Package 'mltest'

May 9, 2026

Title Classification Evaluation Metrics

Version 1.0.3

Description A fast, robust and easy-to-use calculation
of multi-class classification evaluation metrics based on confusion matrix.

License GPL-2

Encoding UTF-8

RoxygenNote 7.2.3

NeedsCompilation no

Maintainer G. Dudnik <gl.dudnik@gmail.com>

Repository CRAN

Date/Publication 2025-02-03 09:30:08 UTC

Author G. Dudnik [aut, cre, cph] (ORCID:
<https://orcid.org/0000-0002-3418-1589>)

Contents

ml_test 1

Index **5**

`ml_test` *multi-class classifier evaluation metrics based on a confusion matrix (contingency table)*

Description

Calculates multi-class classification evaluation metrics: balanced accuracy (**balanced.accuracy**), diagnostic odds ratio (**DOR**), error rate (**error.rate**), F.beta (**F0.5**, **F1** (F-measure, F-score), **F2** with where beta is 0.5, 1 and 2 respectively), false positive rate (**FPR**), false negative rate (**FNR**), false omission rate (**FOR**), false discovery rate (**FDR**), geometric mean (**geometric.mean**), **Jaccard**, positive likelihood ratio (p+, LR(+) or simply **L**), negative likelihood ratio (p-, LR(-) or simply

lambda), Matthews corellation coefficient (**MCC**), markedness (**MK**), negative predictive value (**NPV**), optimization precision **OP**, **precision**, **recall** (sensitivity), **specificity** and finally **Youden**'s index. The function calculates the aforementioned metrics from a confusion matrix (contingency matrix) where *TP*, *TN*, *FP FN* are abbreviations for *true positives*, *true negatives*, *false positives* and *false negatives* respectively.

Usage

```
ml_test(predicted, true, output.as.table = FALSE)
```

Arguments

`predicted`	class labels predicted by the classifier model (a set of classes convertible into type factor with levels representing labels)
`true`	true class labels (a set of classes convertible into type factor of the same length and with the same levels as predicted)
`output.as.table`	the function returns all metrics except for **accuracy** and **error.rate** in a tabular format if this argument is set to *TRUE*

Value

the function returns a list of following metrics:

`accuracy`	= (TP+TN) / (TP+FP+TN+FN) *(doesn't show up when output.as.table = TRUE)*
`balanced.accuracy`	= (TP / (TP+FN)+TN / (TN+FP)) / 2 = (recall+specificity) / 2
`DOR`	= TP*TN / (FP*FN) = L / lambda
`error.rate`	= (FP+FN) / (TP+TN+FP+FN) = 1-accuracy *(doesn't show up when output.as.table = TRUE)*
`F0.5`	= 1.25*recall*precision/(0.25*precision+recall)
`F1`	= 2*recall*precision / (precision+recall)
`F2`	= 5*recall*precision / (4*precision+recall)
`FDR`	= 1-precision
`FNR`	= 1-recall
`FOR`	= 1-NPV
`FPR`	= 1-specificity
`geometric.mean`	= (recall*specificity)^0.5
`Jaccard`	= TP / (TP+FP+FN)
`L`	= recall / (1-specificity)
`lambda`	= (1-recall) / (specificity)
`MCC`	= (TP*TN-FP*FN) / (((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))^0.5)
`MK`	= precision + NPV - 1
`NPV`	= TN / (TN+FN)

OP	= accuracy - |recall-specificity| / (recall+specificity)
precision	= TP / (TP+FP)
recall	= TP / (TP+FN)
specificity	= TN / (TN+FP)
Youden	= recall+specificity-1

Author(s)

G. Dudnik

References

1. Sasaki Y. (2007). The truth of the F-measure.:1–5. https://www.researchgate.net/publication/268185911_The_truth_of_the_F-measure.
2. Powers DMW. (2011). Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness & Correlation. Arch Geschwulstforsch. 2(1):37–63. doi: 10.48550/arXiv.2010.16061.
3. Bekkar M, Djemaa HK, Alitouche TA. (2013). Evaluation Measures for Models Assessment over Imbalanced Data Sets. J Inf Eng Appl. 3(10):27–38. https://www.researchgate.net/publication/292718336_Evaluation_measures_for_models_assessment_over_imbalanced_data_sets.
4. Jeni LA, Cohn JF, De La Torre F. (2013). Facing Imbalanced Data Recommendations for the Use of Performance Metrics. Conference on Affective Computing and Intelligent Interaction. IEEE. p. 245–51. doi: 10.1109/ACII.2013.47.
5. López V, Fernández A, García S, Palade V, Herrera F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Inf Sci. 250:113–41. doi: 10.1016/j.ins.2013.07.007.
6. Tharwat A. (2018). Classification assessment methods. Appl Comput Informatics . doi: 10.1016/j.aci.2018.08.003.

Examples

```
library(mltest)

# class labels ("cat, "dog" and "rat") predicted by the classifier model
predicted_labels <- as.factor(c("dog", "cat", "dog", "rat", "rat"))

# true labels (test set)
true_labels <- as.factor(c("dog", "cat", "dog", "rat", "dog"))

classifier_metrics <- ml_test(predicted_labels, true_labels, output.as.table = FALSE)

# overall classification accuracy
accuracy <- classifier_metrics$accuracy

# F1-measures for classes "cat", "dog" and "rat"
F1 <- classifier_metrics$F1

# tabular view of the metrics (except for 'accuracy' and 'error.rate')
```

```
classifier_metrics <- ml_test(predicted_labels, true_labels, output.as.table = TRUE)
```

Index

∗ **utilities**
ml_test, 1

ml_test, 1