding numbers of true and false positives / negatives and all corresponding metric values.

See Also

Other method functions: maximize_boot_metric(), maximize_gam_metric(), maximize_loess_metric(), maximize_spline_metric(), oc_manual(), oc_mean(), oc_median(), oc_youden_kernel(), oc_youden_normal()

Examples

cutpointr(suicide, dsi, suicide, method = maximize_metric, metric = accuracy)
cutpointr(suicide, dsi, suicide, method = minimize_metric, metric = abs_d_sens_spec)
maximize_spline_metric

Optimize a metric function in binary classification after spline smoothing

Description

Given a function for computing a metric in metric_func, this function smoothes the function of metric value per cutpoint using smoothing splines. Then it optimizes the metric by selecting an optimal cutpoint. For further details on the smoothing spline see \texttt{stats::smooth.spline}. The metric function should accept the following inputs:

- \texttt{tp}: vector of number of true positives
- \texttt{fp}: vector of number of false positives
- \texttt{tn}: vector of number of true negatives
- \texttt{fn}: vector of number of false negatives

Usage

maximize_spline_metric(
    data,
    x,
    class,
    metric_func = youden,
    pos_class = NULL,
    neg_class = NULL,
    direction,
    w = NULL,
    df = NULL,
    spar = 1,
    nknots = cutpoint_knots,
    df_offset = NULL,
    penalty = 1,
    control_spar = list(),
    tol_metric,
    use_midpoints,
    ...
)

minimize_spline_metric(
    data,
    x,
    class,
    metric_func = youden,
    pos_class = NULL,
    neg_class = NULL,
Arguments

- **data**
  - A data frame or tibble in which the columns that are given in x and class can be found.

- **x**
  - (character) The variable name to be used for classification, e.g. predictions or test values.

- **class**
  - (character) The variable name indicating class membership.

- **metric_func**
  - (function) A function that computes a metric to be optimized. See description.

- **pos_class**
  - The value of class that indicates the positive class.

- **neg_class**
  - The value of class that indicates the negative class.

- **direction**
  - (character) Use ">=" or "<=" to select whether an x value >= or <= the cutoff predicts the positive class.

- **w**
  - Optional vector of weights of the same length as x; defaults to all 1.

- **df**
  - The desired equivalent number of degrees of freedom (trace of the smoother matrix). Must be in (1, nx], nx the number of unique x values.

- **spar**
  - Smoothing parameter, typically (but not necessarily) in (0,1]. When spar is specified, the coefficient lambda of the integral of the squared second derivative in the fit (penalized log likelihood) criterion is a monotone function of spar.

- **nknots**
  - Integer or function giving the number of knots. The function should accept data and x (the name of the predictor variable) as inputs. By default nknots = 0.1 * log(n_dat / n_cut) * n_cut where n_dat is the number of observations and n_cut the number of unique predictor values.

- **df_offset**
  - Allows the degrees of freedom to be increased by df_offset in the GCV criterion.

- **penalty**
  - The coefficient of the penalty for degrees of freedom in the GCV criterion.

- **control_spar**
  - Optional list with named components controlling the root finding when the smoothing parameter spar is computed, i.e., NULL. See help("smooth.spline") for further information.

- **tol_metric**
  - All cutpoints will be returned that lead to a metric value in the interval [m_max - tol_metric, m_max + tol_metric] where m_max is the maximum achievable metric value. This can be used to return multiple decent cutpoints and to avoid floating-point problems.
**use_midpoints**  (logical) If TRUE (default FALSE) the returned optimal cutpoint will be the mean of the optimal cutpoint and the next highest observation (for direction = ">") or the next lowest observation (for direction = "<") which avoids biasing the optimal cutpoint.

...  Further arguments that will be passed to metric_func.

**Details**

The above inputs are arrived at by using all unique values in x, Inf, and -Inf as possible cutpoints for classifying the variable in class.

**Value**

A tibble with the columns optimal_cutpoint, the corresponding metric value and roc_curve, a nested tibble that includes all possible cutoffs and the corresponding numbers of true and false positives / negatives and all corresponding metric values.

**See Also**

Other method functions: maximize_boot_metric(), maximize_gam_metric(), maximize_loess_metric(), maximize_metric(), oc_manual(), oc_mean(), oc_median(), oc_youden_kernel(), oc_youden_normal()

**Examples**

```r
oc <- cutpointr(suicide, dsi, suicide, gender, method = maximize_spline_metric, df = 5, metric = accuracy)
plot_metric(oc)
```

---

### metric_constrain

**Metrics that are constrained by another metric**

**Description**

For example, calculate sensitivity where a lower bound (minimal desired value) for specificity can be defined. All returned metric values for cutpoints that lead to values of the constraining metric below the specified minimum will be zero. The inputs must be vectors of equal length.

**Usage**

```r
metric_constrain(
  tp,
  fp,
  tn,
  fn,
  main_metric = sensitivity,
  constrain_metric = specificity,
  min_constrain = 0.5,
  suffix = "_constrain",
)```

...  
)  
sens_constrain(  
    tp,  
    fp,  
    tn,  
    fn,  
    constrain_metric = specificity,  
    min_constrain = 0.5,  
    ...  
)  

spec_constrain(  
    tp,  
    fp,  
    tn,  
    fn,  
    constrain_metric = sensitivity,  
    min_constrain = 0.5,  
    ...  
)  

acc_constrain(  
    tp,  
    fp,  
    tn,  
    fn,  
    constrain_metric = sensitivity,  
    min_constrain = 0.5,  
    ...  
)  

Arguments  
  
  tp             (numeric) number of true positives.  
  fp             (numeric) number of false positives.  
  tn             (numeric) number of true negatives.  
  fn             (numeric) number of false negatives.  
  main_metric    Metric to be optimized.  
  constrain_metric Metric for constraint.  
  min_constrain  Minimum desired value of constrain_metric.  
  suffix         Character string to be added to the name of main_metric.  
  ...            for capturing additional arguments passed by method.
misclassification_cost

Description

Calculate the misclassification cost from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

misclassification_cost = cost_fp * fp + cost_fn * fn

Usage

misclassification_cost(tp, fp, tn, fn, cost_fp = 1, cost_fn = 1, ...)

Arguments

tp (numeric) number of true positives.
fp (numeric) number of false positives.
tn (numeric) number of true negatives.
fn (numeric) number of false negatives.
cost_fp (numeric) the cost of a false positive

cost_fn (numeric) the cost of a false negative

... for capturing additional arguments passed by method.

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cohens_kappa(), cutpoint(), false omission_rate(), misclassification_cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_ppv_npv(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()
Examples

```r
misclassification_cost(10, 5, 20, 10, cost_fp = 1, cost_fn = 5)
misclassification_cost(c(10, 8), c(5, 7), c(20, 12), c(10, 18),
                      cost_fp = 1, cost_fn = 5)
```

---

**multi_cutpointtr**

*Calculate optimal cutpoints and further statistics for multiple predictors*

---

**Description**

Runs `cutpointtr_` over multiple predictor variables. If `x = NULL`, `cutpointtr_` will be run using all numeric columns in the data set as predictors except for the variable in `class` and, if given, `subgroup`.

**Usage**

```r
multi_cutpointtr(data, x = NULL, class, subgroup, silent = FALSE, ...)
```

**Arguments**

- **data**: A data frame.
- **x**: Character vector of predictor variables. If NULL all numeric columns.
- **class**: The name of the outcome / independent variable.
- **subgroup**: An additional covariate that identifies subgroups. Separate optimal cutpoints will be determined per group.
- **silent**: Whether to suppress messages.
- **...**: Further arguments to be passed to `cutpointtr_` (Use a quoted variable name for `subgroup`).

**Details**

The automatic determination of positive / negative classes and `direction` will be carried out separately for every predictor variable. That way, if `direction` and the classes are not specified, the reported AUC for every variable will be >= 0.5. AUC may be < 0.5 if subgroups are specified as `direction` is equal within every subgroup.

**Value**

A data frame.

**See Also**

Other main `cutpointtr` functions: `add_metric()`, `boot_ci()`, `boot_test()`, `cutpointtr()`, `predict.cutpointtr()`, `roc()`
Examples

```r
library(cutpointr)

multi_cutpointr(suicide, x = c("age", "dsi"), class = suicide,
               pos_class = "yes")

mcp <- multi_cutpointr(suicide, x = c("age", "dsi"), class = suicide,
                        subgroup = gender, pos_class = "yes")
mcp

(scp <- summary(mcp))
## Not run:
## The result is a data frame
tibble::print.tbl(scp)
## End(Not run)
```

npv

---

**Calculate the negative predictive value**

Description

Calculate the negative predictive value (NPV) from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
\text{npv} = \frac{\text{tn}}{\text{tn} + \text{fn}}
\]

Usage

```
npv(tp, fp, tn, fn, ...)
```

Arguments

- `tp`: (numeric) number of true positives.
- `fp`: (numeric) number of false positives.
- `tn`: (numeric) number of true negatives.
- `fn`: (numeric) number of false negatives.
- `...`: for capturing additional arguments passed by method.

See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false omission rate()`, `metric_constrain()`, `misclassification_cost()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()`
Examples

npv(10, 5, 20, 10)
npv(c(10, 8), c(5, 7), c(20, 12), c(10, 18))

oc_manual

Set a manual cutpoint for use with cutpointr

Description

This function simply returns cutpoint as the optimal cutpoint. Mainly useful if bootstrap estimates of the out-of-bag performance of a given cutpoint are desired, e.g. taking a cutpoint value from the literature.

Usage

oc_manual(cutpoint, ...)

Arguments

cutpoint (numeric) The fixed cutpoint.
...

To capture further arguments that are always passed to the method function by cutpointr. The cutpointr function passes data, x, class, metric_func, direction, pos_class and neg_class to the method function.

See Also

Other method functions: maximize_boot_metric(), maximize_gam_metric(), maximize_loess_metric(), maximize_metric(), maximize_spline_metric(), oc_mean(), oc_median(), oc_youden_kernel(), oc_youden_normal()

Examples

cutpointr(suicide, dsi, suicide, method = oc_manual, cutpoint = 4)

oc_mean

Use the sample mean as cutpoint

Description

The sample mean is calculated and returned as the optimal cutpoint.

Usage

oc_mean(data, x, trim = 0, ...)

Arguments

- **data**: A data frame or tibble in which the columns that are given in `x` and `class` can be found.
- **x**: (character) The variable name to be used for classification, e.g. predictions or test values.
- **trim**: The fraction (0 to 0.5) of observations to be trimmed from each end of `x` before the mean is computed. Values of `trim` outside that range are taken as the nearest endpoint.
- **...**: To capture further arguments that are always passed to the method function by `cutpointr`. The `cutpointr` function passes `data`, `x`, `class`, `metric_func`, `direction`, `pos_class` and `neg_class` to the method function.

See Also

Other method functions: `maximize_boot_metric()`, `maximize_gam_metric()`, `maximize_loess_metric()`, `maximize_metric()`, `maximize_spline_metric()`, `oc_manual()`, `oc_median()`, `oc_youden_kernel()`, `oc_youden_normal()`

Examples

```r
data(suicide)
oc_mean(suicide, "dsi")
cutpointr(suicide, dsi, suicide, method = oc_mean)
```

---

**oc_median**

*Use the sample median as cutpoint*

Description

The sample median is calculated and returned as the optimal cutpoint.

Usage

`oc_median(data, x, ...)`

Arguments

- **data**: A data frame or tibble in which the columns that are given in `x` and `class` can be found.
- **x**: (character) The variable name to be used for classification, e.g. predictions or test values.
- **...**: To capture further arguments that are always passed to the method function by `cutpointr`. The `cutpointr` function passes `data`, `x`, `class`, `metric_func`, `direction`, `pos_class` and `neg_class` to the method function.
See Also

Other method functions: maximize_boot_metric(), maximize_gam_metric(), maximize_loess_metric(), maximize_metric(), maximize_spline_metric(), oc_manual(), oc_mean(), oc_youden_kernel(), oc_youden_normal()

Examples

data(suicide)
oc_median(suicide, "dsi")
cutpointr(suicide, dsi, suicide, method = oc_median)

oc_youden_kernel

Determine an optimal cutpoint maximizing the Youden-Index based on kernel smoothed densities

Description

Instead of searching for an optimal cutpoint to maximize (sensitivity + specificity - 1) on the ROC curve, this function first smoothes the empirical distributions of \( x \) per class. The smoothing is done using a binned kernel density estimate. The bandwidth is automatically selected using the direct plug-in method.

Usage

oc_youden_kernel(data, x, class, pos_class, neg_class, direction, ...)

Arguments

data A data frame or tibble in which the columns that are given in \( x \) and class can be found.
x (character) The variable name to be used for classification, e.g. predictions or test values.
class (character) The variable name indicating class membership.
pos_class The value of class that indicates the positive class.
neg_class The value of class that indicates the negative class.
direction (character) Use ">=" or "<=" to select whether an \( x \) value >= or <= the cutoff predicts the positive class.
... To capture further arguments that are always passed to the method function by cutpointr. The cutpointr function passes data, \( x \), class, metric_func, direction, pos_class and neg_class to the method function.

Details

The functions for calculating the kernel density estimate and the bandwidth are both from KernSmooth with default parameters, except for the bandwidth selection, which uses the standard deviation as scale estimate.

The cutpoint is estimated as the cutpoint that maximizes the Youden-Index given by

\[
J = \max_c F_N(c) - G_N(c)
\]

where \( J \) and \( G \) are the smoothed distribution functions.
Determine an optimal cutpoint for the Youden-Index assuming normal distributions

An optimal cutpoint maximizing the Youden- or J-Index (sensitivity + specificity - 1) is calculated parametrically assuming normal distributions per class.

Usage

```r
oc_youden_normal(
  data,
  x,
  class,
  pos_class = NULL,
  neg_class = NULL,
  direction,
  ...
)
```

Arguments

data A data frame or tibble in which the columns that are given in x and class can be found.
odds_ratio

x (character) The variable name to be used for classification, e.g. predictions or test values.

class (character) The variable name indicating class membership.

pos_class The value of class that indicates the positive class.

neg_class The value of class that indicates the negative class.

direction (character) Use ">=" or "<=" to select whether an x value >= or <= the cutoff predicts the positive class.

... To capture further arguments that are always passed to the method function by cutpointr. The cutpointr function passes data, x, class, metric_func, direction, pos_class and neg_class to the method function.

See Also

Other method functions: maximize_boot_metric(), maximize_gam_metric(), maximize_loess_metric(), maximize_metric(), maximize_spline_metric(), oc_manual(), oc_mean(), oc_median(), oc_youden_kernel()

Examples

data(suicide)
octouden_normal(suicide, "dsi", "suicide",
pos_class = "yes", neg_class = "no", direction = ">=")
cutpointr(suicide, dsi, suicide, method = oc_youden_normal)

odds_ratio Calculate the odds ratio

Description

Calculate the (diagnostic) odds ratio from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

odds_ratio = (tp / fp) / (fn / tn)

Usage

odds_ratio(tp, fp, tn, fn, ...)

Arguments

tp (numeric) number of true positives.

fp (numeric) number of false positives.

tn (numeric) number of true negatives.

fn (numeric) number of false negatives.

... for capturing additional arguments passed by method.
See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cohens_kappa(), cutpoint(), false omission rate(), metric_constrain(), misclassification_cost(), npv(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_ppv_npv(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()

Examples

odds_ratio(10, 5, 20, 10)
odds_ratio(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
plot.multi_cutpointr

Plotting multi_cutpointr objects is currently not supported

Description

You can try plotting the data manually instead.

Usage

```r
## S3 method for class 'multi_cutpointr'
plot(x, ...)
```

Arguments

- `x` A multi_cutpointr object.
- `...` Further arguments.

plot.roc_cutpointr

Plot ROC curve from a cutpointr or roc_cutpointr object

Description

Given a cutpointr object this function plots the ROC curve(s) per subgroup, if given. Also plots a ROC curve from the output of roc().

Usage

```r
## S3 method for class 'roc_cutpointr'
plot(x, type = "line", ...)
```

Arguments

- `x` A cutpointr or roc_cutpointr object.
- `type` "line" for line plot (default) or "step" for step plot.
- `...` Additional arguments (unused).

See Also

Other cutpointr plotting functions: plot.cutpointr(), plot_cut_boot(), plot_cutpointr(), plot_metric_boot(), plot_metric(), plot_precision_recall(), plot_sensitivity_specificity(), plot_x()
Examples

```r
opt_cut <- cutpointr(suicide, dsi, suicide)
plot_roc(opt_cut, display_cutpoint = FALSE)

opt_cut_2groups <- cutpointr(suicide, dsi, suicide, gender)
plot_roc(opt_cut_2groups, display_cutpoint = TRUE)

roc_curve <- roc(suicide, x = dsi, class = suicide, pos_class = "yes",
                 neg_class = "no", direction = "\geq")
plot(roc_curve)
auc(roc_curve)
```

---

**plot_cutpointr**

*General purpose plotting function for cutpointr or roc_cutpointr objects*

### Description

Flexibly plot various metrics against all cutpoints or any other metric. The function can plot any metric based on a cutpointr or roc_cutpointr object. If cutpointr was run with bootstrapping, bootstrapped confidence intervals can be plotted. These represent the quantiles of the distribution of the y-variable grouped by x-variable over all bootstrap repetitions.

### Usage

```r
plot_cutpointr(
  x, 
  xvar = cutpoint, 
  yvar = sum_sens_spec, 
  conf_lvl = 0.95, 
  aspect_ratio = NULL
)
```

### Arguments

- **x**  
  A cutpointr or roc_cutpointr object.

- **xvar**  
  A function, typically cutpoint or a metric function.

- **yvar**  
  A function, typically a metric function.

- **conf_lvl**  
  (numeric) If bootstrapping was run and x is a cutpointr object, a confidence interval at the level of conf_lvl can be plotted. To plot no confidence interval set conf_lvl = 0.

- **aspect_ratio**  
  (numeric) Set to 1 to obtain a quadratic plot, e.g. for plotting a ROC curve.
Details

The arguments to `xvar` and `yvar` should be metric functions. Any metric function that is suitable for `cutpointr` can also be used in `plot_cutpointr`. Anonymous functions are also allowed. To plot all possible cutpoints, the utility function `cutpoint` can be used.

The functions for `xvar` and `yvar` may accept any or all of the arguments `tp`, `fp`, `tn`, or `fn` and return a numeric vector, a matrix or a data.frame. For more details on metric functions see vignette("cutpointr").

Note that confidence intervals can only be correctly plotted if the values of `xvar` are constant across bootstrap samples. For example, confidence intervals for `tpr` by `fpr` (a ROC curve) cannot be plotted, as the values of the false positive rate vary per bootstrap sample.

See Also

Other `cutpointr` plotting functions: `plot.cutpointr()`, `plot_cut_boot()`., `plot_metric_boot()`, `plot_metric()`, `plot_precision_recall()`, `plot_roc()`, `plot_sensitivity_specificity()`, `plot_x()`

Examples

```r
set.seed(1)
oc <- cutpointr(suicide, dsi, suicide, boot_runs = 10)

plot_cutpointr(o, cutpoint, F1_score)

## ROC curve
plot_cutpointr(o, fpr, tpr, aspect_ratio = 1)

## Custom function
plot_cutpointr(o, cutpoint, function(tp, tn, fp, fn, ...) tp / fp) +
   ggplot2::ggtitle("Custom metric") + ggplot2::ylab("value")
```

---

**plot_cut_boot**

Plot the bootstrapped distribution of optimal cutpoints from a cutpointr object

Description

Given a cutpointr object this function plots the bootstrapped distribution of optimal cutpoints. `cutpointr` has to be run with `boot_runs` > 0 to enable bootstrapping.

Usage

```r
plot_cut_boot(x, ...)
```
### plot_metric

**Plot a metric over all possible cutoffs from a cutpointr object**

#### Arguments

- `x` A cutpointr object.
- `conf_lvl` The confidence level of the bootstrap confidence interval. Set to 0 to draw no bootstrap confidence interval.
- `add_unsmoothed` Add the line of unsmoothed metric values to the plot. Applicable for some smoothing methods, e.g. maximize_gam_metric.

#### Description

If `maximize_metric` is used as method function in cutpointr the computed metric values over all possible cutoffs can be plotted. Generally, this works for method functions that return a ROC-curve including the metric value for every cutpoint along with the optimal cutpoint.

#### Usage

```r
plot_metric(x, conf_lvl = 0.95, add_unsmoothed = TRUE)
```

#### See Also

Other cutpointr plotting functions: `plot.cutpointr()`, `plot_cut_boot()`, `plot_cutpointr()`, `plot_metric_boot()`, `plot_precision_recall()`, `plot_roc()`, `plot_sensitivity_specificity()`, `plot_x()`
**plot_metric_boot**  
*Plot the bootstrapped metric distribution from a cutpointr object*

**Description**
Given a `cutpointr` object this function plots the bootstrapped metric distribution, i.e. the distribution of out-of-bag metric values. The metric depends on the function that was supplied to `metric` in the call to `cutpointr`. The `cutpointr` function has to be run with `boot_runs > 0` to enable bootstrapping.

**Usage**

```r
plot_metric_boot(x, ...)
```

**Arguments**

- `x`: A `cutpointr` object.
- `...`: Additional arguments (unused)

**See Also**
Other `cutpointr` plotting functions: `plot.cutpointr()`, `plot_cut_boot()`, `plot_cutpointr()`, `plot_metric()`, `plot_precision_recall()`, `plot_roc()`, `plot_sensitivity_specificity()`, `plot_x()`

**Examples**

```r
set.seed(300)
opt_cut <- cutpointr(suicide, dsi, suicide, boot_runs = 10)
plot_metric_boot(opt_cut)
```

**plot_precision_recall**  
*Precision recall plot from a cutpointr object*

**Description**
Given a `cutpointr` object this function plots the precision recall curve(s) per subgroup, if given.

**Usage**

```r
plot_precision_recall(x, display_cutpoint = TRUE, ...)
```
Arguments

x A cutpointr or roc_cutpointr object.
display_cutpoint (logical) Whether or not to display the optimal cutpoint as a dot on the ROC curve.

See Also

Other cutpointr plotting functions: plot.cutpointr(), plot_cut_boot(), plot_cutpointr(), plot_metric_boot(), plot_metric(), plot_roc(), plot_sensitivity_specificity(), plot_x()

Examples

library(cutpointr)

## Optimal cutpoint for dsi
data(suicide)
opt_cut <- cutpointr(suicide, dsi, suicide)
plot_precision_recall(opt_cut)

# plot_roc

plot_roc from a cutpointr or roc_cutpointr object

Description

Given a cutpointr object this function plots the ROC curve(s) per subgroup, if given. Also plots a ROC curve from the output of roc().

Usage

plot_roc(x, ...)

## S3 method for class 'cutpointr'
plot_roc(x, display_cutpoint = TRUE, type = "line", ...)

## S3 method for class 'roc_cutpointr'
plot_roc(x, type = "line", ...)

Arguments

x A cutpointr or roc_cutpointr object.
... Additional arguments (unused).
display_cutpoint (logical) Whether or not to display the optimal cutpoint as a dot on the ROC curve for cutpointr objects.
type "line" for line plot (default) or "step" for step plot.
**plot_sensitivity_specificity**  

*See Also*

Other cutpointr plotting functions: `plot.cutpointr()`, `plot_cut_boot()`, `plot_cutpointr()`, `plot_metric_boot()`, `plot_metric()`, `plot_precision_recall()`, `plot_sensitivity_specificity()`, `plot_x()`

**Examples**

```r
opt_cut <- cutpointr(suicide, dsi, suicide)
plot_roc(opt_cut, display_cutpoint = FALSE)

opt_cut_2groups <- cutpointr(suicide, dsi, suicide, gender)
plot_roc(opt_cut_2groups, display_cutpoint = TRUE)

roc_curve <- roc(suicide, x = dsi, class = suicide, pos_class = "yes",
                 neg_class = "no", direction = ">="
plot(roc_curve)
auc(roc_curve)
```

---

**plot_sensitivity_specificity**  

*Sensitivity and specificity plot from a cutpointr object*

**Description**

Given a cutpointr object this function plots the sensitivity and specificity curve(s) per subgroup, if the latter is given.

**Usage**

```r
plot_sensitivity_specificity(x, display_cutpoint = TRUE, ...)
```

**Arguments**

- `x`  
  A cutpointr object.
- `display_cutpoint`  
  (logical) Whether or not to display the optimal cutpoint as a dot on the precision recall curve.
- `...`  
  Additional arguments (unused).

**See Also**

Other cutpointr plotting functions: `plot.cutpointr()`, `plot_cut_boot()`, `plot_cutpointr()`, `plot_metric_boot()`, `plot_metric()`, `plot_precision_recall()`, `plot_roc()`, `plot_x()`
Examples

```r
library(cutpointr)

## Optimal cutpoint for dsi
data(suicide)
opt_cut <- cutpointr(suicide, dsi, suicide)
plot_sensitivity_specificity(opt_cut)
```

---

`plot_x`  
*Plot the distribution of the independent variable per class from a cutpointr object*

---

**Description**

Given a `cutpointr` object this function plots the distribution(s) of the independent variable(s) and the respective cutpoints per class.

**Usage**

```r
plot_x(x, display_cutpoint = TRUE, ...)
```

**Arguments**

- `x`: A `cutpointr` object.
- `display_cutpoint` (logical): Whether or not to display the optimal cutpoint as a vertical line.
- `...`: Additional arguments (unused).

**See Also**

Other `cutpointr` plotting functions: `plot.cutpointr()`, `plot_cut_boot()`, `plot_cutpointr()`, `plot_metric_boot()`, `plot_metric()`, `plot_precision_recall()`, `plot_roc()`, `plot_sensitivity_specificity()`

**Examples**

```r
opt_cut <- cutpointr(suicide, dsi, suicide)
plot_x(opt_cut)

## With subgroup
opt_cut_2groups <- cutpointr(suicide, dsi, suicide, gender)
plot_x(opt_cut_2groups)
```
Calculate the positive or negative likelihood ratio from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
\text{plr} = \frac{\text{tpr}}{\text{fpr}} \\
\text{nlr} = \frac{\text{fnr}}{\text{tnr}}
\]

Usage

\[
\text{plr}(\text{tp}, \text{fp}, \text{tn}, \text{fn}, \ldots) \\
\text{nlr}(\text{tp}, \text{fp}, \text{tn}, \text{fn}, \ldots)
\]

Arguments

- **tp** (numeric) number of true positives.
- **fp** (numeric) number of false positives.
- **tn** (numeric) number of true negatives.
- **fn** (numeric) number of false negatives.
- **...** for capturing additional arguments passed by method.

See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()

Examples

\[
\text{plr}(10, 5, 20, 10) \\
\text{plr}(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
\]
Calculate the positive predictive value (PPV) from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[ \text{ppv} = \frac{tp}{tp + fp} \]

**Usage**

\[ \text{ppv}(tp, fp, tn, fn, ...) \]

**Arguments**

- `tp` (numeric) number of true positives.
- `fp` (numeric) number of false positives.
- `tn` (numeric) number of true negatives.
- `fn` (numeric) number of false negatives.
- `...` for capturing additional arguments passed by method.

**See Also**

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false omission rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()`

**Examples**

```
ppv(10, 5, 20, 10)
ppv(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
```
Calculate precision (equal to the positive predictive value) from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[ \text{precision} = \frac{tp}{tp + fp} \]

### Usage

```r
precision(tp, fp, tn, fn, ...)
```

### Arguments

- **tp**: (numeric) number of true positives.
- **fp**: (numeric) number of false positives.
- **tn**: (numeric) number of true negatives.
- **fn**: (numeric) number of false negatives.
- **...**: for capturing additional arguments passed by method.

### See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohen_s_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()`

### Examples

```r
precision(10, 5, 20, 10)
precision(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
```
predict.cutpointr  Predict using a cutpointr object

Description

Predictions are made on the data.frame in newdata using either the variable name or by applying the same transformation to the data as in cutpointr. The class of the output will be identical to the class of the predictor.

Usage

## S3 method for class 'cutpointr'
predict(object, newdata, cutpoint_nr = 1, ...)

Arguments

- **object**: a cutpointr object.
- **newdata**: a data.frame with a column that contains the predictor variable.
- **cutpoint_nr**: if multiple optimal cutpoints were found this parameter defines which one should be used for predictions. Can be a vector if different cutpoint numbers are desired for different subgroups.
- **...**: further arguments.

See Also

Other main cutpointr functions: `add_metric()`, `boot_ci()`, `boot_test()`, `cutpointr()`, `multi_cutpointr()`, `roc()`

Examples

oc <- cutpointr(suicide, dsi, suicide)
## Return in-sample predictions
predict(oc, newdata = data.frame(dsi = oc$data[[1]]$dsi))

print.cutpointr  Print cutpointr objects

Description

Prints the cutpointr object with full width like a tbl_df.

Usage

## S3 method for class 'cutpointr'
print(x, width = 1000, n = 50, sigfig = 6, ...)


print.multi_cutpointr

Arguments

x             a cutpointr object.
width         width of output.
n             number of rows to print.
sigfig        Number of significant digits to print. Temporarily overrides options("pillar.sigfig").
...           further arguments.

Source


print.multi_cutpointr  Print multi_cutpointr objects

Description

Prints the multi_cutpointr object with infinite width like a tbl_df.

Usage

## S3 method for class 'multi_cutpointr'
print(x, n = Inf, ...)

Arguments

x             a multi_cutpointr object.
n             number of rows to print.
...           further arguments.

Source

prod_ppv_npv

Calculate the product of positive and negative predictive value

Description

Calculate the product of positive predictive value (PPV) and negative predictive value (NPV) from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
ppv = \frac{tp}{tp + fp} \\
npv = \frac{tn}{tn + fn} \\
prod_{ppv\_npv} = ppv \times npv
\]

Usage

\[
prod_{ppv\_npv}(tp, fp, tn, fn, \ldots)
\]

Arguments

- **tp** (numeric): number of true positives.
- **fp** (numeric): number of false positives.
- **tn** (numeric): number of true negatives.
- **fn** (numeric): number of false negatives.
- **...** for capturing additional arguments passed by method.

See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `precision()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()`

Examples

\[
prod_{ppv\_npv}(10, 5, 20, 10) \\
prod_{ppv\_npv}(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
\]
prod_sens_spec

*Calculate the product of sensitivity and specificity*

**Description**

Calculate the product of sensitivity and specificity from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
\text{sensitivity} = \frac{tp}{tp + fn} \\
\text{specificity} = \frac{tn}{tn + fp} \\
\text{prod_sens_spec} = \text{sensitivity} \times \text{specificity}
\]

**Usage**

```r
prod_sens_spec(tp, fp, tn, fn, ...)
```

**Arguments**

- `tp` (numeric) number of true positives.
- `fp` (numeric) number of false positives.
- `tn` (numeric) number of true negatives.
- `fn` (numeric) number of false negatives.
- `...` for capturing additional arguments passed by method.

**See Also**

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()`

**Examples**

```r
prod_sens_spec(10, 5, 20, 10)
prod_sens_spec(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
```
prostate_nodal  

**Nodal involvement and acid phosphatase levels in 53 prostate cancer patients**

**Description**

Prostatic acid phosphatase (PAP) emerged as the first clinically useful tumor marker in the 1940s and 1950s. This data set contains the serum levels of acid phosphatase of 53 patients that were confirmed to have prostate cancer and whether the neighboring lymph nodes were involved.

**Usage**

```r
prostate_nodal
```

**Format**

A data frame with 53 rows and 2 variables:

- **acid_phosphatase** (numeric) Blood serum level of acid phosphatase
- **nodal_involvement** (logical) Whether neighboring lymph nodes were involved

**Source**


---

**p_chisquared**

**Calculate the p-value of a chi-squared test**

**Description**

Calculate the p-value of a chi-squared test from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

**Usage**

```r
p_chisquared(tp, fp, tn, fn, ...)
```

**Arguments**

- **tp** (numeric) number of true positives.
- **fp** (numeric) number of false positives.
- **tn** (numeric) number of true negatives.
- **fn** (numeric) number of false negatives.
- `...` for capturing additional arguments passed by method.
See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()

Examples

```
p_chisquared(10, 5, 20, 10)
p_chisquared(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
```

Recall

Recall (equal to sensitivity) from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

```
recall = tp / (tp + fn)
```

Usage

```
recall(tp, fp, tn, fn, ...)
```

Arguments

- `tp` (numeric) number of true positives.
- `fp` (numeric) number of false positives.
- `tn` (numeric) number of true negatives.
- `fn` (numeric) number of false negatives.
- `...` for capturing additional arguments passed by method.

See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()

Examples

```
recall(10, 5, 20, 10)
recall(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
```
Calculate the risk ratio (relative risk)

Description

Calculate the risk ratio (or relative risk) from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
risk\_ratio = \frac{tp / (tp + fn)}{fp / (fp + tn)}
\]

Usage

\[risk\_ratio(tp, fp, tn, fn, \ldots)\]

Arguments

- **tp**: (numeric) number of true positives.
- **fp**: (numeric) number of false positives.
- **tn**: (numeric) number of true negatives.
- **fn**: (numeric) number of false negatives.
- **\ldots**: for capturing additional arguments passed by method.

See Also

Other metric functions: \texttt{F1\_score()}, \texttt{Jaccard()}, \texttt{abs\_d\_ppv\_npv()}, \texttt{abs\_d\_sens\_spec()}, \texttt{accuracy()}, \texttt{cohens\_kappa()}, \texttt{cutpoint()}, \texttt{false\_omission\_rate()}, \texttt{metric\_constrain()}, \texttt{misclassification\_cost()}, \texttt{npv()}, \texttt{odds\_ratio()}, \texttt{p\_chisquared()}, \texttt{plr()}, \texttt{ppv()}, \texttt{precision()}, \texttt{prod\_ppv\_npv()}, \texttt{prod\_sens\_spec()}, \texttt{recall()}, \texttt{roc01()}, \texttt{sensitivity()}, \texttt{specificity()}, \texttt{sum\_ppv\_npv()}, \texttt{sum\_sens\_spec()}, \texttt{total\_utility()}, \texttt{tpr()}, \texttt{tp()}, \texttt{youden()}

Examples

\[risk\_ratio(10, 5, 20, 10)\]
\[risk\_ratio(c(10, 8), c(5, 7), c(20, 12), c(10, 18))\]
Calculate a ROC curve

Description

Given a data.frame with a numeric predictor variable and a binary outcome variable this function returns a data.frame that includes all elements of the confusion matrix (true positives, false positives, true negatives, and false negatives) for every unique value of the predictor variable. Additionally, the true positive rate (tpr), false positive rate (fpr), true negative rate (tnr) and false negative rate (fnr) are returned.

Usage

roc(data, x, class, pos_class, neg_class, direction = ">=", silent = FALSE)

Arguments

data
A data.frame or matrix. Will be converted to a data.frame.

x
The name of the numeric predictor variable.

class
The name of the binary outcome variable.

pos_class
The value of 'class' that represents the positive cases.

neg_class
The value of 'class' that represents the negative cases.

direction
(character) One of ">=" or "<=". Specifies if the positive class is associated with higher values of x (default).

silent
If FALSE and the ROC curve contains no positives or negatives, a warning is generated.

Details

To enable classifying all observations as belonging to only one class the predictor values will be augmented by Inf or -Inf. The returned object can be plotted with plot_roc.

This function uses tidyeval to support unquoted arguments. For programming with roc the operator !! can be used to unquote an argument, see the examples.

Value

A data frame with the columns x.sorted, tp, fp, tn, fn, tpr, tnr, fpr, and fnr.

Source

Forked from the ROCR package

See Also

Other main cutpointr functions: add_metric(), boot_ci(), boot_test(), cutpointr(), multi_cutpointr(), predict.cutpointr()
Examples

```r
roc_curve <- roc(data = suicide, x = dsi, class = suicide,
                  pos_class = "yes", neg_class = "no", direction = ">=")
roc_curve
plot_roc(roc_curve)
auc(roc_curve)

## Unquoting an argument
myvar <- "dsi"
roc(suicide, x = !!myvar, suicide, pos_class = "yes", neg_class = "no")
```

---

**roc01**

_Calculate the distance between points on the ROC curve and (0,1)_

**Description**

Calculate the distance on the ROC space between points on the ROC curve and the point of perfect discrimination from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length. To be used with `method = minimize_metric`.

```r
sensitivity = tp / (tp + fn)
specificity = tn / (tn + fp)
roc01 = sqrt((1 - sensitivity)^2 + (1 - specificity)^2)
```

**Usage**

```r
roc01(tp, fp, tn, fn, ...)
```

**Arguments**

- `tp` (numeric) number of true positives.
- `fp` (numeric) number of false positives.
- `tn` (numeric) number of true negatives.
- `fn` (numeric) number of false negatives.
- `...` for capturing additional arguments passed by method.

**See Also**

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `tp()`, `youden()`
Examples

roc01(10, 5, 20, 10)
roc01(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
oc <- cutpointr(suicide, dsi, suicide,
method = minimize_metric, metric = roc01)
plot_roc(o)

sensitivity

Calculate sensitivity

Description

Calculate sensitivity from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
sensitivity = \frac{tp}{tp + fn}
\]

Usage

sensitivity(tp, fn, ...)

Arguments

tp (numeric) number of true positives.
fn (numeric) number of false negatives.
... for capturing additional arguments passed by method.

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cohens_kappa(), cutpoint(), false omission rate(), metric_constrain(), misclassification cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), specificity(), sum_ppv_npv(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()

Examples

sensitivity(10, 5, 20, 10)
sensitivity(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
**specificity**  
*Calculate specificity*

**Description**

Calculate specificity from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
\text{specificity} = \frac{\text{tn}}{\text{tn} + \text{fp}}
\]

**Usage**

```r
specificity(fp, tn, ...)
```

**Arguments**

- `fp` (numeric) number of false positives.
- `tn` (numeric) number of true negatives.
- `...` for capturing additional arguments passed by method.

**See Also**

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tp()`, `tp()`, `youden()`

**Examples**

```r
specificity(10, 5, 20, 10)
specificity(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
```

---

**suicide**  
*Suicide attempts and DSI sum scores of 532 subjects*

**Description**

Various personality and clinical psychological characteristics were assessed as part of an online-study preventing suicide. To identify persons at risk for attempting suicide, various demographic and clinical characteristics were assessed. Depressive Symptom Inventory - Suicidality Subscale (DSA-SS) sum scores and past suicide attempts from 532 subjects are included as a demonstration set to calculate optimal cutpoints. Two additional demographic variables (age, gender) are also included to test for group differences.
Usage

suicide

Format

A data frame with 532 rows and 4 variables:

age  (numeric) Age of participants in years

gender  (factor) Gender

dsi  (numeric) Sum-score (0 = low suicidality, 12 = high suicidality)
suicide  (factor) Past suicide attempt (no = no attempt, yes = at least one attempt)

Source


Description

Calculate the sum of positive and negative predictive value (PPV) and negative predictive value (NPV) from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

ppv = tp / (tp + fp)
npv = tn / (tn + fn)
sum_ppv_npv = ppv + npv

Usage

sum_ppv_npv(tp, fp, tn, fn, ...)

Arguments

tp  (numeric) number of true positives.
fp  (numeric) number of false positives.
tn  (numeric) number of true negatives.
fn  (numeric) number of false negatives.
...  for capturing additional arguments passed by method.
sum_sens_spec

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cohens_kappa(), cutpoint(), false omission rate(), metric_constrain(), misclassification_cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()

Examples

sum_ppv_npv(10, 5, 20, 10)
sum_ppv_npv(c(10, 8), c(5, 7), c(20, 12), c(10, 18))

sum_sens_spec (Calculate the sum of sensitivity and specificity

Description

Calculate the sum of sensitivity and specificity from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

sensitivity = tp / (tp + fn)
specificity = tn / (tn + fp)
sum_sens_spec = sensitivity + specificity

Usage

sum_sens_spec(tp, fp, tn, fn, ...)

Arguments

tp (numeric) number of true positives.
fp (numeric) number of false positives.
tn (numeric) number of true negatives.
fn (numeric) number of false negatives.
... for capturing additional arguments passed by method.

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cohens_kappa(), cutpoint(), false omission rate(), metric_constrain(), misclassification_cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_sens_spec(), total_utility(), tpr(), tp(), youden()
total_utility

Examples

sum_sens_spec(10, 5, 20, 10)
sum_sens_spec(c(10, 8), c(5, 7), c(20, 12), c(10, 18))

total_utility  Calculate the total utility

Description

Calculate the total utility from true positives, false positives, true negatives and false negatives.

total_utility = utility_tp * tp + utility_tn * tn - cost_fp * fp - cost_fn * fn

The inputs must be vectors of equal length.

Usage

total_utility(
  tp,
  fp,
  tn,
  fn,
  utility_tp = 1,
  utility_tn = 1,
  cost_fp = 1,
  cost_fn = 1,
  ...
)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tp</td>
<td>(numeric) number of true positives.</td>
</tr>
<tr>
<td>fp</td>
<td>(numeric) number of false positives.</td>
</tr>
<tr>
<td>tn</td>
<td>(numeric) number of true negatives.</td>
</tr>
<tr>
<td>fn</td>
<td>(numeric) number of false negatives.</td>
</tr>
<tr>
<td>utility_tp</td>
<td>(numeric) the utility of a true positive</td>
</tr>
<tr>
<td>utility_tn</td>
<td>(numeric) the utility of a true negative</td>
</tr>
<tr>
<td>cost_fp</td>
<td>(numeric) the cost of a false positive</td>
</tr>
<tr>
<td>cost_fn</td>
<td>(numeric) the cost of a false negative</td>
</tr>
<tr>
<td>...</td>
<td>for capturing additional arguments passed by method.</td>
</tr>
</tbody>
</table>
See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `tpr()`, `tp()`, `youden()`

Examples

total_utility(10, 5, 20, 10, utility_tp = 3, utility_tn = 3, cost_fp = 1, cost_fn = 5)
total_utility(c(10, 8), c(5, 7), c(20, 12), c(10, 18),
              utility_tp = 3, utility_tn = 3, cost_fp = 1, cost_fn = 5)

---

tp  Extract number true / false positives / negatives

Description

Extract the number of true positives (tp), false positives (fp), true negatives (tn), or false negatives (fn). The inputs must be vectors of equal length. Mainly useful for `plot_cutpoint()`.

Usage

    tp(tp, ...)
    tn(tn, ...)
    fp(fp, ...)
    fn(fn, ...)

Arguments

    tp  (numeric) number of true positives.
    ... for capturing additional arguments passed by method.
    tn  (numeric) number of true negatives.
    fp  (numeric) number of false positives.
    fn  (numeric) number of false negatives.

See Also

Other metric functions: `F1_score()`, `Jaccard()`, `abs_d_ppv_npv()`, `abs_d_sens_spec()`, `accuracy()`, `cohens_kappa()`, `cutpoint()`, `false_omission_rate()`, `metric_constrain()`, `misclassification_cost()`, `npv()`, `odds_ratio()`, `p_chisquared()`, `plr()`, `ppv()`, `precision()`, `prod_ppv_npv()`, `prod_sens_spec()`, `recall()`, `risk_ratio()`, `roc01()`, `sensitivity()`, `specificity()`, `sum_ppv_npv()`, `sum_sens_spec()`, `total_utility()`, `tpr()`, `youden()`
Examples

tp(10, 5, 20, 10)
tp(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
fp(10, 5, 20, 10)
tn(10, 5, 20, 10)
fn(10, 5, 20, 10)

Description

Calculate the true positive rate (tpr, equal to sensitivity and recall), the false positive rate (fpr, equal
to fall-out), the true negative rate (tnr, equal to specificity), or the false negative rate (fnr) from true
positives, false positives, true negatives and false negatives. The inputs must be vectors of equal
length.

tpr = tp / (tp + fn)
fpr = fp / (fp + tn)
tnr = tn / (tn + fp)
fnr = fn / (fn + tp)

Usage

tpr(tp, fn, ...)
fpr(fp, tn, ...)
tnr(fp, tn, ...)
fnr(tp, fn, ...)

Arguments

tp (numeric) number of true positives.
fn (numeric) number of false negatives.
... for capturing additional arguments passed by method.
fp (numeric) number of false positives.
 tn (numeric) number of true negatives.

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(),
cohens_kappa(), cutpoint(), false_omission_rate(), metric_constrain(), misclassification_cost(),
npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(),
recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_ppv_npv(), sum_sens_spec(),
total_utility(), tp(), youden()
Examples

\[
tpr(10, 5, 20, 10)
tpr(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
\]

user_span_cutpointr Calculate bandwidth for LOESS smoothing of metric functions by rule of thumb

Description

This function implements a rule of thumb for selecting the bandwidth when smoothing a function of metric values per cutpoint value, particularly in maximize_loess_metric and minimize_loess_metric.

Usage

user_span_cutpointr(data, x)

Arguments

data A data frame
x The predictor variable

Details

The function used for calculating the bandwidth is \(0.1 \times \text{xsd} / \sqrt{\text{xn}}\), where \(\text{xsd}\) is the standard deviation of the unique values of the predictor variable (i.e. all cutpoints) and \(\text{xn}\) is the number of unique predictor values.

youden Calculate the Youden-Index

Description

Calculate the Youden-Index (J-Index) from true positives, false positives, true negatives and false negatives. The inputs must be vectors of equal length.

\[
sensitivity = \frac{tp}{tp + fn} \\
\text{specificity} = \frac{tn}{tn + fp} \\
youden\_index = sensitivity + specificity - 1
\]

Usage

youden(tp, fp, tn, fn, ...)

\[
sensitivity = \frac{tp}{tp + fn} \\
\text{specificity} = \frac{tn}{tn + fp} \\
youden\_index = sensitivity + specificity - 1
\]
Arguments

tp (numeric) number of true positives.
fp (numeric) number of false positives.
tn (numeric) number of true negatives.
fn (numeric) number of false negatives.
... for capturing additional arguments passed by method.

See Also

Other metric functions: F1_score(), Jaccard(), abs_d_ppv_npv(), abs_d_sens_spec(), accuracy(), cohens_kappa(), cutpoint(), false omission rate(), metric_constrain(), misclassification_cost(), npv(), odds_ratio(), p_chisquared(), plr(), ppv(), precision(), prod_ppv_npv(), prod_sens_spec(), recall(), risk_ratio(), roc01(), sensitivity(), specificity(), sum_ppv_npv(), sum_sens_spec(), total_utility(), tpr(), tp() 

Examples

youden(10, 5, 20, 10)
youden(c(10, 8), c(5, 7), c(20, 12), c(10, 18))
Index

*Topic datasets
prostate_nodal, 60
suicide, 66

abs_d_ppv_npv, 3, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
abs_d_sens_spec, 4, 4, 5, 10, 11, 20–22, 36,38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
acc_constrain (metric_constrain), 34
accuracy, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
add_metric, 6, 8, 9, 16, 37, 56, 63
auc, 6

boot_ci, 6, 7, 9, 16, 37, 56, 63
boot_test, 6, 8, 8, 16, 37, 56, 63

cohens_kappa, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
cutpoint, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
cutpoint_knots, 19
cutpointr, 6, 8, 9, 11, 37, 56, 63
cutpointr_, 17
cutpoints (cutpoint), 11

F1_score, 4, 5, 10, 11, 20, 21, 22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
false_discovery_rate (false_omission_rate), 21
false_omission_rate, 4, 5, 10, 11, 20, 21, 22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
fn (tp), 70

fpr (tp), 71
fp (tp), 70
fpr (tpr), 71

Jaccard, 4, 5, 10, 11, 20, 21, 22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
maximize_boot_metric, 23, 37, 29, 31, 34, 39–43
maximize_gam_metric, 25, 29, 31, 34, 39–43
maximize_loess_metric, 25, 27, 29, 31, 34, 39–43
maximize_metric, 25, 27, 29, 30, 34, 39–43
maximize_spline_metric, 25, 27, 29, 31, 32, 39–43

metric_constrain, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
minimize_boot_metric
(maximize_boot_metric), 23
minimize_gam_metric
(maximize_gam_metric), 25
minimize_loess_metric
(maximize_loess_metric), 27
minimize_metric (maximize_metric), 30
minimize_spline_metric
(maximize_spline_metric), 32
misclassification_cost, 4, 5, 10, 11, 20–22, 36, 38, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
multi_cutpoint, 6, 8, 9, 16, 37, 56, 63

nlr (plr), 53
npv, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73

oc_manual, 25, 27, 29, 31, 34, 39, 40–43
oc_mean, 25, 27, 29, 31, 34, 39, 39, 41–43
INDEX

oc_median, 25, 27, 29, 31, 34, 39, 40, 42, 43
oc_youden_kernel, 25, 27, 29, 31, 34, 39–41, 41, 43
oc_youden_normal, 25, 27, 29, 31, 34, 39–42, 42
odds_ratio, 4, 5, 10, 11, 20–22, 36, 38, 43, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
p_chisquared, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 60, 61, 62, 64–66, 68, 70, 71, 73
plot.cutpointr, 44, 45, 47–52
plot.multi_cutpointr, 45
plot roc cutpointr, 45
plot_cut_boot, 44, 45, 47, 48–52
plot_cutpointr, 44, 45, 46, 48–52
plot_metric, 44, 45, 47, 48, 49, 49–52
plot_metric_boot, 44, 45, 47, 48, 49, 50–52
plot_precision_recall, 44, 45, 47–49, 49, 51, 52
plot_roc, 44, 47–50, 50, 51, 52
plot_sensitivity_specificity, 44, 45, 47–51, 51, 52
plot_x, 44, 45, 47–51, 52
plr, 4, 5, 10, 11, 20–22, 36, 38, 44, 53, 54, 55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
ppv, 4, 5, 10, 11, 20–22, 36, 38, 44, 53, 54, 55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
precision, 4, 5, 10, 11, 20–22, 36, 38, 44, 53, 54, 55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
predict.cutpointr, 6, 8, 9, 16, 37, 56, 63
print.cutpointr, 56
print.multi_cutpointr, 57
prod_ppv_npv, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
prod_sens_spec, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
prostate_nodal, 60
recall, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
risk_ratio, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
roc, 6, 8, 9, 16, 37, 56, 63
roc01, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64, 65, 66, 68, 70, 71, 73
sens_constrain (metric_constrain), 34
spec_constrain (metric_constrain), 34
specificity, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64, 65, 66, 68, 70, 71, 73
suicide, 66
sum_sens_spec, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 68, 70, 71, 73
tn (tp), 70
tnr (tpr), 71
total_utility, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 69, 70, 71, 73
tp, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
tpr, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 73
user_span_cutpointr, 72
youden, 4, 5, 10, 11, 20–22, 36, 38, 44, 53–55, 58, 59, 61, 62, 64–66, 68, 70, 71, 72

71, 73