Package ‘palmtree’

May 15, 2020

Title Partially Additive (Generalized) Linear Model Trees

Date 2020-05-15

Version 0.9-1

Description This is an implementation of model-based trees with global model parameters (PALM trees). The PALM tree algorithm is an extension to the MOB algorithm (implemented in the 'partykit' package), where some parameters are fixed across all groups. Details about the method can be found in Seibold, Hothorn, Zeileis (2016) <arXiv:1612.07498>. The package offers coef(), logLik(), plot(), and predict() functions for PALM trees.

Depends R (>= 3.1.0), partykit

Imports Formula (>= 1.2-1)

Suggests mvtnorm, psychotools

License GPL-2 | GPL-3

RoxygenNote 6.0.1

NeedsCompilation no

Author Heidi Seibold [aut, cre] (<https://orcid.org/0000-0002-8960-9642>), Torsten Hothorn [aut] (<https://orcid.org/0000-0001-8301-0471>), Achim Zeileis [aut] (<https://orcid.org/0000-0003-0918-3766>)

Maintainer Heidi Seibold <heidi@seibold.co>

Repository CRAN

Date/Publication 2020-05-15 12:00:02 UTC

R topics documented:

palmtree ........................................... 2

Index 5
palmtree \hspace{1cm} \textit{Partially Additive (Generalized) Linear Model Trees}

Description

Model-based recursive partitioning based on (generalized) linear models with some local (i.e., leaf-specific) and some global (i.e., constant throughout the tree) regression coefficients.

Usage

```r
palmtree(formula, data, weights = NULL, family = NULL,
  lmstart = NULL, abstol = 0.001, maxit = 100,
  dfsplit = TRUE, verbose = FALSE, plot = FALSE, ...)
```

Arguments

- **formula**: formula specifying the response variable and a three-part right-hand-side describing the local (i.e., leaf-specific) regressors, the global regressors (i.e., with constant coefficients throughout the tree), and partitioning variables, respectively. For details see below.
- **data**: data.frame to be used for estimating the model tree.
- **weights**: numeric. An optional numeric vector of weights. (Note that this is passed with standard evaluation, i.e., it is not enough to pass the name of a column in data.)
- **family**: either NULL so that \texttt{lm/lmtree} are used or family specification for \texttt{glm/glmtree}. See \texttt{glm} documentation for families.
- **lmstart**: numeric. A vector of length \texttt{nrow(data)}, to be used as an offset in estimation of the first tree. NULL by default, which results in an initialization with the global model.
- **abstol**: numeric. The convergence criterion used for estimation of the model. When the difference in log-likelihoods of the model from two consecutive iterations is smaller than abstol, estimation of the model tree has converged.
- **maxit**: numeric. The maximum number of iterations to be performed in estimation of the model tree.
- **dfsplit**: logical or numeric. \texttt{as.integer(dfsplit)} is the degrees of freedom per selected split employed when extracting the log-likelihood.
- **verbose**: Should the log-likelihood value of the estimated model be printed for every iteration of the estimation?
- **plot**: Should the tree be plotted at every iteration of the estimation? Note that selecting this option slows down execution of the function.
- **...**: Additional arguments to be passed to \texttt{lmtree()} or \texttt{glmtree()}. See \texttt{mob_control} documentation for details.
Details

Partially additive (generalized) linear model (PALM) trees learn a tree where each terminal node is associated with different regression coefficients while adjusting for additional global regression effects. This allows for detection of subgroup-specific coefficients with respect to selected covariates, while keeping the remaining regression coefficients constant throughout the tree. The estimation algorithm iterates between (1) estimation of the tree given an offset of the global effects, and (2) estimation of the global regression effects given the tree structure.

To specify all variables in the model a formula such as $y \sim x_1 + x_2 | x_3 | z_1 + z_2 + z_3$ is used, where $y$ is the response, $x_1$ and $x_2$ are the regressors in every node of the tree, $x_3$ has a global regression coefficients, and $z_1$ to $z_3$ are the partitioning variables considered for growing the tree.

The code is still under development and might change in future versions.

Value

The function returns a list with the following objects:

- `formula` The formula as specified with the `formula` argument.
- `call` The matched call.
- `tree` The final `lmtree/glmmtree`.
- `palm` The final `lm/glm` model.
- `data` The dataset specified with the `data` argument including added auxiliary variables `.lm` and `.tree` from the last iteration.
- `nobs` Number of observations.
- `loglik` The log-likelihood value of the last iteration.
- `df` Degrees of freedom.
- `dfsplit` degrees of freedom per selected split as specified with the `dfsplit` argument.
- `iterations` The number of iterations used to estimate the `palmtree`.
- `maxit` The maximum number of iterations specified with the `maxit` argument.
- `lmstart` Offset in estimation of the first tree as specified in the `lmstart` argument.
- `abstol` The prespecified value for the change in log-likelihood to evaluate convergence, as specified with the `abstol` argument.
- `intercept` Logical specifying if an intercept was computed.
- `family` The `family` object used.
- `mob.control` A list containing control parameters passed to `lmtree()`, as specified with . . .

References


See Also

`lm`, `glm`, `lmtree`, `glmmtree`
Examples

```r
## one DGP from Sies and Van Mechelen (2015)
dgp <- function(nobs = 1000, nreg = 5, creg = 0.4, ptreat = 0.5, sd = 1,
    coef = c(1, 0.25, 0.25, 0, 0, -0.25), eff = 1)
{
d <- mvtnorm::rmvnorm(nobs,
    mean = rep(0, nreg),
    sigma = diag(1 - creg, nreg) + creg)
colnames(d) <- paste0("x", 1:nreg)
d <- as.data.frame(d)
d$a <- rbinom(nobs, size = 1, prob = ptreat)
d$err <- rnorm(nobs, mean = 0, sd = sd)
go <- function(d) {
  as.numeric(d$x1 > -0.545) * as.numeric(d$x2 < 0.545)
}
d$y <- coef[1] + drop(as.matrix(d[, paste0("x", 1:5)]) %*% coef[-1]) -
eff * (d$a - go(d))^2 + d$err
d$a <- factor(d$a)
return(d)
}
set.seed(1)
d <- dgp()

## estimate PALM tree with correctly specified global (partially
## additive) regressors and all variables considered for partitioning
palm <- palmtree(y ~ a | x1 + x2 + x5 | x1 + x2 + x3 + x4 + x5, data = d)
print(palm)
plot(palm)

## query coefficients
coef(palm, model = "tree")
coef(palm, model = "palm")
coef(palm, model = "all")
```
Index

*Topic tree
  palmtree, 2
  coef.palmtree (palmtree), 2
  glm, 2, 3
  glmtree, 3
  lm, 3
  lmtree, 3
  logLik.palmtree (palmtree), 2
  mob_control, 2
  palmtree, 2
  plot.palmtree (palmtree), 2
  predict.palmtree (palmtree), 2
  print.palmtree (palmtree), 2