Package ‘evabic’

June 12, 2020

Title Evaluation of Binary Classifiers

Version 0.0.3

Description Evaluates the performance of binary classifiers.
Computes confusion measures (TP, TN, FP, FN), derived measures (TPR, FDR, accuracy, F1, DOR, ..), and area under the curve. Outputs are well suited for nested dataframes.

License GPL-3


BugReports https://github.com/abichat/evabic/issues

Suggests testthat (>= 2.1.0)

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.1.0

NeedsCompilation no

Author Antoine Bichat [aut, cre] (<https://orcid.org/0000-0001-6599-7081>)

Maintainer Antoine Bichat <a.bichat@yahoo.fr>

Repository CRAN

Date/Publication 2020-06-12 15:00:03 UTC

R topics documented:

add_names ................................................................. 2
area_rect ................................................................. 2
ebc_allmeasures .......................................................... 3
ebc_AUC ................................................................. 5
ebc_confusion ............................................................. 6
ebc_tidy ................................................................. 7
ebc_tidy_by_threshold .................................................. 8
add_names

Add names to a vector

Description
Add names to a vector, with default names.

Usage
add_names(x, names = NULL, prefix = "x")

Arguments
- x: A vector.
- names: Vector of names to add. If NULL, default names are added.
- prefix: The prefix to add before default names. Useful only if names is set to NULL.

Value
A named vector

Examples
add_names(month.name)

area_rect

Area Under the Curve

Description
Compute area under the curve by using the rectangle’s method.

Usage
area_rect(x, y)

Arguments
- x: Numeric.
- y: Numeric.
Value

A numeric.

Examples

```r
x <- c(0, 0, 0.5, 1)
y <- c(0, 0.5, 0.5, 1)
plot(x, y, type = "b")
evabic::area_rect(x, y)
```

description

Available measures in evabic

Usage

```r
ebc_allmeasures
```

Format

An object of class character of length 18.

Details

<table>
<thead>
<tr>
<th>Detection</th>
<th>True condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected positive</td>
<td>Condition positive</td>
</tr>
<tr>
<td>Detected positive</td>
<td>Condition negative</td>
</tr>
<tr>
<td>Detected negative</td>
<td>Condition positive</td>
</tr>
<tr>
<td>Detected negative</td>
<td>Condition negative</td>
</tr>
</tbody>
</table>

**TP** True Positive
FP False Positive
FN False Negative
TN True Negative
TPR True Positive Rate or Sensitivity or Recall or Power
\[ TPR = \frac{TP}{TP + FN} = 1 - FNR \]
TNR True Negative Rate or Specificity
\[ TNR = \frac{TN}{FP + TN} = 1 - FPR \]
PPV Positive Predictive Value or Precision
\[ PPV = \frac{TP}{TP + FP} = 1 - FDR \]
NPV Negative Predictive Value
\[ NPV = \frac{TN}{TN + FN} = 1 - FOR \]
FNR False Negative Rate or Type II Error Rate or Miss Rate
\[ FNR = \frac{FN}{TP + FN} = 1 - TPR \]
FPR False Positive Rate or Type I Errors Rate or Fall-out
\[ FPR = \frac{FP}{FP + TN} = 1 - TNR \]
FDR False Discovery Rate
\[ FDR = \frac{FP}{FP + TP} = 1 - PPV \]
FOR False Omission Rate
\[ FOR = \frac{FN}{TN + FN} = 1 - NPV \]
ACC Accuracy
\[ ACC = \frac{TP + TN}{TP + FP + FN + TN} \]
BACC Balanced Accuracy
\[ BACC = \frac{\frac{TP}{TP + FN} + \frac{TN}{FP + TN}}{2} \]
F1 F1 Score
\[ F1 = \frac{2TP}{2TP + FP + FN} = \frac{2}{\frac{TPR}{TPR} + \frac{PPV}{PPV}} \]
PLR Positive Likelihood Ratio or LR+ or Likelihood Ratio for Positive Results
\[ PLR = \frac{TPR}{1 - TNR} \]
**NLR**  Negative Likelihood Ratio or LR- or Likelihood Ratio for Negative Results

\[ NLR = \frac{1 - TPR}{TNR} \]

**DOR**  Diagnostic Odds Ratio

\[ DOR = \frac{TP}{FP} \times \frac{TN}{FN} = \frac{PLR}{NLR} \]

**References**


**Examples**

```r
ebc_allmeasures
```

---

**Description**

Compute the Area Under the Curve for a classification.

**Usage**

```r
ebc_AUC(  
  detection_values,
  true,
  all,
  m = length(all),
  direction = c("<", ">", "<=", ">=")
)
ebc_AUC_from_measures(df_measures)
```

**Arguments**

- `detection_values`  Values corresponding to elements that are detected. Must be named.
- `true`  Vector of element that are supposed to be detected.
- `all`  Vector of all elements.
- `m`  Total number of elements.
- `direction`  With < (default), detected elements are those which are strictly less than the threshold. Could be change to ">", "=" or ">=".
- `df_measures`  A dataframe with TPR and FPR columns. E.g. the output of `ebc_tidy_by_threshold`. 
Value

A numeric.

Examples

```
set.seed(42)
X1 <- rnorm(50)
X2 <- rnorm(50)
X3 <- rnorm(50)
predictors <- paste0("X", 1:3)
df_lm <- data.frame(X1 = X1, X2 = X2, X3 = X3,
                     X4 = X1 + X2 + X3 + rnorm(50, sd = 0.5),
                     X5 = X1 + 3 * X3 + rnorm(50, sd = 0.5),
                     X6 = X2 - 2 * X3 + rnorm(50, sd = 0.5),
                     X7 = X1 - X2 + rnorm(50, sd = 2),
                     Y = X1 - X2 + 3 * X3 + rnorm(50))
model <- lm(Y ~ ., data = df_lm)
pvalues <- summary(model)$coefficients[-1, 4]
ebc_AUC(pvalues, predictors, m = 7)
```

```
df_measures <- ebc_tidy_by_threshold(pvalues, predictors, m = 7)
ebc_AUC_from_measures(df_measures)
```

ebc_confusion

**Confusion matrix**

Description

Compute the the confusion matrix

Usage

```
ebc_confusion(detected, true, all, m = length(all), prop = FALSE)
```

Arguments

- `detected` Vector of elements that are detected.
- `true` Vector of element that are supposed to be detected.
- `all` Vector of all elements.
- `m` Total number of elements.
- `prop` Logical, default to FALSE. Should the matrix sum to one?

Details

See `ebc_allmeasures` for the description of the measures.
Value
A 2*2 named matrix.

Examples

```r
ebc_confusion(detected = c("A", "C", "D"), true = c("A", "B", "C"), m = 6)
```

Description
Construct a single row summary of the classifier.

Usage

```r
ebc_tidy(
  detected,
  true,
  all,
  m = length(all),
  measures = c("TPR", "FPR", "FDR", "ACC", "F1")
)
```

Arguments

- `detected`: Vector of elements that are detected.
- `true`: Vector of element that are supposed to be detected.
- `all`: Vector of all elements.
- `m`: Total number of elements.
- `measures`: Desired measures of performance.

Details
See `ebc_allmeasures` for the available measures and their descriptions.

Value
A single-row data.frame with one column per element in measures.

See Also
`ebc_TP`, `ebc_TPR`, `ebc_allmeasures`

Examples

```r
ebc_tidy(detected = c("A", "C", "D"), true = c("A", "B", "C"),
          all = LETTERS[1:6], measures = c("ACC", "FDR"))
```
**Description**

Computes measures according to a moving threshold.

**Usage**

\[
\text{ebc\_tidy\_by\_threshold}(\text{detection\_values}, \text{true}, \text{all}, m = \text{length(}\text{all}), \text{measures} = \text{c("TPR", "FPR", "FDR", "ACC", "F1")}, \text{direction} = \text{c("<", ",","<="", ",">")})
\]

**Arguments**

- **detection\_values**
  - Values corresponding to elements that are detected. Must be named.
- **true**
  - Vector of element that are supposed to be detected.
- **all**
  - Vector of all elements.
- **m**
  - Total number of elements.
- **measures**
  - Desired measures of performance.
- **direction**
  - With < (default), detected elements are those which are strictly less than the threshold. Could be change to ">", "<=" or ">=".

**Details**

See \texttt{ebc\_allmeasures} for the available measures and their descriptions.

**Value**

A dataframe with one column called threshold and other corresponding to those specified in \texttt{measures}.

**Examples**

```r
set.seed(42)
X1 <- rnorm(50)
X2 <- rnorm(50)
X3 <- rnorm(50)
predictors <- paste0("X", 1:3)
df\_lm <- data\_frame(X1 = X1, X2 = X2, X3 = X3,
  X4 = X1 + X2 + X3 + rnorm(50, sd = 0.5),
  X5 = X1 + 3 \times X3 + rnorm(50, sd = 0.5),
)```

X6 = X2 - 2 * X3 + rnorm(50, sd = 0.5),
X7 = X1 - X2 + rnorm(50, sd = 2),
Y = X1 - X2 + 3 * X3 + rnorm(50))

model <- lm(Y ~ ., data = df_lm)
pvalues <- summary(model)$coefficients[-1, 4]
ebc_tidy_by_threshold(pvalues, predictors, m = 7)

---

**Description**

Basic measures from the confusion matrix.

**Usage**

- `ebc_TP(detected, true)`
- `ebc_FP(detected, true)`
- `ebc_FN(detected, true)`
- `ebc_TN(detected, true, all, m = length(all))`

**Arguments**

- `detected` Vector of elements that are detected.
- `true` Vector of element that are supposed to be detected.
- `all` Vector of all elements.
- `m` Total number of elements.

**Details**

See `ebc_allmeasures` for the description of the measures.

**Value**

An integer.

**See Also**

`ebc_TPR, ebc_tidy, ebc_allmeasures`
Examples

```
ebc_TP(detected = c("A", "C", "D"), true = c("A", "B", "C"))
ebc_FP(detected = c("A", "C", "D"), true = c("A", "B", "C"))
ebc_FN(detected = c("A", "C", "D"), true = c("A", "B", "C"),
        all = LETTERS[1:6])
ebc_TN(detected = c("A", "C", "D"), true = c("A", "B", "C"),
        m = 6)
```

---

ebc_TPR          Derived measures.

Description

Measures derived from confusion matrix.

Usage

```
ebc_TPR(detected, true)
ebc_TNR(detected, true, all, m = length(all))
ebc_PPV(detected, true)
ebc_NPV(detected, true, all, m = length(all))
ebc_FNR(detected, true)
ebc_FPR(detected, true, all, m = length(all))
ebc_FDR(detected, true)
ebc_FOR(detected, true, all, m = length(all))
ebc_ACC(detected, true, all, m = length(all))
ebc_BACC(detected, true, all, m = length(all))
ebc_F1(detected, true)
ebc_PLR(detected, true, all, m = length(all))
ebc_NLR(detected, true, all, m = length(all))
ebc_DOR(detected, true, all, m = length(all))
```
**Arguments**

- **detected**: Vector of elements that are detected.
- **true**: Vector of elements that are supposed to be detected.
- **all**: Vector of all elements.
- **m**: Total number of elements.

**Details**

See [ebc_allmeasures](#) for the description of the measures.

**Value**

A numeric.

**See Also**

[ebc_TP](#), [ebc_tidy](#), [ebc_allmeasures](#)

**Examples**

```r
ebc_TPR(detected = c("A", "C", "D"), true = c("A", "B", "C"))
ebc_ACC(detected = c("A", "C", "D"), true = c("A", "B", "C"),
          all = LETTERS[1:5])
```

---

**nl2c**

*Named logical to character*

**Description**

Extracts names of `x` where `x` is TRUE

**Usage**

`nl2c(x)`

**Arguments**

- **x**: vector

**Value**

A vector of the same size

**Examples**

```r
x <- c(a = TRUE, b = FALSE, c = FALSE, d = TRUE)
evabic:::nl2c(x)
```
Description
Repeat 0

Usage
rep_0(n)

Arguments
n The number of time to repeat

Value
A character.
Index

*Topic **datasets**
   ebc_allmeasures, 3

add_names, 2
area_rect, 2

ebc_ACC (ebc_TPR), 10
ebc_allmeasures, 3, 6–9, 11
ebc_AUC, 5
ebc_AUC_from_measures (ebc_AUC), 5
ebc_BACC (ebc_TPR), 10
ebc_confusion, 6
ebc_DOR (ebc_TPR), 10
ebc_F1 (ebc_TPR), 10
ebc_FDR (ebc_TPR), 10
ebc_FN (ebc_TP), 9
ebc_FNR (ebc_TPR), 10
ebc_FOR (ebc_TPR), 10
ebc_FP (ebc_TP), 9
ebc_FPR (ebc_TPR), 10
ebc_NLR (ebc_TPR), 10
ebc_NPV (ebc_TPR), 10
ebc_PLR (ebc_TPR), 10
ebc_PPV (ebc_TPR), 10
ebc_tidy, 7, 9, 11
ebc_tidy_by_threshold, 5, 8
ebc_TN (ebc_TP), 9
ebc_TNR (ebc_TPR), 10
ebc_TP, 7, 9, 11
ebc_TPR, 7, 9, 10

nl2c, 11

rep_0, 12