

Package ‘ROCnGO’

July 17, 2025

Title Fast Analysis of ROC Curves

Version 0.1.0

Description A toolkit for analyzing classifier performance by using receiver operating characteristic (ROC) curves. Performance may be assessed on a single classifier or multiple ones simultaneously, making it suitable for comparisons. In addition, different metrics allow the evaluation of local performance when working within restricted ranges of sensitivity and specificity. For details on the different implementations, see McClish D. K. (1989) <[doi:10.1177/0272989X8900900307](https://doi.org/10.1177/0272989X8900900307)>, Vivo J.-M., Franco M. and Vicari D. (2018) <[doi:10.1007/S11634-017-0295-9](https://doi.org/10.1007/S11634-017-0295-9)>, Jiang Y., et al (1996) <[doi:10.1148/radiology.201.3.8939225](https://doi.org/10.1148/radiology.201.3.8939225)>, Franco M. and Vivo J.-M. (2021) <[doi:10.3390/math9212826](https://doi.org/10.3390/math9212826)> and Carrington, André M., et al (2020) <[doi:10.1186/s12911-019-1014-6](https://doi.org/10.1186/s12911-019-1014-6)>.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.3.2

Imports cli, dplyr, forcats, ggplot2, magrittr, purrr, rlang, stringr, SummarizedExperiment, tibble, tidyr

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

URL <https://pablopnc.github.io/ROCnGO/>,
<https://github.com/pabloPNC/ROCnGO>

Depends R (>= 4.1.0)

BugReports <https://github.com/pabloPNC/ROCnGO/issues>

NeedsCompilation no

Author Pablo Navarro [aut, cre, cph],
Juana-María Vivo [aut],
Manuel Franco [aut]

Maintainer Pablo Navarro <pablo.navarrocarpio@gmail.com>

Repository CRAN

Date/Publication 2025-07-17 20:30:18 UTC

Contents

add_chance_line	2
add_fpauc_partially_proper_lower_bound	3
add_fpr_threshold_line	5
add_partial_roc_curve	6
add_partial_roc_points	7
add_roc_curve	9
add_roc_points	10
add_tpauc_concave_lower_bound	11
auc	13
calc_curve_shape	14
calc_partial_roc_points	16
concordance_indexes	17
hide_legend	19
npauc_lower_bounds	20
pauc	21
plot_partial_roc_curve	23
plot_partial_roc_points	24
plot_roc_curve	26
plot_roc_points	27
roc_points	28
sensitivity_indexes	29
spauc_lower_bounds	31
sp_auc	32
sumexp_to_df	34
summarize_dataset	35
summarize_predictor	36
Index	39

add_chance_line	<i>Show chance line in a ROC plot</i>
-----------------	---------------------------------------

Description

Plot chance line in a ROC plot.

Usage

add_chance_line()

Value

A ggplot layer instance object.

Examples

```
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +  
  add_chance_line()
```

```
add_fpauc_partially_proper_lower_bound  
  Add FpAUC lower bound to a ROC plot
```

Description

Calculate and plot lower bound defined by FpAUC sensitivity index.

- `add_fpauc_lower_bound()` provides an upper level function which automatically calculates curve shape and plots a lower bound that better fits it.
- `add_fpauc_partially_proper_lower_bound()` and `add_fpauc_concave_lower_bound()` are lower level functions that enforce the plot of specific bounds.

First one plots lower bound when curve shape is partially proper (presents some kind of hook).
Second one plots lower bound when curve shape is concave in the region of interest.

Usage

```
add_fpauc_partially_proper_lower_bound(  
  data,  
  response = NULL,  
  predictor = NULL,  
  threshold,  
  .condition = NULL,  
  .label = NULL  
)
```

```
add_fpauc_concave_lower_bound(  
  data,  
  response = NULL,  
  predictor = NULL,  
  threshold,  
  .condition = NULL,  
  .label = NULL  
)
```

```
add_fpauc_lower_bound(  
  data,  
  response = NULL,  
  predictor = NULL,  
  threshold,  
  .condition = NULL,  
  .label = NULL  
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
threshold	<p>A number between 0 and 1, inclusive. This number represents the lower value of TPR for the region where to calculate and plot lower bound.</p> <p>Because of definition of <code>fp_auc()</code>, region upper bound will be established as 1.</p>
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
.label	<p>A string representing the name used in labels.</p> <p>If NULL, variable name from predictor will be used as label.</p>

Value

A ggplot layer instance object.

Examples

```
# Add lower bound based on curve shape (Concave)
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_fpauc_lower_bound(
    data = iris,
    response = Species,
    predictor = Sepal.Width,
    threshold = 0.9
  )
```

`add_fpr_threshold_line`*Add a threshold line to a ROC plot*

Description

Include a threshold line on an specified axis.

Usage

```
add_fpr_threshold_line(threshold)
```

```
add_tpr_threshold_line(threshold)
```

```
add_threshold_line(threshold, ratio = NULL)
```

Arguments

- | | |
|-----------|---|
| threshold | <p>A number between 0 and 1, both inclusive, which represents the region bound where to calculate partial area under curve.</p> <p>If ratio = "tpr", it represents lower bound of the TPR region, being its upper limit equal to 1.</p> <p>If ratio = "fpr", it represents the upper bound of the FPR region, being its lower limit equal to 0.</p> |
| ratio | <p>Ratio in which to display threshold.</p> <ul style="list-style-type: none">• If "tpr" threshold will be displayed in TPR, y axis• If "fpr" it will be displayed in FPR, x axis. |

Value

A ggplot layer instance object.

Examples

```
# Add two threshold line in TPR = 0.9 and FPR = 0.1
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_threshold_line(threshold = 0.9, ratio = "tpr") +
  add_threshold_line(threshold = 0.1, ratio = "fpr")
# Add threshold line in TPR = 0.9
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_tpr_threshold_line(threshold = 0.9)
# Add threshold line in FPR = 0.1
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_fpr_threshold_line(threshold = 0.1)
```

add_partial_roc_curve *Add a section of a ROC curve to an existing one*

Description

Add an specific region of a ROC curve to an existing ROC plot.

Usage

```
add_partial_roc_curve(
  data,
  response = NULL,
  predictor = NULL,
  ratio,
  threshold,
  .condition = NULL,
  .label = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
ratio	<p>Ratio or axis where to apply calculations.</p> <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations. • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
threshold	<p>A number between 0 and 1, both inclusive, which represents the region bound where to calculate partial area under curve.</p> <p>If ratio = "tpr", it represents lower bound of the TPR region, being its upper limit equal to 1.</p> <p>If ratio = "fpr", it represents the upper bound of the FPR region, being its lower limit equal to 0.</p>

<code>.condition</code>	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See vignette("selecting-condition") for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
<code>.label</code>	<p>A string representing the name used in labels.</p> <p>If NULL, variable name from predictor will be used as label.</p>

Value

A ggplot layer instance object.

Examples

```
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_partial_roc_curve(
    iris,
    response = Species,
    predictor = Sepal.Length,
    ratio = "tpr",
    threshold = 0.9
  )
```

`add_partial_roc_points`

Add points in a section of a ROC curve to an existing plot

Description

Add points in a specific ROC region to an existing ROC plot.

Usage

```
add_partial_roc_points(
  data,
  response = NULL,
  predictor = NULL,
  ratio,
  threshold,
  .condition = NULL,
  .label = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
ratio	<p>Ratio or axis where to apply calculations.</p> <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations. • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
threshold	<p>A number between 0 and 1, both inclusive, which represents the region bound where to calculate partial area under curve.</p> <p>If ratio = "tpr", it represents lower bound of the TPR region, being its upper limit equal to 1.</p> <p>If ratio = "fpr", it represents the upper bound of the FPR region, being its lower limit equal to 0.</p>
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
.label	<p>A string representing the name used in labels.</p> <p>If NULL, variable name from predictor will be used as label.</p>

Value

A ggplot layer instance object.

Examples

```
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_partial_roc_points(
    iris,
    response = Species,
    predictor = Sepal.Length,
```



```

    ratio = "tpr",
    threshold = 0.9
  )

```

add_roc_curve

Add a ROC curve plot to an existing one

Description

Add a ROC curve to an existing ROC plot.

Usage

```

add_roc_curve(
  data,
  response = NULL,
  predictor = NULL,
  .condition = NULL,
  .label = NULL
)

```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
.label	<p>A string representing the name used in labels.</p> <p>If NULL, variable name from predictor will be used as label.</p>

Value

A ggplot layer instance object.

Examples

```
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_roc_curve(iris, response = Species, predictor = Sepal.Length)
```

add_roc_points	<i>Add ROC points plot to an existing one</i>
----------------	---

Description

Add ROC points to an existing ROC plot.

Usage

```
add_roc_points(
  data,
  response = NULL,
  predictor = NULL,
  .condition = NULL,
  .label = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See .condition for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See vignette("selecting-condition") for further information on how automatic selection is performed and details on selecting the condition of interest.</p>

.label A string representing the name used in labels.
 If NULL, variable name from predictor will be used as label.

Value

A ggplot layer instance object.

Examples

```
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_roc_points(iris, response = Species, predictor = Sepal.Length)
```

```
add_tpauc_concave_lower_bound
      Add TpAUC lower bound to a ROC plot
```

Description

Calculate and plot lower bound defined by TpAUC specificity index.

- add_tpauc_lower_bound() provides a upper level function which automatically calculates curve shape and plots a lower bound that better fits it.

Additionally, several lower level functions are provided to plot specific lower bounds:

- add_tpauc_concave_lower_bound(). Plot lower bound corresponding to a ROC curve with concave shape in selected region.
- add_tpauc_partially_proper_lower_bound. Plot lower bound corresponding to a ROC curve with partially proper (presence of some hook) in selected region.
- add_tpauc_under_chance_lower_bound. Plot lower bound corresponding to a ROC curve with a hook under chance line in selected region.

Usage

```
add_tpauc_concave_lower_bound(
  data,
  response = NULL,
  predictor = NULL,
  lower_threshold,
  upper_threshold,
  .condition = NULL,
  .label = NULL
)

add_tpauc_partially_proper_lower_bound(
  data,
  response = NULL,
  predictor = NULL,
```

```

    lower_threshold,
    upper_threshold,
    .condition = NULL,
    .label = NULL
  )

add_tpauc_under_chance_lower_bound(
  data,
  response = NULL,
  predictor = NULL,
  lower_threshold,
  upper_threshold,
  .condition = NULL,
  .label = NULL
)

add_tpauc_lower_bound(
  data,
  response = NULL,
  predictor = NULL,
  lower_threshold,
  upper_threshold,
  .condition = NULL,
  .label = NULL
)

```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See .condition for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
lower_threshold, upper_threshold	Two numbers between 0 and 1, inclusive. These numbers represent lower and upper values of FPR region where to calculate and plot lower bound.
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p>

	Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted. See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.
<code>.label</code>	A string representing the name used in labels. If NULL, variable name from predictor will be used as label.

Value

A ggplot layer instance object.

Examples

```
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +  
  add_tpauc_lower_bound(  
    data = iris,  
    response = Species,  
    predictor = Sepal.Width,  
    upper_threshold = 0.1,  
    lower_threshold = 0  
  )
```

auc	<i>Calculate area under ROC curve</i>
-----	---------------------------------------

Description

Calculates area under curve (AUC) of a predictor’s ROC curve.

Usage

```
auc(data = NULL, response, predictor, .condition = NULL)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>). If the variable presents more than two possible outcomes, classes or categories: <ul style="list-style-type: none">• The outcome of interest (the one to be predicted) will remain distinct.• All other categories will be combined into a single category. New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.

`.condition` A value from response that represents class, category or condition of interest which wants to be predicted.
 If NULL, condition of interest will be selected automatically depending on response type.
 Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.
 See `vignette("selecting-condition")` for further information on how automatic selection is performed and details on selecting the condition of interest.

Value

A numerical value representing the area under ROC curve.

Examples

```
# Calc AUC of Sepal.Width as a classifier of setosa species
auc(iris, Species, Sepal.Width)
# Change class to predict to virginica
auc(iris, Species, Sepal.Width, .condition = "virginica")
```

calc_curve_shape	<i>Calculate curve shape over an specific region</i>
------------------	--

Description

`calc_curve_shape()` calculates ROC curve shape over a specified region.

Usage

```
calc_curve_shape(
  data = NULL,
  response = NULL,
  predictor = NULL,
  lower_threshold,
  upper_threshold,
  ratio,
  .condition = NULL
)
```

Arguments

<code>data</code>	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
<code>response</code>	A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>). If the variable presents more than two possible outcomes, classes or categories: <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct.

- All other categories will be combined into a single category.

New combined category represents the "absence" of the condition to predict. See `.condition` for more information.

<code>predictor</code>	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
<code>lower_threshold, upper_threshold</code>	Two numbers between 0 and 1, inclusive. These numbers represent lower and upper bounds of the region where to apply calculations.
<code>ratio</code>	Ratio or axis where to apply calculations. <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations. • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
<code>.condition</code>	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>

Value

A string indicating ROC curve shape in the specified region. Result can take any of the following values:

- "Concave". ROC curve is concave over the entire specified region.
- "Partially proper". ROC curve loses concavity at some point of the specified region.
- "Hook under chance". ROC curve loses concavity at some point of the region and it lies below chance line.

Examples

```
# Calc ROC curve shape of Sepal.Width as a classifier of setosa species
# in TPR = (0.9, 1)
calc_curve_shape(iris, Species, Sepal.Width, 0.9, 1, "tpr")
# Change class to virginica
calc_curve_shape(iris, Species, Sepal.Width, 0.9, 1, "tpr", .condition = "virginica")
```

 calc_partial_roc_points

Calculate ROC curve partial points

Description

Calculates a series pairs of (FPR, TPR) which correspond to ROC curve points in a specified region.

Usage

```
calc_partial_roc_points(
  data = NULL,
  response = NULL,
  predictor = NULL,
  lower_threshold,
  upper_threshold,
  ratio,
  .condition = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
lower_threshold, upper_threshold	Two numbers between 0 and 1, inclusive. These numbers represent lower and upper bounds of the region where to apply calculations.
ratio	<p>Ratio or axis where to apply calculations.</p> <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations. • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
.condition	A value from response that represents class, category or condition of interest which wants to be predicted.

If NULL, condition of interest will be selected automatically depending on response type.

Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.

See vignette("selecting-condition") for further information on how automatic selection is performed and details on selecting the condition of interest.

Value

A tibble with two columns:

- "tpr". Containing "true positive ratio", or y, values of points within the specified region.
- "fpr". Containing "false positive ratio", or x, values of points within the specified region.

Examples

```
# Calc ROC points of Sepal.Width as a classifier of setosa species
# in TPR = (0.9, 1)
calc_partial_roc_points(
  iris,
  response = Species,
  predictor = Sepal.Width,
  lower_threshold = 0.9,
  upper_threshold = 1,
  ratio = "tpr"
)

# Change class to virginica
calc_partial_roc_points(
  iris,
  response = Species,
  predictor = Sepal.Width,
  lower_threshold = 0.9,
  upper_threshold = 1,
  ratio = "tpr",
  .condition = "virginica"
)
```

concordance_indexes *Concordance indexes*

Description

Concordance derived indexes allow calculation and explanation of area under ROC curve in a specific region. They use a dual perspective since they consider both TPR and FPR ranges which enclose the region of interest.

`cp_auc()` applies *concordant partial area under curve* (CpAUC), while `ncp_auc()` applies its normalized version by dividing by the total area.

Usage

```
cp_auc(
  data = NULL,
  response,
  predictor,
  lower_threshold,
  upper_threshold,
  ratio,
  .condition = NULL
)
```

```
nep_auc(
  data = NULL,
  response,
  predictor,
  lower_threshold,
  upper_threshold,
  ratio,
  .condition = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
lower_threshold, upper_threshold	Two numbers between 0 and 1, inclusive. These numbers represent lower and upper bounds of the region where to apply calculations.
ratio	<p>Ratio or axis where to apply calculations.</p> <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations. • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p>

Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.

See `vignette("selecting-condition")` for further information on how automatic selection is performed and details on selecting the condition of interest.

Value

A numeric value representing index score for the partial area under ROC curve.

References

Carrington, André M., et al. A new concordant partial AUC and partial c statistic for imbalanced data in the evaluation of machine learning algorithms. *BMC medical informatics and decision making* 20 (2020): 1-12.

Examples

```
# Calculate cp_auc of Sepal.Width as a classifier of setosa especies in
# FPR = (0, 0.1)
cp_auc(
  iris,
  response = Species,
  predictor = Sepal.Width,
  lower_threshold = 0,
  upper_threshold = 0.1,
  ratio = "fpr"
)
# Calculate ncp_auc of Sepal.Width as a classifier of setosa especies in
# FPR = (0, 0.1)
ncp_auc(
  iris,
  response = Species,
  predictor = Sepal.Width,
  lower_threshold = 0,
  upper_threshold = 0.1,
  ratio = "fpr"
)
```

hide_legend

Hide legend in a ROC plot

Description

Hide legend showing name of plotted classifiers and bounds in a ROC curve plot.

Usage

```
hide_legend()
```

Value

A ggplot theme object.

npauc_lower_bounds	<i>Add NpAUC lower bound to a ROC plot</i>
--------------------	--

Description

Calculate and plot lower bound defined by NpAUC specificity index.

- `add_npauc_normalized_lower_bound()` allows to plot normalized lower bound, which is used to formally calculate NpAUC.
- `add_npauc_lower_bound()` is a lower level function providing a way to plot lower bound previous to normalization.

Usage

```
add_npauc_lower_bound(
  data,
  response = NULL,
  predictor = NULL,
  threshold,
  .condition = NULL,
  .label = NULL
)

add_npauc_normalized_lower_bound(
  data,
  response = NULL,
  predictor = NULL,
  threshold,
  .condition = NULL,
  .label = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>

predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
threshold	A number between 0 and 1, inclusive. This number represents the lower value of TPR for the region where to calculate and plot lower bound. Because of definition of <code>np_auc()</code> , region upper bound will be established as 1.
.condition	A value from response that represents class, category or condition of interest which wants to be predicted. If NULL, condition of interest will be selected automatically depending on response type. Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted. See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.
.label	A string representing the name used in labels. If NULL, variable name from predictor will be used as label.

Value

A ggplot layer instance object.

Examples

```
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_npauc_lower_bound(
    iris,
    response = Species,
    predictor = Sepal.Width,
    threshold = 0.9
  )
```

pauc

Calculate partial area under curve

Description

Calculates area under curve curve in an specific TPR or FPR region.

Usage

```
pauc(
  data = NULL,
  response,
  predictor,
  ratio,
  lower_threshold,
  upper_threshold,
  .condition = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
ratio	<p>Ratio or axis where to apply calculations.</p> <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations. • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
lower_threshold, upper_threshold	Two numbers between 0 and 1, inclusive. These numbers represent lower and upper bounds of the region where to apply calculations.
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>

Value

A numeric value representing the area under ROC curve in the specified region.

Examples

```
# Calculate pauc of Sepal.Width as a classifier of setosa species in
# in TPR = (0.9, 1)
pauc(
  iris,
  response = Species,
  predictor = Sepal.Width,
  ratio = "tpr",
  lower_threshold = 0.9,
  upper_threshold = 1
)
# Calculate pauc of Sepal.Width as a classifier of setosa species in
```

```
# in FPR = (0, 0.1)
pauc(
  iris,
  response = Species,
  predictor = Sepal.Width,
  ratio = "fpr",
  lower_threshold = 0,
  upper_threshold = 0.1
)
```

plot_partial_roc_curve

Plot a section of a classifier ROC curve

Description

Create a curve plot using points in an specific region of ROC curve.

Usage

```
plot_partial_roc_curve(
  data,
  response = NULL,
  predictor = NULL,
  ratio,
  threshold,
  .condition = NULL,
  .label = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See .condition for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
ratio	<p>Ratio or axis where to apply calculations.</p> <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations.

	<ul style="list-style-type: none"> • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
threshold	<p>A number between 0 and 1, both inclusive, which represents the region bound where to calculate partial area under curve.</p> <p>If ratio = "tpr", it represents lower bound of the TPR region, being its upper limit equal to 1.</p> <p>If ratio = "fpr", it represents the upper bound of the FPR region, being its lower limit equal to 0.</p>
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See vignette("selecting-condition") for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
.label	<p>A string representing the name used in labels.</p> <p>If NULL, variable name from predictor will be used as label.</p>

Value

A ggplot object.

Examples

```
plot_partial_roc_curve(
  iris,
  response = Species,
  predictor = Sepal.Width,
  ratio = "tpr",
  threshold = 0.9
)
```

```
plot_partial_roc_points
```

Plot points in a region of a ROC curve

Description

Create an scatter plot using points in an specific region of ROC curve.

Usage

```
plot_partial_roc_points(
  data,
  response = NULL,
  predictor = NULL,
  ratio,
  threshold,
  .condition = NULL,
  .label = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See .condition for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
ratio	<p>Ratio or axis where to apply calculations.</p> <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations. • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
threshold	<p>A number between 0 and 1, both inclusive, which represents the region bound where to calculate partial area under curve.</p> <p>If ratio = "tpr", it represents lower bound of the TPR region, being its upper limit equal to 1.</p> <p>If ratio = "fpr", it represents the upper bound of the FPR region, being its lower limit equal to 0.</p>
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See vignette("selecting-condition") for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
.label	<p>A string representing the name used in labels.</p> <p>If NULL, variable name from predictor will be used as label.</p>

Value

A ggplot object.

Examples

```
plot_partial_roc_points(
  iris,
  response = Species,
  predictor = Sepal.Width,
  ratio = "tpr",
  threshold = 0.9
)
```

plot_roc_curve	<i>Plot a classifier ROC curve</i>
----------------	------------------------------------

Description

Create a curve plot using ROC curve points.

Usage

```
plot_roc_curve(
  data,
  response = NULL,
  predictor = NULL,
  .condition = NULL,
  .label = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See .condition for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.

.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See vignette("selecting-condition") for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
.label	<p>A string representing the name used in labels.</p> <p>If NULL, variable name from predictor will be used as label.</p>

Value

A ggplot object.

Examples

```
plot_roc_curve(iris, response = Species, predictor = Sepal.Width)
```

plot_roc_points	<i>Plot classifier points of a ROC curve</i>
-----------------	--

Description

Create an scatter plot using ROC curve points.

Usage

```
plot_roc_points(
  data,
  response = NULL,
  predictor = NULL,
  .condition = NULL,
  .label = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category.

	New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.
<code>predictor</code>	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
<code>.condition</code>	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
<code>.label</code>	<p>A string representing the name used in labels.</p> <p>If NULL, variable name from predictor will be used as label.</p>

Value

A ggplot object.

Examples

```
plot_roc_points(iris, response = Species, predictor = Sepal.Width)
```

<code>roc_points</code>	<i>Calculate ROC curve points</i>
-------------------------	-----------------------------------

Description

Calculates a series pairs of (FPR, TPR) which correspond to points displayed by ROC curve. "false positive ratio" will be represented on x axis, while "true positive ratio" on y one.

Usage

```
roc_points(data = NULL, response, predictor, .condition = NULL)
```

Arguments

<code>data</code>	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
<code>response</code>	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category.

	New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.
<code>predictor</code>	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
<code>.condition</code>	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>

Value

A tibble with two columns:

- `"tpr"`. Containing values for "true positive ratio", or y axis.
- `"fpr"`. Containing values for "false positive ratio", or x axis.

Examples

```
# Calc ROC points of Sepal.Width as a classifier of setosa species
roc_points(iris, Species, Sepal.Width)
# Change class to predict to virginica
roc_points(iris, Species, Sepal.Width, .condition = "virginica")
```

sensitivity_indexes	<i>Sensitivity indexes</i>
---------------------	----------------------------

Description

Sensitivity indexes provide different ways of calculating area under ROC curve in a specific TPR region. Two different approaches to calculate this area are available:

- `fp_auc()` applies *fitted partial area under curve* index (FpAUC). This one calculates area under curve adjusting to points defined by the curve in the selected region.
- `np_auc()` applies *normalized partial area under curve* index (NpAUC), which calculates area under curve over the whole specified region.

Usage

```
fp_auc(data = NULL, response, predictor, lower_tpr, .condition = NULL)

np_auc(data, response, predictor, lower_tpr, .condition = NULL)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
lower_tpr	<p>A numeric value between 0 and 1, inclusive, which represents lower value of TPR for the region where to calculate the partial area under curve.</p> <p>Because of definition of sensitivity indexes, upper bound of the region will be established as 1.</p>
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>

Value

A numeric value representing the index score for the partial area under ROC curve.

References

- Franco M. y Vivo J.-M. Evaluating the Performances of Biomarkers over a Restricted Domain of High Sensitivity. *Mathematics* 9, 2826 (2021).
- Jiang Y., Metz C. E. y Nishikawa R. M. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology* 201, 745-750 (1996).

Examples

```
# Calculate fp_auc of Sepal.Width as a classifier of setosa species
# in TPR = (0.9, 1)
fp_auc(iris, response = Species, predictor = Sepal.Width, lower_tpr = 0.9)
# Calculate np_auc of Sepal.Width as a classifier of setosa species
# in TPR = (0.9, 1)
np_auc(iris, response = Species, predictor = Sepal.Width, lower_tpr = 0.9)
```

spauc_lower_bounds	Add SpAUC lower bound to a ROC plot
--------------------	-------------------------------------

Description

Calculate and plot lower bound defined by SpAUC specificity index.

Usage

```
add_spauc_lower_bound(
  data,
  response = NULL,
  predictor = NULL,
  lower_threshold,
  upper_threshold,
  .condition = NULL,
  .label = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
lower_threshold, upper_threshold	Two numbers between 0 and 1, inclusive. These numbers represent lower and upper bounds of the region where to apply calculations.
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
.label	<p>A string representing the name used in labels.</p> <p>If NULL, variable name from predictor will be used as label.</p>

Details

SpAUC presents some limitations regarding its lower bound. Lower bound defined by this index cannot be applied to sections where ROC curve is defined under chance line.

`add_spauc_lower_bound()` doesn't make any check to ensure the index can be safely applied. Consequently, it allows to enforce the representation even though SpAUC could't be calculated in the region.

Value

A ggplot layer instance object.

Examples

```
plot_roc_curve(iris, response = Species, predictor = Sepal.Width) +
  add_spauc_lower_bound(
    iris,
    response = Species,
    predictor = Sepal.Width,
    lower_threshold = 0,
    upper_threshold = 0.1
  )
```

sp_auc

Specificity indexes

Description

Specificity indexes provide different ways of calculating area under ROC curve in a specific FPR region. Two different approaches to calculate this area are available:

- `tp_auc()` applies *tighter partial area under curve* index (SpAUC). This one calculates area under curve adjusting to points defined by the curve in the selected region.
- `sp_auc()` applies *standardized partial area under curve* index (TpAUC), which calculates area under curve over the whole specified region.

Usage

```
sp_auc(
  data = NULL,
  response,
  predictor,
  lower_fpr,
  upper_fpr,
  .condition = NULL,
  .invalid = FALSE
)
```



```

tp_auc(
  data = NULL,
  response,
  predictor,
  lower_fpr,
  upper_fpr,
  .condition = NULL
)

```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See .condition for more information.</p>
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
lower_fpr, upper_fpr	Two numbers between 0 and 1, inclusive. These numbers represent lower and upper values of FPR region where to calculate partial area under curve.
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See vignette("selecting-condition") for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
.invalid	If FALSE, the default, sp_auc() will return NA when ROC curve does not fit theoretical bounds and index cannot be applied. If TRUE, function will force the calculation and return a value despite probably being incorrect.

Value

A numeric value representing the index score for the partial area under ROC curve.

References

McClish D. K. Analyzing a Portion of the ROC Curve. *Medical Decision Making* 9, 190-195 (1989).

Vivo J.-M., Franco M. y Vicari D. Rethinking an ROC partial area index for evaluating the classification performance at a high specificity range. *Advances in Data Analysis and Classification* 12, 683-704 (2018).

Examples

```
# Calculate sp_auc of Sepal.Width as a classifier of setosa species
# in FPR = (0.9, 1)
sp_auc(
  iris,
  response = Species,
  predictor = Sepal.Width,
  lower_fpr = 0,
  upper_fpr = 0.1
)
# Calculate tp_auc of Sepal.Width as a classifier of setosa species
# in FPR = (0.9, 1)
tp_auc(
  iris,
  response = Species,
  predictor = Sepal.Width,
  lower_fpr = 0,
  upper_fpr = 0.1
)
```

sumexp_to_df

Transform data in a SummarizedExperiment to a data.frame

Description

Transforms a SummarizedExperiment into a data.frame which can be used as input for other functions.

Usage

```
sumexp_to_df(se, .n = NULL)
```

Arguments

se	A SummarizedExperiment object.
.n	An integer or string, representing the index or name of the assay to use. Same as <code>i</code> in <code>SummarizedExperiment::assay()</code> function. By default, function combines every assay in <code>se</code> argument.

Value

A data.frame created from combining assays and colData in a SummarizedExperiment.

summarize_dataset	<i>Summarize classifiers performance in a dataset</i>
-------------------	---

Description

Calculate a series of metrics describing global and local performance for selected classifiers in a dataset.

Usage

```
summarize_dataset(
  data,
  predictors = NULL,
  response,
  ratio,
  threshold,
  .condition = NULL,
  .progress = FALSE
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
predictors	<p>A vector of numeric data variables which represents the different classifiers or predictors in data to be summarized.</p> <p>If NULL and by default, predictors will match all numeric variables in data with the exception of response, given that it has a numeric type.</p>
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See .condition for more information.</p>
ratio	<p>Ratio or axis where to apply calculations.</p> <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations. • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
threshold	<p>A number between 0 and 1, both inclusive, which represents the region bound where to calculate partial area under curve.</p> <p>If ratio = "tpr", it represents lower bound of the TPR region, being its upper limit equal to 1.</p>

	If <code>ratio = "fpr"</code> , it represents the upper bound of the FPR region, being its lower limit equal to 0.
<code>.condition</code>	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>
<code>.progress</code>	If TRUE, show progress of calculations.

Value

A list with different elements:

- Performance metrics for each of evaluated classifiers.
- Overall description of performance metrics in the dataset.

Examples

```
summarize_dataset(iris, response = Species, ratio = "tpr", threshold = 0.9)
```

<code>summarize_predictor</code>	<i>Summarize classifier performance</i>
----------------------------------	---

Description

Calculates a series of metrics describing global and local classifier performance.

Usage

```
summarize_predictor(
  data = NULL,
  predictor,
  response,
  ratio,
  threshold,
  .condition = NULL
)
```

Arguments

data	A data.frame or extension (e.g. a tibble) containing values for predictors and response variables.
predictor	A data variable which must be numeric, representing values of a classifier or predictor for each observation.
response	<p>A data variable which must be a factor, integer or character vector representing the prediction outcome on each observation (<i>Gold Standard</i>).</p> <p>If the variable presents more than two possible outcomes, classes or categories:</p> <ul style="list-style-type: none"> • The outcome of interest (the one to be predicted) will remain distinct. • All other categories will be combined into a single category. <p>New combined category represents the "absence" of the condition to predict. See <code>.condition</code> for more information.</p>
ratio	<p>Ratio or axis where to apply calculations.</p> <ul style="list-style-type: none"> • If "tpr", only points within the specified region of TPR, y axis, will be considered for calculations. • If "fpr", only points within the specified region of FPR, x axis, will be considered for calculations.
threshold	<p>A number between 0 and 1, both inclusive, which represents the region bound where to calculate partial area under curve.</p> <p>If ratio = "tpr", it represents lower bound of the TPR region, being its upper limit equal to 1.</p> <p>If ratio = "fpr", it represents the upper bound of the FPR region, being its lower limit equal to 0.</p>
.condition	<p>A value from response that represents class, category or condition of interest which wants to be predicted.</p> <p>If NULL, condition of interest will be selected automatically depending on response type.</p> <p>Once the class of interest is selected, rest of them will be collapsed in a common category, representing the "absence" of the condition to be predicted.</p> <p>See <code>vignette("selecting-condition")</code> for further information on how automatic selection is performed and details on selecting the condition of interest.</p>

Value

A single row tibble with different predictor with following metrics as columns:

- Area under curve (AUC) as a metric of global performance.
- Partial are under curve (pAUC) as a metric of local performance.
- Indexes derived from pAUC, depending on the selected ratio. [Sensitivity indexes](#) will be used for TPR and [specificity indexes](#) for FPR.
- [Curve shape](#) in the specified region.

Examples

```
# Summarize Sepal.Width as a classifier of setosa species
# and local performance in TPR (0.9, 1)
summarize_predictor(
  data = iris,
  predictor = Sepal.Width,
  response = Species,
  ratio = "tpr",
  threshold = 0.9
)
# Summarize Sepal.Width as a classifier of setosa species
# and local performance in FPR (0, 0.1)
summarize_predictor(
  data = iris,
  predictor = Sepal.Width,
  response = Species,
  ratio = "fpr",
  threshold = 0.1
)
```

Index

add_chance_line, 2
 add_fpauc_concave_lower_bound
 (add_fpauc_partially_proper_lower_bound), 3
 add_fpauc_lower_bound
 (add_fpauc_partially_proper_lower_bound), 3
 add_fpauc_partially_proper_lower_bound, 3
 add_fpr_threshold_line, 5
 add_npauc_lower_bound
 (npauc_lower_bounds), 20
 add_npauc_normalized_lower_bound
 (npauc_lower_bounds), 20
 add_partial_roc_curve, 6
 add_partial_roc_points, 7
 add_roc_curve, 9
 add_roc_points, 10
 add_spauc_lower_bound
 (spauc_lower_bounds), 31
 add_threshold_line
 (add_fpr_threshold_line), 5
 add_tpauc_concave_lower_bound, 11
 add_tpauc_lower_bound
 (add_tpauc_concave_lower_bound), 11
 add_tpauc_partially_proper_lower_bound
 (add_tpauc_concave_lower_bound), 11
 add_tpauc_under_chance_lower_bound
 (add_tpauc_concave_lower_bound), 11
 add_tpr_threshold_line
 (add_fpr_threshold_line), 5
 auc, 13
 calc_curve_shape, 14
 calc_partial_roc_points, 16
 concordance_indexes, 17
 cp_auc (concordance_indexes), 17
 Curve shape, 37
 fp_auc (sensitivity_indexes), 29
 fp_auc(), 4
 fpauc_lower_bounds
 (add_fpauc_partially_proper_lower_bound), 3
 hide_legend, 19
 ncp_auc (concordance_indexes), 17
 np_auc (sensitivity_indexes), 29
 np_auc(), 21
 npauc_lower_bounds, 20
 pauc, 21
 plot_partial_roc_curve, 23
 plot_partial_roc_points, 24
 plot_roc_curve, 26
 plot_roc_points, 27
 plot_thresholds
 (add_fpr_threshold_line), 5
 roc_points, 28
 Sensitivity indexes, 37
 sensitivity_indexes, 29
 sp_auc, 32
 spauc_lower_bounds, 31
 specificity_indexes, 37
 specificity_indexes (sp_auc), 32
 sumexp_to_df, 34
 summarize_dataset, 35
 summarize_predictor, 36
 SummarizedExperiment::assay(), 34
 tp_auc (sp_auc), 32
 tpauc_lower_bounds
 (add_tpauc_concave_lower_bound), 11