Package ‘correctR’

August 20, 2023

Type Package

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

Version 0.1.3

Date 2023-08-20

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 7.2.2

VignetteBuilder knitr

Config/testthat/edition 3

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

NeedsCompilation no

Author Trent Henderson [cre, aut]

Repository CRAN

Date/Publication 2023-08-20 13:02:33 UTC
R topics documented:

- `correctR` ................................................................. 2
- `kfold_ttest` ............................................................ 2
- `repkfold_ttest` ................................................................ 3
- `resampled_ttest` .......................................................... 4

Index 5

<table>
<thead>
<tr>
<th>correctR</th>
<th>Corrections For Correlated Test Statistics</th>
</tr>
</thead>
</table>

**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)`

**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

**Value**

object of class `data.frame`

**Author(s)**

Trent Henderson

**References**

**Description**

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

**Usage**

```r
repkfold_ttest(data, n1, n2, k, r)
```

**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 "r"
- `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

**Value**

object of class `data.frame`

**Author(s)**

Trent Henderson

**References**


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)

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

Value
object of class data.frame

Author(s)
Trent Henderson

References
Index

correctR, 2
correctR-package (correctR), 2

kfold_ttest, 2
repkold_ttest, 3
resampled_ttest, 4