

Package ‘hypoRF’

May 7, 2021

Type Package

Title Random Forest Two-Sample Tests

Version 1.0.0

Author Hediger Simon, Loris Michel, Jeffrey Naef

Maintainer Hediger Simon <simon.hediger@uzh.ch>

Description An implementation of Random Forest-based two-sample tests as introduced in Hediger & Michel & Naef (2020) <arXiv:1903.06287>.

License GPL-3

Imports stats, ranger

Encoding UTF-8

Suggests testthat

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2021-05-07 09:50:02 UTC

R topics documented:

hypoRF	1
Index	4

hypoRF	<i>HypoRF; a Random Forest based Two Sample Test</i>
--------	--

Description

Performs a permutation two sample test based on the out-of-bag-error of random forest

Usage

```

hypoRF(
  data1,
  data2,
  K = 100,
  statistic = "PerClassOOB",
  normalapprox = F,
  seed = NULL,
  alpha = 0.05,
  ...
)

```

Arguments

data1	An object of type "data.frame". The first sample.
data2	An object of type "data.frame". The second sample.
K	A numeric value specifying the number of times the created label is permuted. For K = 1 a binomial test is carried out. The Default is K = 100.
statistic	A character value specifying the statistic for permutation testing. Two options available <ul style="list-style-type: none"> • PerClassOOB Sum of OOB per class errors. • OverallOOB OOB-error. . Default is statistic = "PerClassOOB".
normalapprox	A logical value asking for the use of a normal approximation. Default is normalapprox = FALSE.
seed	A numeric value for reproducibility.
alpha	The level of the test. Default is alpha = 0.05.
...	Arguments to be passed to ranger

Value

A list with elements

- pvalue: The p-value of the test.
- obs: The OOB-statistic in case of K>1 or the out-of-sample error in case of K=1 (binomial test).
- val: The OOB-statistic of the permuted random forests in case of K>1 (otherwise NULL).
- varest: The estimated variance of the permuted random forest OOB-statistic in case of K>1 (otherwise NULL).
- statistic: The used OOB-statistic
- importance_ranking: The variable importance measure, when importance == "impurity".
- cutoff: The quantile of the importance distribution at level alpha.
- call: Call to the function.

See Also

[ranger](#)

Examples

```
# Using the default testing procedure (permutation test)
x1 <- data.frame(x=stats::rt(100, df=1.5))
x2 <- data.frame(x=stats::rnorm(100))
hypoRF(x1, x2, K=2)
# Using the exact binomial test
hypoRF(x1, x2, K=1)
```

Index

[hypoRF](#), [1](#)

[ranger](#), [3](#)