Package 'corrRF'

March 20, 2025

Type Package

Title Clustered Random Forests for Optimal Prediction and Inference of Clustered Data

Version 1.1.0

Maintainer Elliot H. Young <ey244@cam.ac.uk>

Description A clustered random forest algorithm for fitting random forests for data of independent clusters, that exhibit within cluster dependence.
Details of the method can be found in Young and Buehlmann (2025) <doi:10.48550/arXiv.2503.12634>.

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.3

LinkingTo Rcpp

Imports Rcpp, rpart

Depends R (>= 4.2.0)

Suggests knitr, rmarkdown, testthat

NeedsCompilation yes

Author Elliot H. Young [aut, cre]

Repository CRAN

Date/Publication 2025-03-20 09:20:06 UTC

Contents

crf *Clustered random forest fitting*

Description

Clustered random forest fitting

Usage

```
crf(
  formula,
  data,
  B = 500,
  L = 100,
  beta = 0.9,
  weight_optimiser = "Training MSE",
  correlation = "equicorr",
  maxdepth = 30,
  minbucket = 10,
  cp = 0,
  x0 = NULL,
  test_data = NULL,
  fixrho = FALSE,
  honesty = TRUE,
  verbose = TRUE,
  seed = NULL
)
```

Arguments

formula	an object of class 'formula' describing the model to fit.
data	training dataset for fitting the CRF. Note that group ID must be given by the column id.
B	the total number of trees (or trees per little bag if $L \neq$'NULL'). Default is 500.
L	the total number of little bags if providing a bootstrap of little bags estimate for inference. To not include set $L =$'NULL'. Default is 'NULL'.
beta	the subsampling rate. Default is $beta = 0.9$.
weight_optimiser	the method used to construct weights. Options are 'Pointwise variance', 'Training MSE' or 'Test MSE'. Default is 'Training MSE'.
correlation	the weight structure implemented. Currently supported options are 'ar1' and 'equicorr'. Default is 'equicorr'.
maxdepth	the maximum depth of the decision tree fitting. Default is 30.
minbucket	the minbucket of the decision tree fitting. Default is 10.

cp	the complexity paramter for decision tree fitting. Default is 0.
x0	the covariate point to optimise weights towards if 'weightoptimiser' set to 'Pointwise variance'.
test_data	the test dataset to optimise weights towards if 'weightoptimiser' set to 'Test MSE'.
fixrho	fixes a pre-specified weight structure, given by the relevant 'ar1' or 'equicorr' parameter. Default is 'FALSE' (optimise weights).
honesty	whether honest or dishonest trees to be fit. Default is 'TRUE'.
verbose	Logical indicating whether or not to print computational progress. Default is 'TRUE'.
seed	Random seed for sampling. Default is NULL.

Value

A clustered random forest fitted object

`predict.crf` *Predictions from a crf given newdata*

Description

Predictions from a fitted `crf` clustered random forest on newdata `newdata`.

Usage

```
## S3 method for class 'crf'
predict(object, newdata, sderr = FALSE, ...)
```

Arguments

object	a fitted `crf` clustered random forest object fitted by `crf`.
newdata	dataset on which predictions are to be performed.
sderr	whether 'bootstrap of little bags' standard errors should be additionally outputted. Default is `FALSE`.
...	additional arguments

Value

Fitted values, potentially alongside standard errors (see `sderr`).

summary.crf *Summary for a crf fitted object*

Description

Summary of a fitted `crf` clustered random forest object fitted by `crf`.

Usage

```
## S3 method for class 'crf'
summary(object, ...)
```

Arguments

`object`	a fitted `crf` clustered random forest object fitted by `crf`.
`...`	additional arguments

Value

Prints summary output for `crf` object

Index