Package ‘acepack’

August 22, 2023

Maintainer Shawn Garbett <shawn.garbett@vumc.org>
Version 1.4.2
Author Phil Spector, Jerome Friedman, Robert Tibshirani, Thomas Lumley, Shawn Garbett, Jonathan Baron

Title ACE and AVAS for Selecting Multiple Regression Transformations
License MIT + file LICENSE
Suggests testthat
Repository CRAN
Date/Publication 2023-08-22 09:10:02 UTC

\textbf{R topics documented:}

\begin{verbatim}
ace ................................................................. 2
avas ............................................................... 4
\end{verbatim}

Index 7
Alternating Conditional Expectations

Description

Uses the alternating conditional expectations algorithm to find the transformations of y and x that maximise the proportion of variation in y explained by x. When x is a matrix, it is transformed so that its columns are equally weighted when predicting y.

Usage

```r
ace(x, y, wt = rep(1, nrow(x)), cat = NULL, mon = NULL, lin = NULL, 
circ = NULL, delrsq = 0.01)
```

Arguments

- `x`: a matrix containing the independent variables.
- `y`: a vector containing the response variable.
- `wt`: an optional vector of weights.
- `cat`: an optional integer vector specifying which variables assume categorical values. Positive values in `cat` refer to columns of the x matrix and zero to the response variable. Variables must be numeric, so a character variable should first be transformed with `as.numeric()` and then specified as categorical.
- `mon`: an optional integer vector specifying which variables are to be transformed by monotone transformations. Positive values in `mon` refer to columns of the x matrix and zero to the response variable.
- `lin`: an optional integer vector specifying which variables are to be transformed by linear transformations. Positive values in `lin` refer to columns of the x matrix and zero to the response variable.
- `circ`: an integer vector specifying which variables assume circular (periodic) values. Positive values in `circ` refer to columns of the x matrix and zero to the response variable.
- `delrsq`: termination threshold. Iteration stops when R-squared changes by less than `delrsq` in 3 consecutive iterations (default 0.01).

Value

A structure with the following components:

- `x`: the input x matrix.
- `y`: the input y vector.
- `tx`: the transformed x values.
- `ty`: the transformed y values.
- `rsq`: the multiple R-squared value for the transformed values.
- `l`: the codes for `cat`, `mon`, ...
- `m`: not used in this version of `ace`
References


The R code is adapted from S code for avas() by Tibshirani, in the Statlib S archive; the FORTRAN is a double-precision version of FORTRAN code by Friedman and Spector in the Statlib general archive.

Examples

\[
\text{TWOPI} \leftarrow 8 \cdot \text{atan}(1) \\
\text{x} \leftarrow \text{runif}(200, 0, \text{TWOPI}) \\
\text{y} \leftarrow \exp(\sin(\text{x}) + \text{rnorm}(200)/2) \\
\text{a} \leftarrow \text{ace(x, y)} \\
\text{par(mfrow} = \text{c(3,1))} \\
\text{plot(a}$y, a$ty) \quad \# \text{view the response transformation} \\
\text{plot(a}$x, a$tx) \quad \# \text{view the carrier transformation} \\
\text{plot(a$tx, a$ty)} \quad \# \text{examine the linearity of the fitted model}
\]

\# example when x is a matrix
\[
\text{X1} \leftarrow 1:10 \\
\text{X2} \leftarrow \text{X1}^2 \\
\text{X} \leftarrow \text{cbind(X1,X2)} \\
\text{Y} \leftarrow 3 \times \text{X1} + \text{X2} \\
\text{a1} \leftarrow \text{ace(X,Y)} \\
\text{plot(rowSums(a1$tx), a1$y)} \\
(\text{lm(a1$y - a1$tx)}) \# \text{shows that the columns of X are equally weighted}
\]

\[
\text{X1} \leftarrow \text{runif(100)} \times 2 - 1 \\
\text{X2} \leftarrow \text{runif(100)} \times 2 - 1 \\
\text{X3} \leftarrow \text{runif(100)} \times 2 - 1 \\
\text{X4} \leftarrow \text{runif(100)} \times 2 - 1 \\
\]

\# Original equation of Y:
\[
\text{Y} \leftarrow \log(4 + \sin(3 \times \text{X1}) + \text{abs(X2)} + \text{X3}^2 + \text{X4} + 0.1 \times \text{rnorm(100)})
\]

\# Transformed version so that Y, after transformation, is a linear function of transforms of the X variables:
\# \exp(Y) = 4 + \sin(3 \times \text{X1}) + \text{abs(X2)} + \text{X3}^2 + \text{X4}
\#
\text{a1} \leftarrow \text{ace(cbind(X1,X2,X3,X4), Y)}
\#
\text{For each variable, show its transform as a function of the original variable and the of the transform that created it,}
\text{showing that the transform is recovered.}
\text{par(mfrow} = \text{c(2,1))}
\text{plot(}\text{X1, a1$tx}[{,1}]) \\
\text{plot(}\sin(3 \times \text{X1}), a1$tx}[{,1}])
avas

Additivity and variance stabilization for regression

Description

Estimate transformations of \( x \) and \( y \) such that the regression of \( y \) on \( x \) is approximately linear with constant variance

Usage

```
avas(x, y, wt = rep(1, nrow(x)), cat = NULL, mon = NULL, 
lin = NULL, circ = NULL, delrsq = 0.01, yspan = 0)
```

Arguments

- **x**: a matrix containing the independent variables.
- **y**: a vector containing the response variable.
- **wt**: an optional vector of weights.
- **cat**: an optional integer vector specifying which variables assume categorical values. Positive values in `cat` refer to columns of the \( x \) matrix and zero to the response variable. Variables must be numeric, so a character variable should first be transformed with `as.numeric()` and then specified as categorical.
- **mon**: an optional integer vector specifying which variables are to be transformed by monotone transformations. Positive values in `mon` refer to columns of the \( x \) matrix and zero to the response variable.
- **lin**: an optional integer vector specifying which variables are to be transformed by linear transformations. Positive values in `lin` refer to columns of the \( x \) matrix and zero to the response variable.
- **circ**: an integer vector specifying which variables assume circular (periodic) values. Positive values in `circ` refer to columns of the \( x \) matrix and zero to the response variable.
- **delrsq**: termination threshold. Iteration stops when R-squared changes by less than `delrsq` in 3 consecutive iterations (default 0.01).
- **yspan**: Optional window size parameter for smoothing the variance. Range is \([0, 1]\). Default is 0 (cross validated choice). .5 is a reasonable alternative to try.
Value

A structure with the following components:

- **x**: the input x matrix.
- **y**: the input y vector.
- **tx**: the transformed x values.
- **ty**: the transformed y values.
- **rsq**: the multiple R-squared value for the transformed values.
- **l**: the codes for cat, mon, ...
- **m**: not used in this version of avas
- **yspan**: span used for smoothing the variance
- **iters**: iteration number and rsq for that iteration
- **niters**: number of iterations used

References


Examples

```r
TWOPI <- 8*atan(1)
x <- runif(200,0,TWOPI)
y <- exp(sin(x)+rnorm(200)/2)
a <- avas(x,y)
par(mfrow=c(3,1))
plot(a$y,a$ty) # view the response transformation
plot(a$x,a$tx) # view the carrier transformation
plot(a$tx,a$ty) # examine the linearity of the fitted model

# From D. Wang and M. Murphy (2005), Identifying nonlinear relationships
# regression using the ACE algorithm. Journal of Applied Statistics, 
# 32, 243-258, adapted for avas.
X1 <- runif(100)*2-1
X2 <- runif(100)*2-1
X3 <- runif(100)*2-1
X4 <- runif(100)*2-1

# Original equation of Y:
Y <- log(4 + sin(3*X1) + abs(X2) + X3^2 + X4 + .1*rnorm(100))

# Transformed version so that Y, after transformation, is a 
# linear function of transforms of the X variables:
# exp(Y) = 4 + sin(3*X1) + abs(X2) + X3^2 + X4
a1 <- avas(cbind(X1,X2,X3,X4),Y)
par(mfrow=c(2,1))
```
# For each variable, show its transform as a function of
# the original variable and the of the transform that created it,
# showing that the transform is recovered.
plot(X1,a1$tx[,1])
plot(sin(3*X1),a1$tx[,1])

plot(X2,a1$tx[,2])
plot(abs(X2),a1$tx[,2])

plot(X3,a1$tx[,3])
plot(X3^2,a1$tx[,3])

plot(X4,a1$tx[,4])
plot(X4,a1$tx[,4])

plot(Y,a1$ty)
plot(exp(Y),a1$ty)
Index

* models
  ace, 2
  avas, 4

ace, 2
avas, 4