

Package ‘rpartScore’

February 20, 2015

Type Package

Title Classification trees for ordinal responses

Version 1.0-1

Date 2012-12-08

Author Giuliano Galimberti, Gabriele Soffritti, Matteo Di Maso

Maintainer Giuliano Galimberti <giuliano.galimberti@unibo.it>

Depends R (>= 2.14.2), rpart

Suggests MASS

Description This package contains functions that allow to build classification trees for ordinal responses within the CART framework. The trees are grown using the Generalized Gini impurity function, where the misclassification costs are given by the absolute or squared differences in scores assigned to the categories of the response. Pruning is based on the total misclassification rate or on the total misclassification cost.

License GPL (>= 2)

LazyLoad yes

Repository CRAN

Date/Publication 2012-12-08 12:56:58

NeedsCompilation no

R topics documented:

rpartScore-package	2
evalMedian	4
evalMode	4
initScore	5
rpartScore	5
splitAbs	8
splitQuad	8

Index	9
--------------	----------

rpartScore-package *Classification trees for ordinal responses*

Description

This package contains functions that allow the user to build classification trees for ordinal responses within the CART framework.

The trees are grown using the Generalized Gini impurity function, where the misclassification costs are given by the absolute or squared differences in scores assigned to the categories of the response. Pruning is based on the total misclassification rate or on the total misclassification cost.

Details

Package:	rpartScore
Type:	Package
Version:	1.0-1
Date:	2012-12-08
License:	GPL (>=2)
LazyLoad:	yes

This package contains functions that allow the user to build classification trees for ordinal responses within the CART framework.

It is assumed that a set of numerical scores has been assigned to the ordered categories of the response.

Two splitting functions are implemented, both based on the generalized Gini impurity function. They use the absolute and the squared differences in scores, respectively, as misclassification costs. In order to select the optimal tree size, pruning can be performed, using two different measures of prediction performance: the total misclassification rate or the total misclassification cost.

This package requires the `rpart` package. The main function in this package is `rpartScore`. The use of this function is almost the same as the `rpart` function. The main difference is the presence of two arguments (`split` and `prune`) instead of the `method` argument.

The argument `split` controls the splitting function used to grow the classification tree, by setting the misclassification costs equal to the absolute ("`abs`" - default option) or to the squared ("`quad`") differences in scores.

The argument `prune` allows the user to select the prediction performance measure used to prune the classification tree, and can take two values: "`mr`" (total misclassification rate) or "`mc`" (total misclassification cost - default option).

Author(s)

Giuliano Galimberti, Gabriele Soffritti, Matteo Di Maso

Maintainer: Giuliano Galimberti <giuliano.galimberti@unibo.it>

References

- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. 1984 *Classification and Regression Trees*. Wadsworth International.
- Galimberti G., Soffritti G., Di Maso M. 2012 Classification Trees for Ordinal Responses in R: The rpartScore Package. *Journal of Statistical Software*, **47**(10), 1-25. URL <http://www.jstatsoft.org/v47/i10/>.
- Piccarreta R. 2008 Classification Trees for Ordinal Variables. *Computational Statistics*, **23**, 407-427.
- Therneau T.M., Atkinson E.J. 1997 An Introduction to Recursive Partitioning Using rpart Routines. *Technical Report 61*, Section of Biostatistics, Mayo Clinic, Rochester. URL <http://www.mayo.edu/hsr/techrpt/61.pdf>.

See Also

[rpart](#)

Examples

```
data("birthwt",package="MASS")

birthwt$Category.s <- ifelse(birthwt$bwt <= 2500, 3,
  ifelse(birthwt$bwt <= 3000, 2,
    ifelse(birthwt$bwt <= 3500, 1, 0)))

T.abs.mc <- rpartScore(Category.s ~ age + lwt + race + smoke +
  ptl + ht + ui + ftv, data = birthwt)

plotcp(T.abs.mc)

T.abs.mc.pruned<-prune(T.abs.mc,cp=0.02)

plot(T.abs.mc.pruned)

text(T.abs.mc.pruned)

T.abs.mr <- rpartScore(Category.s ~ age + lwt + race + smoke +
  ptl + ht + ui + ftv, data = birthwt, prune = "mr")

T.quad.mc <- rpartScore(Category.s ~ age + lwt + race + smoke +
  ptl + ht + ui + ftv, split = "quad", data = birthwt)

T.quad.mr <- rpartScore(Category.s ~ age + lwt + race + smoke + ptl + ht +
  ui + ftv, split = "quad", prune = "mr", data = birthwt)
```

evalMedian

Pruning a classification tree using the total misclassification cost

Description

This function is not invoked directly by the user but is used for its effects in the pruning procedure. See Galimberti *et al.* (2012) for further details.

Author(s)

Giuliano Galimberti, Gabriele Soffritti, Matteo Di Maso

References

Galimberti G., Soffritti G., Di Maso M. 2012 Classification Trees for Ordinal Responses in R: The rpartScore Package. *Journal of Statistical Software*, 47(10), 1-25. URL <http://www.jstatsoft.org/v47/i10/>.

See Also

[rpartScore](#)

evalMode

Pruning a classification tree using the total misclassification rate

Description

This function is not invoked directly by the user but is used for its effects in the pruning procedure. See Galimberti *et al.* (2012) for further details.

Author(s)

Giuliano Galimberti, Gabriele Soffritti, Matteo Di Maso

References

Galimberti G., Soffritti G., Di Maso M. 2012 Classification Trees for Ordinal Responses in R: The rpartScore Package. *Journal of Statistical Software*, 47(10), 1-25. URL <http://www.jstatsoft.org/v47/i10/>.

See Also

[rpartScore](#)

initScore

Summarizing and visualizing classification trees

Description

This function is not invoked directly by the user but is used for summarizing and visualizing a classification tree. See Galimberti *et al.* (2012) for further details.

Author(s)

Giuliano Galimberti, Gabriele Soffritti, Matteo Di Maso

References

Galimberti G., Soffritti G., Di Maso M. 2012 Classification Trees for Ordinal Responses in R: The rpartScore Package. *Journal of Statistical Software*, **47**(10), 1-25. URL <http://www.jstatsoft.org/v47/i10/>.

See Also

[rpartScore](#)

rpartScore

Classification trees for ordinal responses

Description

This function allows the user to build classification trees for ordinal responses within the CART framework. The trees are grown using the Generalized Gini impurity function, where the misclassification costs are given by the absolute or squared differences in scores assigned to the categories of the response. Pruning is based on the total misclassification rate or on the total misclassification cost.

Usage

```
rpartScore(formula, data, weights, subset, na.action = na.rpart,  
  split = "abs", prune = "mc",  
  model = FALSE, x = FALSE, y = TRUE,  
  control, ...)
```

Arguments

<code>formula</code>	a formula, as in the <code>lm</code> function.
<code>data</code>	an optional data frame in which to interpret the variables named in the formula.
<code>weights</code>	optional case weights.
<code>subset</code>	optional expression saying that only a subset of the rows of the data should be used in the fit.
<code>na.action</code>	The default action deletes all observations for which <code>y</code> is missing, but keeps those in which one or more predictors are missing.
<code>split</code>	One of "abs" or "quad".
<code>prune</code>	One of "mc" or "mr".
<code>model</code>	if logical: keep a copy of the model frame in the result? If the input value for <code>model</code> is a model frame (likely from an earlier call to the <code>rpart</code> or <code>rpartScore</code> function), then this frame is used rather than constructing new data.
<code>x</code>	keep a copy of the <code>x</code> matrix in the result.
<code>y</code>	keep a copy of the dependent variable in the result. If missing and <code>model</code> is supplied this defaults to <code>FALSE</code> .
<code>control</code>	options that control details of the <code>rpart</code> algorithm.
<code>...</code>	arguments to <code>rpart.control</code> may also be specified in the call to <code>rpartScore</code> . They are checked against the list of valid arguments.

Details

The use of this function is almost the same as the `rpart` function.

It is assumed that a set of (not necessarily linear) numerical scores has been assigned to the ordered categories of the response.

The main difference with respect to the `rpart` function is the presence of two arguments (`split` and `prune`) instead of the `method` argument.

The argument `split` controls the splitting function used to grow the classification tree, by setting the misclassification costs in the generalized Gini impurity function equal to the absolute ("abs" - is the default option) or to the squared ("quad") differences in scores.

The argument `prune` allows the user to select the prediction performance measure used to prune the classification tree, and can take two values: "mr" (total misclassification rate) or "mc" (total misclassification cost - is the default option).

Value

An object of class `rpart`, a superset of class `tree`.

Author(s)

Giuliano Galimberti, Gabriele Soffritti, Matteo Di Maso

References

- Breiman L., Friedman J.H., Olshen R.A., Stone C.J. 1984 *Classification and Regression Trees*. Wadsworth International.
- Galimberti G., Soffritti G., Di Maso M. 2012 Classification Trees for Ordinal Responses in R: The rpartScore Package. *Journal of Statistical Software*, **47**(10), 1-25. URL <http://www.jstatsoft.org/v47/i10/>.
- Piccarreta R. 2008 Classification Trees for Ordinal Variables. *Computational Statistics*, **23**, 407-427.
- Therneau T.M., Atkinson E.J. 1997 An Introduction to Recursive Partitioning Using rpart Routines. *Technical Report 61*, Section of Biostatistics, Mayo Clinic, Rochester. URL <http://www.mayo.edu/hsr/techrpt/61.pdf>.

See Also

[rpart](#), [rpart.control](#), [rpart.object](#), [summary.rpart](#), [print.rpart](#)

Examples

```
data("birthwt", package="MASS")

birthwt$Category.s <- ifelse(birthwt$bwt <= 2500, 3,
  ifelse(birthwt$bwt <= 3000, 2,
    ifelse(birthwt$bwt <= 3500, 1, 0)))

T.abs.mc <- rpartScore(Category.s ~ age + lwt + race + smoke +
  ptl + ht + ui + ftv, data = birthwt)

plotcp(T.abs.mc)

T.abs.mc.pruned <- prune(T.abs.mc, cp=0.02)

plot(T.abs.mc.pruned)

text(T.abs.mc.pruned)

T.abs.mr <- rpartScore(Category.s ~ age + lwt + race + smoke +
  ptl + ht + ui + ftv, data = birthwt, prune = "mr")

T.quad.mc <- rpartScore(Category.s ~ age + lwt + race + smoke +
  ptl + ht + ui + ftv, split = "quad", data = birthwt)

T.quad.mr <- rpartScore(Category.s ~ age + lwt + race + smoke + ptl + ht +
  ui + ftv, split = "quad", prune = "mr", data = birthwt)
```

splitAbs *Generalized Gini splitting function based on absolute differences in scores*

Description

This function is not invoked directly by the user but is used for its effects in the tree growing procedure. See Galimberti *et al.* (2012) for further details.

Author(s)

Giuliano Galimberti, Gabriele Soffritti, Matteo Di Maso

References

Galimberti G., Soffritti G., Di Maso M. 2012 Classification Trees for Ordinal Responses in R: The rpartScore Package. *Journal of Statistical Software*, 47(10), 1-25. URL <http://www.jstatsoft.org/v47/i10/>.

See Also

[rpartScore](#)

splitQuad *Generalized Gini splitting function based on quadratic differences in scores*

Description

This function is not invoked directly by the user but is used for its effects in the tree growing procedure. See Galimberti *et al.* (2012) for further details.

Author(s)

Giuliano Galimberti, Gabriele Soffritti, Matteo Di Maso

References

Galimberti G., Soffritti G., Di Maso M. 2012 Classification Trees for Ordinal Responses in R: The rpartScore Package. *Journal of Statistical Software*, 47(10), 1-25. URL <http://www.jstatsoft.org/v47/i10/>.

See Also

[rpartScore](#)

Index

*Topic **tree**

 rpartScore, 5

 rpartScore-package, 2

evalMedian, 4

evalMode, 4

initScore, 5

print.rpart, 7

rpart, 3, 7

rpart.control, 7

rpart.object, 7

rpartScore, 4, 5, 5, 8

rpartScore-package, 2

splitAbs, 8

splitQuad, 8

summary.rpart, 7