Package 'correctR'

July 22, 2025

Type Package

Title Corrected Test Statistics for Comparing Machine Learning Models on Correlated Samples

Version 0.3.1

Date 2025-02-05

Maintainer Trent Henderson <then6675@uni.sydney.edu.au>

Description Calculate a set of corrected test statistics for cases when samples are not independent, such as when classification accuracy values are obtained over resamples or through k-fold cross-validation, as proposed by Nadeau and Bengio (2003) <doi:10.1023/A:1024068626366> and presented in Bouckaert and Frank (2004) <doi:10.1007/978-3-540-24775-3_3>.

BugReports https://github.com/hendersontrent/correctR/issues

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 3.5.0)

Imports stats

Suggests knitr, markdown, rmarkdown, pkgdown, testthat (>= 3.0.0)

RoxygenNote 7.2.3

VignetteBuilder knitr

Config/testthat/edition 3

URL https://hendersontrent.github.io/correctR/

NeedsCompilation no

Author Trent Henderson [cre, aut]

Repository CRAN

Date/Publication 2025-02-05 10:50:06 UTC

Contents

---

correctR	*Corrections For Correlated Test Statistics*

Description

Corrections For Correlated Test Statistics

---

kfold_ttest	*Compute correlated t-statistic and p-value for k-fold cross-validated results*

Description

Compute correlated t-statistic and p-value for k-fold cross-validated results

Usage

```
kfold_ttest(x, y, n, k, tailed = c("two", "one"), greater = NULL)
```

Arguments

x	numeric vector of values for model A
y	numeric vector of values for model B
n	integer denoting total sample size
k	integer denoting number of folds used in k-fold
tailed	character denoting whether to perform a two-tailed or one-tailed test. Can be one of `"two"` or `"one"`. Defaults to `"two"`
greater	character specifying whether `"x"` or `"y"` is greater for the one-tailed test if `tailed = "one"`. Defaults to `NULL`

Value

`data.frame` containing the test statistic and p-value

Author(s)

Trent Henderson

References

Nadeau, C., and Bengio, Y. Inference for the Generalization Error. Machine Learning 52, (2003).

Corani, G., Benavoli, A., Demsar, J., Mangili, F., and Zaffalon, M. Statistical comparison of classi-fiers through Bayesian hierarchical modelling. Machine Learning, 106, (2017).

Examples

```
x <- rnorm(100, mean = 95, sd = 0.5)
y <- rnorm(100, mean = 90, sd = 1)
kfold_ttest(x = x, y = y, n = 100, k = 5, tailed = "two")
```

---

repkfold_ttest	*Compute correlated t-statistic and p-value for repeated k-fold cross-validated results*

Description

Compute correlated t-statistic and p-value for repeated k-fold cross-validated results

Usage

```
repkfold_ttest(data, n1, n2, k, r, tailed = c("two", "one"), greater = NULL)
```

Arguments

data	data.frame of values for model A and model B over repeated k-fold cross-validation. Four named columns are expected: "model", "values", "k", and "k"
n1	integer denoting train set size
n2	integer denoting test set size
k	integer denoting number of folds used in k-fold
r	integer denoting number of repeats per fold
tailed	character denoting whether to perform a two-tailed or one-tailed test. Can be one of "two" or "one". Defaults to "two"
greater	value specifying which value in the "model" column is greater for the one-tailed test if tailed = "one". Defaults to NULL

Value

data.frame containing the test statistic and p-value

Author(s)

Trent Henderson

References

Nadeau, C., and Bengio, Y. Inference for the Generalization Error. Machine Learning 52, (2003).

Bouckaert, R. R., and Frank, E. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. Advances in Knowledge Discovery and Data Mining. PAKDD 2004. Lecture Notes in Computer Science, 3056, (2004).

Examples

```
tmp <- data.frame(model = rep(c(1, 2), each = 60),
  values = c(stats::rnorm(60, mean = 0.6, sd = 0.1),
  stats::rnorm(60, mean = 0.4, sd = 0.1)),
  k = rep(c(1, 1, 2, 2), times = 15),
  r = rep(c(1, 2), times = 30))

repkfold_ttest(data = tmp, n1 = 80, n2 = 20, k = 2, r = 2, tailed = "two")
```

---

resampled_ttest	*Compute correlated t-statistic and p-value for resampled data*

Description

Compute correlated t-statistic and p-value for resampled data

Usage

```
resampled_ttest(x, y, n, n1, n2, tailed = c("two", "one"), greater = NULL)
```

Arguments

x	numeric vector of values for model A
y	numeric vector of values for model B
n	integer denoting number of repeat samples. Defaults to length(x)
n1	integer denoting train set size
n2	integer denoting test set size
tailed	character denoting whether to perform a two-tailed or one-tailed test. Can be one of "two" or "one". Defaults to "two"
greater	character specifying whether "x" or "y" is greater for the one-tailed test if tailed = "one". Defaults to NULL

Value

data.frame containing the test statistic and p-value

Author(s)

Trent Henderson

References

Nadeau, C., and Bengio, Y. Inference for the Generalization Error. Machine Learning 52, (2003).

Bouckaert, R. R., and Frank, E. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. Advances in Knowledge Discovery and Data Mining. PAKDD 2004. Lecture Notes in Computer Science, 3056, (2004).

Examples

```
x <- rnorm(100, mean = 95, sd = 0.5)
y <- rnorm(100, mean = 90, sd = 1)
resampled_ttest(x = x, y = y, n = 100, n1 = 80, n2 = 20, tailed = "two")
```

Index