

Package ‘Boruta’

May 16, 2026

Title Wrapper Algorithm for All Relevant Feature Selection

Version 10.0.0

Imports fru

Suggests rFems, randomForest, ranger, survival

Description An all relevant feature selection wrapper algorithm.

It finds relevant features by comparing original attributes' importance with importance achievable at random, estimated using their permuted copies (shadows).

License GPL (>= 2)

URL <https://gitlab.com/mbq/Boruta/>

RoxygenNote 7.3.3

Encoding UTF-8

NeedsCompilation no

Author Miron Bartosz Kursa [aut, cre] (ORCID:
<<https://orcid.org/0000-0001-7672-648X>>),
Witold Remigiusz Rudnicki [aut]

Maintainer Miron Bartosz Kursa <m@mbq.me>

Repository CRAN

Date/Publication 2026-05-16 08:10:16 UTC

Contents

attStats	2
Boruta	3
conditionalTransdapter	5
decohereTransdapter	5
getConfirmedFormula	6
getImpFems	6
getImpFru	7
getImpLegacyRf	8
getImpRf	9
getSelectedAttributes	10

imputeTransdapter	10
plot.Boruta	11
plotImpHistory	12
print.Boruta	13
srx	13
TentativeRoughFix	14
Index	15

attStats	<i>Extract attribute statistics</i>
----------	-------------------------------------

Description

attStats shows a summary of a Boruta run in an attribute-centred way. It produces a data frame containing some importance stats as well as the number of hits that attribute scored and the decision it was given.

Usage

```
attStats(x)
```

Arguments

x an object of a class Boruta, from which attribute stats should be extracted.

Value

A data frame containing, for each attribute that was originally in information system, mean, median, maximal and minimal importance, number of hits normalised to number of importance source runs performed and the decision copied from finalDecision.

Note

When using a Boruta object generated by a [TentativeRoughFix](#), the resulting data frame will consist a rough-fixed decision.

x has to be made with holdHistory set to TRUE for this code to run.

Description

Boruta is an all relevant feature selection wrapper algorithm, capable of working with any classification method that output variable importance measure (VIM); by default, Boruta uses Random Forest. The method performs a top-down search for relevant features by comparing original attributes' importance with importance achievable at random, estimated using their permuted copies, and progressively eliminating irrelevant features to stabilise that test.

Usage

```
Boruta(x, ...)  
  
## Default S3 method:  
Boruta(  
  x,  
  y,  
  pValue = 0.01,  
  mcAdj = TRUE,  
  maxRuns = 100,  
  doTrace = 0,  
  holdHistory = TRUE,  
  getImp = getImpFruZ,  
  ...  
)  
  
## S3 method for class 'formula'  
Boruta(formula, data, ...)
```

Arguments

x	data frame of predictors.
...	additional parameters passed to getImp.
y	response vector; factor for classification, numeric vector for regression, Surv object for survival (support depends on importance adapter capabilities).
pValue	confidence level. Default value should be used.
mcAdj	if set to TRUE, a multiple comparisons adjustment using the Bonferroni method will be applied. Default value should be used; older (1.x and 2.x) versions of Boruta were effectively using FALSE.
maxRuns	maximal number of importance source runs. You may increase it to resolve attributes left Tentative.

<code>doTrace</code>	verbosity level. 0 means no tracing, 1 means reporting decision about each attribute as soon as it is justified, 2 means the same as 1, plus reporting each importance source run, 3 means the same as 2, plus reporting of hits assigned to yet undecided attributes.
<code>holdHistory</code>	if set to TRUE, the full history of importance is stored and returned as the <code>ImpHistory</code> element of the result. Can be used to decrease a memory footprint of Boruta in case this side data is not used, especially when the number of attributes is huge; yet it disables plotting of such made Boruta objects and the use of the <code>TentativeRoughFix</code> function.
<code>getImp</code>	function used to obtain attribute importance, called importance adapter in this documentation. The default is <code>getImpFruZ</code> , which runs Random Forest implementation from the <code>fru</code> package and gathers Z-scores of mean decrease accuracy measure. It should return a numeric vector of a size identical to the number of columns of its first argument, containing importance measure of respective attributes. Any order-preserving transformation of this measure will yield the same result. It is assumed that more important attributes get higher importance. $+\infty$ are accepted, NaNs and NAs are treated as 0s, with a warning.
<code>formula</code>	alternatively, formula describing model to be analysed.
<code>data</code>	in which to interpret formula.

Details

Boruta iteratively compares importances of attributes with importances of shadow attributes, created by shuffling original ones. Attributes that have significantly worst importance than shadow ones are being consecutively dropped. On the other hand, attributes that are significantly better than shadows are admitted to be Confirmed. Shadows are re-created in each iteration. Algorithm stops when only Confirmed attributes are left, or when it reaches `maxRuns` importance source runs. If the second scenario occurs, some attributes may be left without a decision. They are claimed Tentative. You may try to extend `maxRuns` or lower `pValue` to clarify them, but in some cases their importances do fluctuate too much for Boruta to converge. Instead, you can use `TentativeRoughFix` function, which will perform other, weaker test to make a final decision, or simply treat them as undecided in further analysis.

Value

An object of class Boruta, which is a list with the following components:

<code>finalDecision</code>	a factor of three value: Confirmed, Rejected or Tentative, containing final result of feature selection.
<code>ImpHistory</code>	a data frame of importances of attributes gathered in each importance source run. Beside predictors' importances, it contains maximal, mean and minimal importance of shadow attributes in each run. Rejected attributes get $-\infty$ importance. Set to NULL if <code>holdHistory</code> was given FALSE.
<code>timeTaken</code>	time taken by the computation.
<code>impSource</code>	string describing the source of importance, equal to a comment attribute of the <code>getImp</code> argument.
<code>call</code>	the original call of the Boruta function.

References

Miron B. Kursa, Witold R. Rudnicki (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11), p. 1-13. URL: [doi:10.18637/jss.v036.i11](https://doi.org/10.18637/jss.v036.i11)

conditionalTransdapter

Conditional transdapter

Description

Applies downstream importance source on a given object strata and averages their outputs.

Usage

```
conditionalTransdapter(groups, adapter = getImpFruZ)
```

Arguments

groups	groups.
adapter	importance adapter to transform.

Value

transformed importance adapter which can be fed into getImp argument of the [Boruta](#) function.

decohereTransdapter

Decohere transdapter

Description

Applies the decoherence transformation to the input, destroying all multivariate interactions. It will trash the Boruta result, only apply if you know what are you doing! Works only for categorical decision.

Usage

```
decohereTransdapter(adapter = getImpFruZ)
```

Arguments

adapter	importance adapter to transform.
---------	----------------------------------

Value

transformed importance adapter which can be fed into getImp argument of the [Boruta](#) function.

getConfirmedFormula	<i>Export Boruta result as a formula</i>
---------------------	--

Description

Functions which convert the Boruta selection into a formula, so that it could be passed further to other functions.

Usage

```
getConfirmedFormula(x)
```

```
getNonRejectedFormula(x)
```

Arguments

x an object of a class Boruta, made using a formula interface.

Value

Formula, corresponding to the Boruta results. `getConfirmedFormula` returns only Confirmed attributes, `getNonRejectedFormula` also adds Tentative ones.

Note

This operation is possible only when Boruta selection was invoked using a formula interface.

getImpFerns	<i>Random Ferns importance</i>
-------------	--------------------------------

Description

This function is intended to be given to a `getImp` argument of `Boruta` function to be called by the Boruta algorithm as an importance source.

Usage

```
getImpFerns(x, y, ...)
```

Arguments

x data frame of predictors including shadows.

y response vector.

... parameters passed to the underlying `rFerns` call; they are relayed from ... of `Boruta`.

Note

Random Ferns importance calculation should be much faster than using Random Forest; however, one must first optimize the value of the depth parameter and it is quite likely that the number of ferns in the ensemble required for the importance to converge will be higher than the number of trees in case of Random Forest.

 getImpFru

Fru Random Forest importance adapters

Description

Those function is intended to be given to a getImp argument of [Boruta](#) function to be called by the Boruta algorithm as an importance source. getImpFruZ generates default, normalized permutation importance while getImpFruRaw raw permutation importance. Fru does not support Gini index importance.

Usage

```
getImpFruZ(
  x,
  y,
  ntree = 500,
  trees = ntree,
  num.threads = 0,
  threads = num.threads,
  ...
)
```

```
getImpFruRaw(
  x,
  y,
  ntree = 500,
  trees = ntree,
  num.threads = 0,
  threads = num.threads,
  ...
)
```

Arguments

x	data frame of predictors including shadows.
y	response vector.
ntree	Number of trees in the forest; copied into fru 's native trees, put to retain transparent compatibility with randomForest.
trees	Number of trees in the forest, as according to fru 's nomenclature. If not given, set to ntree value. If both are given, trees takes precedence.

num.threads	Number of computing threads to use; copied into <code>fru</code> 's native threads, put to retain transparent compatibility with <code>ranger</code> .
threads	Number of computing threads to use, as according to <code>fru</code> 's nomenclature. If not given, set to <code>num.threads</code> . If both are given, <code>threads</code> takes precedence. Default value of zero means all available threads.
...	parameters passed to the underlying <code>fru</code> call; they are relayed from ... of <code>Boruta</code> .

Note

In prior versions of `Boruta`, other implementations of Random Forest were used as a default; prior to 5.0.0, `randomForest` package was used, and the default adapter used back then is available as `getImpLegacyRf`. Prior to 10.0, `ranger` package was used with `getImpRfZ`.

<code>getImpLegacyRf</code>	<i>randomForest importance adapters</i>
-----------------------------	---

Description

Those function is intended to be given to a `getImp` argument of `Boruta` function to be called by the `Boruta` algorithm as an importance source. `getImpLegacyRfZ` generates default, normalized permutation importance, `getImpLegacyRfRaw` raw permutation importance, finally `getImpLegacyRfGini` generates Gini index importance, all using `randomForest` as a Random Forest algorithm implementation.

Usage

```
getImpLegacyRfZ(x, y, ...)
```

```
getImpLegacyRfRaw(x, y, ...)
```

```
getImpLegacyRfGini(x, y, ...)
```

Arguments

<code>x</code>	data frame of predictors including shadows.
<code>y</code>	response vector.
...	parameters passed to the underlying <code>randomForest</code> call; they are relayed from ... of <code>Boruta</code> .

Note

The `getImpLegacyRfZ` function was a default importance source in `Boruta` versions prior to 5.0; since then `ranger` Random Forest implementation is used instead of `randomForest`, for speed, memory conservation and an ability to use multithreading. Both importance sources should generally lead to the same results, yet there are differences.

Most notably, `ranger` by default treats factor attributes as ordered (and works very slow if instructed otherwise with respect `.unordered.factors=TRUE`); on the other hand it lifts 32 levels limit specific to `randomForest`. To this end, Boruta decision for factor attributes may be different.

Random Forest methods has two main parameters, number of attributes tried at each split and the number of trees in the forest; first one is called `mtry` in both implementations, but the second `ntree` in `randomForest` and `num.trees` in `ranger`. To this end, to maintain compatibility, `getImpRf*` functions still accept `ntree` parameter relaying it into `num.trees`. Still, both parameters take the same defaults in both implementations (square root of the number all all attributes and 500 respectively).

Moreover, `ranger` brings some addition capabilities to Boruta, like analysis of survival problems or sticky variables which are always considered on splits.

Finally, the results for the same PRNG seed will be different.

getImpRf

Ranger Random Forest importance adapters

Description

Those function is intended to be given to a `getImp` argument of `Boruta` function to be called by the Boruta algorithm as an importance source. `getImpRfZ` generates default, normalized permutation importance, `getImpRfRaw` raw permutation importance, finally `getImpRfGini` generates Gini index importance.

Usage

```
getImpRfZ(x, y, ntree = 500, num.trees = ntree, ...)
```

```
getImpRfGini(x, y, ntree = 500, num.trees = ntree, ...)
```

```
getImpRfRaw(x, y, ntree = 500, num.trees = ntree, ...)
```

Arguments

<code>x</code>	data frame of predictors including shadows.
<code>y</code>	response vector.
<code>ntree</code>	Number of trees in the forest; copied into <code>ranger</code> 's native <code>num.trees</code> , put to retain transparent compatibility with <code>randomForest</code> .
<code>num.trees</code>	Number of trees in the forest, as according to <code>ranger</code> 's nomenclature. If not given, set to <code>ntree</code> value. If both are given, <code>num.trees</code> takes precedence.
<code>...</code>	parameters passed to the underlying <code>ranger</code> call; they are relayed from <code>...</code> of <code>Boruta</code> .

Note

`getImpRfZ` was the default importance source between versions 5.0.0 and 10.0.0. Prior to Boruta 5.0, `getImpLegacyRfZ` function was a default; see `getImpLegacyRf` for more details.

`getSelectedAttributes` *Extract names of the selected attributes*

Description

`getSelectedAttributes` returns a vector of names of attributes selected during a Boruta run.

Usage

```
getSelectedAttributes(x, withTentative = FALSE)
```

Arguments

`x` an object of a class `Boruta`, from which relevant attributes names should be extracted.

`withTentative` if set to `TRUE`, Tentative attributes will be also returned.

Value

A character vector with names of the relevant attributes.

`imputeTransdapter` *Impute transdapter*

Description

Wraps the importance adapter to accept NAs in input.

Usage

```
imputeTransdapter(adapter = getImpFruZ)
```

Arguments

`adapter` importance adapter to transform.

Value

transformed importance adapter which can be fed into `getImp` argument of the [Boruta](#) function.

Note

An all-NA feature will be converted to all zeroes, which should be ok as a totally non-informative value with most methods, but it is not universally correct. Ideally, one should avoid having such features in input altogether.

plot.Boruta	<i>Plot Boruta object</i>
-------------	---------------------------

Description

Default plot method for Boruta objects, showing boxplots of attribute importances over run.

Usage

```
## S3 method for class 'Boruta'
plot(
  x,
  colCode = c("green", "yellow", "red", "blue"),
  sort = TRUE,
  whichShadow = c(TRUE, TRUE, TRUE),
  col = NULL,
  xlab = "Attributes",
  ylab = "Importance",
  ...
)
```

Arguments

<code>x</code>	an object of a class Boruta.
<code>colCode</code>	a vector containing colour codes for attribute decisions, respectively Confirmed, Tentative, Rejected and shadow.
<code>sort</code>	controls whether boxplots should be ordered, or left in original order.
<code>whichShadow</code>	a logical vector controlling which shadows should be drawn; switches respectively max shadow, mean shadow and min shadow.
<code>col</code>	standard col attribute. If given, suppresses effects of colCode.
<code>xlab</code>	X axis label that will be passed to <code>boxplot</code> .
<code>ylab</code>	Y axis label that will be passed to <code>boxplot</code> .
<code>...</code>	additional graphical parameter that will be passed to <code>boxplot</code> .

Value

Invisible copy of x.

Note

If col is given and sort is TRUE, the col will be permuted, so that its order corresponds to attribute order in ImpHistory.

This function will throw an error when x lacks importance history, i.e., was made with holdHistory set to FALSE.

plotImpHistory *Plot Boruta object as importance history*

Description

Alternative plot method for Boruta objects, showing matplot of attribute importances over run.

Usage

```
plotImpHistory(
  x,
  colCode = c("green", "yellow", "red", "blue"),
  col = NULL,
  type = "l",
  lty = 1,
  pch = 0,
  xlab = "Classifier run",
  ylab = "Importance",
  ...
)
```

Arguments

x	an object of a class Boruta.
colCode	a vector containing colour codes for attribute decisions, respectively Confirmed, Tentative, Rejected and shadow.
col	standard col attribute, passed to <code>matplot</code> . If given, suppresses effects of colCode.
type	Plot type that will be passed to <code>matplot</code> .
lty	Line type that will be passed to <code>matplot</code> .
pch	Point mark type that will be passed to <code>matplot</code> .
xlab	X axis label that will be passed to <code>matplot</code> .
ylab	Y axis label that will be passed to <code>matplot</code> .
...	additional graphical parameter that will be passed to <code>matplot</code> .

Value

Invisible copy of x.

Note

This function will throw an error when x lacks importance history, i.e., was made with `holdHistory` set to FALSE.

print.Boruta	<i>Print Boruta object</i>
--------------	----------------------------

Description

Print method for the Boruta objects.

Usage

```
## S3 method for class 'Boruta'
print(x, ...)
```

Arguments

x an object of a class Boruta.
 ... additional arguments passed to `print`.

Value

Invisible copy of x.

srx	<i>Small redundant XOR data</i>
-----	---------------------------------

Description

A synthetic data set with 32 rows corresponding to all combinations of values of five logical features, A, B, N1, N2 and N3. The decision Y is equal to A xor B, hence N1–N3 are irrelevant attributes. The set also contains 3 additional features, A or B (AoB), A and B (AnB) and not A (nA), which provide a redundant, but still relevant way to reconstruct Y.

Usage

```
data(srx)
```

Format

A data frame with 8 predictors, 4 relevant: A, B, AoB, AnB and nA, as well as 3 irrelevant N1, N2 and N3, and decision attribute Y.

Details

This set is an easy way to demonstrate the difference between all relevant feature selection methods, which should select all features except N1–N3, and minimal optimal ones, which will probably ignore most of them.

Source

<https://blog.m bq.me/relevance-and-redundancy/>

TentativeRoughFix *Rough fix of Tentative attributes*

Description

In some circumstances (too short Boruta run, unfortunate mixing of shadow attributes, tricky dataset. . .), Boruta can leave some attributes Tentative. TentativeRoughFix performs a simplified, weaker test for judging such attributes.

Usage

```
TentativeRoughFix(x, averageOver = Inf)
```

Arguments

x	an object of a class Boruta.
averageOver	Either number of last importance source runs to average over or Inf for averaging over the whole Boruta run.

Details

Function claims as Confirmed those attributes that have median importance higher than the median importance of maximal shadow attribute, and the rest as Rejected. Depending of the user choice, medians for the test are count over last round, all rounds or N last importance source runs.

Value

A Boruta class object with modified finalDecision element. Such object has few additional elements:

originalDecision	Original finalDecision.
averageOver	Copy of averageOver parameter.

Note

This function should be used only when strict decision is highly desired, because this test is much weaker than Boruta and can lower the confidence of the final result.

x has to be made with holdHistory set to TRUE for this code to run.

Index

- * **datasets**
 - srx, [13](#)
- attStats, [2](#)
- Boruta, [3](#), [5–10](#)
- boxplot, [11](#)
- conditionalTransdapter, [5](#)
- decohereTransdapter, [5](#)
- fru, [7](#), [8](#)
- getConfirmedFormula, [6](#)
- getImpFruRaw (getImpFru), [7](#)
- getImpFerns, [6](#)
- getImpFru, [7](#)
- getImpFruRaw (getImpFru), [7](#)
- getImpFruZ (getImpFru), [7](#)
- getImpLegacyRf, [8](#), [8](#), [9](#)
- getImpLegacyRfGini (getImpLegacyRf), [8](#)
- getImpLegacyRfRaw (getImpLegacyRf), [8](#)
- getImpLegacyRfZ (getImpLegacyRf), [8](#)
- getImpRf, [9](#)
- getImpRfGini (getImpRf), [9](#)
- getImpRfRaw (getImpRf), [9](#)
- getImpRfZ, [8](#)
- getImpRfZ (getImpRf), [9](#)
- getLegacyImpRfRaw (getImpLegacyRf), [8](#)
- getNonRejectedFormula
(getConfirmedFormula), [6](#)
- getSelectedAttributes, [10](#)
- imputeTransdapter, [10](#)
- matplot, [12](#)
- plot.Boruta, [11](#)
- plotImpHistory, [12](#)
- print, [13](#)
- print.Boruta, [13](#)
- randomForest, [8](#), [9](#)
- ranger, [8](#), [9](#)
- rFerns, [6](#)
- srx, [13](#)
- TentativeRoughFix, [2](#), [4](#), [14](#)