Package 'marginalmodelplots'

January 2, 2012

Type Package

Title Marginal Mean Model Plots

Version 0.4.2

Date 12/15/2009

Author Andrew Redd

Maintainer Andrew Redd <aredd@stat.tamu.edu>

Depends R (>= 2.0.1), locfit (>= 1.5-4), grid

Description Marginal mean model plots for linear (lm) and generalized linear models (glm). Including tools for bandwidth exploration. Built on the work from the alr3 and locfit packages.

License GPL (>= 2)

Repository CRAN

Date/Publication 2009-12-16 07:37:39

R topics documented:

`marginalmodelplots-package` *Marginal Model Plots*

Description

Marginal mean model plots for linear (lm) and generalized linear models (glm). Including tools for bandwidth exploration. Built on the work from the alr3 and locfit package.

Details

Package:	marginalmodelplots
Type:	Package
Version:	0.6
Date:	2007-05-15
License:	GPL version 2 or newer.

The primary function of interest is `mmplot`, which creates a layout of marginal mean model plots for a given model. The package also includes functions for exploring marginal model plots and experimentation with bandwidths through the functions `mmpslider` and `mmpdualslider`.\ marginalmodelplots is based on the work of marginal model plots by Cook and Weisburg, and extends the their work in from the package `alr3`. Unlike `alr3` the marginalmodelplots package only performs mean curve smooths. For variance curve estimation use the `alr3` package.

Author(s)

Andrew Redd

Maintainer: Andrew Redd <aredd at stat.tamu.edu>

References

S. Weisberg (2005), *Applied Linear Regression*, third edition, Wiley, Chapter 8

`aicc` *Akaike's Information Criteria - Corrected for Non-Parametic Models*

Description

Calculates Akaike's information criteria corrected for non parametric models made with `locfit`.

Usage

```
aicc(x,...,pen=2)
```

Arguments

x	A model or model formula.
...	Other arguments to locfit
pen	Penalty for the degrees of freedom term.

Value

vector of length 4 with the value of the aicc in the fourth spot.

Author(s)

Andrew Redd, <aredd at stat.tamu.edu>

References

Chiou, J. and Tsai, C. (2006). Smoothing parameter selection in quasi-likelihood models. *Nonparametric Statistics*, 18(3), 307-314.

See Also

aic, gcv

Bordeaux *Ratings of Bordeaux Wines*

Description

A dataset comparing the rating of expensive bordeaux wines from two prominent wine critics.

Usage

```
Bordeaux
```

Format

A dataframe with 72 observations.

Source

Compiled by Dr. Simon Sheather.

MichelinNY *Data from the Michelin Guide for New York City Resaurants*

Description

The Michelin Guide publishes travel information world wide. An award that the Michelin Guide give is the coveted michelin star, a star in the michelin resaurant guide. The awards are given based on a confidential evaluation process. The Zagat Survey on the other hand is reviews and ratings based on comsumer input.\ The dataset MichelinNY contains information from both the Michelin Guide and the Zagat Survey.

Usage

```
data(MichelinNY)
```

Format

A dataframe with 165 observations.

Source

Michelin Guide New York City 2006, Michelin Travel Publications, Greenville, South Carolina\ Zagat survey:2006 New York City Restaurants, Available online at www.zagat.com

MissAmerica08 *Data for finalists in the Miss America Pageant*

Description

A dataset for the miss america pageant for the years of 2000 through 2008. Included demographic and geographic data for each of the 50 states plus DC.

Usage

```
data(MissAmerica08)
```

Format

A dataframe with 51 observations.

Source

Compiled by Dr. Simon Sheather.

`mmplot` *Marginal model plot*

Description

Marginal model plots are a non parametric form of verifying model adequacy. The Structure of a marginal model plot is a plot with the response variable along the vertical axis, and a predictor variable, or any linear combination of predictor variables allong the horizontal axis. Two non parametric smooths are added to the plots. One for the response on the predictor and another for the fitted values on the predictor. If the two curves match closely then the model is determined to be adaquate.

The `mmplot` fuction, as implimented here, is capable of computing an optimal bandwidth using generalized cross validation (`gcv`), Akaike's information criteria (`aic`), or Akaike's information criteria corrected (`aicc`).

Usage

```
mmplot(object,exclude=NULL,layout=NULL,ask,...)

mmplot1(object, ...)
mmplot1.lm(object,u=predict(object),
label=deparse(substitute(u)), locfit.control=list(nn=2/3),
    colors=c("blue","red"), ...)
mmplot1.glm(object,u=predict(object),mean=TRUE,
    label=deparse(substitute(u)), locfit.control=list(nn=2/3), subset=NULL,
    colors=c("blue","red"),family=NULL,link=NULL,yhat.autoquasi=TRUE,
inc.legend=TRUE,...)
```

Arguments

`object`	A regression object, of type either lm or glm, for which there is a `predict` method defined.
`exclude`	A vector of indices of variables for which a plot is not required. For example, `c(2,4)` would exclude the second and fourth terms in the model, excluding the intercept.
`layout`	A reasonable layout for the plots in the window is determined by the program. If you don't like the default you can set your own layout: `c(2,3)` means two rows and three columns.
`ask`	If TRUE, ask before clearing the graph window to draw more plots.
`u`	The quantity to be plotted on the horizontal axis. The default is the predicted values `predict(object)`. Can be any other vector of length equal to the number of observations in the object.
`mean`	If TRUE, compare mean smooths. Depreciated, should allways be left as TRUE.
`label`	label for horizontal axis

`locfit.control`	A list of control parameters to pass to the locfit function to adjust the nature of the curve. the primary purpose of this will be to set the bandwidth or nearest neighbor parameter for smoothing, although any of the `locfit` parameters may be set here. To use a optimally computed bandwidth use `h='aic'` or `nn='gcv'` for an AIC optimal fixed bandwidth or a GCV optimal nearest neighbor, respectivly.
`subset`	logical vector indicating the observations to use. See subset.
`colors`	colors for response and fitted values smooths, respectively
`family`	Overrides the family used for the local glm fit in glm models.
`link`	Overrides the link used for the local glm fit in glm models.
`yhat.autoquasi`	Determines if the quasi version of the family should be used to fit the fitted values curve. The default is True, since the fitted values do not have constant variance.
`inc.legend`	Determines if there is a legend included on the plot. Defaults to TRUE.
`...`	additional arguments passed to `plot`

Details

`mmplot1` draws one marginal model plot. `mmplot` draws all marginal model plots, versus each of the terms in the model and versus fitted values. `mmplot` should not be used if some of the terms in the model are factors or interactions.

Special attention is paid to logistic regression models. The vertical axis is labeled with the levels of the factor if the response was a factor. Also points are colored if the predicted outcome differs from the observed response.

The `mmplot` implimentation here uses the locfit package to generate the non parametric smooths. A `lm` model is equivalent to using a local polynomial smooth, but with `glm` models the smooth is a local glm fit of the same family as the model, unless overridden by user.

Value

`mmplot1` functions return (invisible) an object with the smothing parameters used. The first componenent being the parameters for the data (named `y`), the second for the fitted values (named `yhat`). The components of the smoothing parameters are first the bandwidth (`h`) if used, otherwise `NULL`. Second the nearest neighbor parameter(`nn`) if used, otherwise `NULL`. The `mmplot` function returns an unnamed list of the smoothing parameters in the order the graph appear.

Note

A large number of warnings can be produced from locfit when running mmplot, especially with optimized smoothing parameters. Warnings that are passed from the locfit procedures are for the final fit.

Author(s)

Andrew Redd, `<aredd at stat.tamu.edu>`

References

S. Weisberg (2005), *Applied Linear Regression*, third edition, Wiley, Chapter 8

See Also

locfit, plot

Examples

```
data(Bordeaux)
m<-lm(Price~ParkerPoints+CoatesPoints+FirstGrowth+Pomerol+VintageSuperstar+CultWine,data=Bordeaux)
mmplot(m,locfit.control=list(nn=2/3))

data(MichelinNY)
m4 <- glm(InMichelin~Food+Decor*Service+log(Cost),family=binomial(),data=MichelinNY)
mmplot(m4)
mmplot(m4,locfit.control=list(nn=.7))
mmplot(m4,locfit.control=list(nn='aic'))
```

---

mmpslider *Margial Model Plots with Single and Dual Slider Bars*

Description

`mmpslider` Creates an interactive marginal model plot for the supplied model and predictor variable. The graph provides an area at the bottom of the plot with a slider bar that can be used to interactivly adjust the fixed bandwidth used in generating the nonparametric smooths in the plot. `mmpdualslider` behaves behaves like `mmpslider` but creates two slider bars one for each fit, the response and the fitted valeues. The slider bar is not actually a slider bar but an interactive region of the graph that will adjust the fixed bandwidth(s) of the non parametric smooth when clicked. It does not actually slide.

Usage

```
mmpslider(m, ...)
mmpslider.lm(m,pred=predict(m), bw=NULL , label=deparse(substitute(u)), colors=c('blue','red'),...)
mmpslider.glm(m,pred=predict(m), bw=NULL , label=deparse(substitute(u)), colors=c('blue','red'), fami
```

Arguments

`m`	Model to use to generate marginal model plot from.
`pred`	The predictor variable or cobination of variables to use for the plot. Defaults to the linear predictor of the model.
`bw`	Vector of length 3 that contains the minimun allowable bandwidth, the starting bandwidth and the maximum allowable bandwidth. This setting sets up the range of the slider bar. The default for the minumum bandwidth is the maximum

	of the differences of sequential points. The default starting value is the aic optimum value and te default maximum is twice the aic optimum bandwidth. Either a vector of length 3 must be provided of left null. `mmpdualslider` uses a local glm fit for marginal model plots with glm models.
`label`	The x axis label. So named to reflect the same parameter in similar mmp code.
`colors`	Vector of length 2 containg the colors for the smooth against the response and the fitted values, respectivly.
`family`	Overrides the family used for the local glm fit in glm models.
`link`	Overrides the link used for the local glm fit in glm models.
`...`	Passes arguments onto other functions

Details

To use `mmpslider` run the function then click in the slider bar area to adjust the bandwidth of the curve. Clicking the stop button, or outside of the slider bars area will finalize the plot and return control to the consol.\ The plot is a custom plot generated from scratch with the grid utilities. Any further customization of the plot will have to be used with in the grid framework.

Value

none. Used for side effects of generating a plot.

Author(s)

Andrew Redd <aredd at tamu.edu>

References

S. Weisberg (2005), *Applied Linear Regression*, third edition, Wiley, Chapter 8

See Also

`marginalmodelplots-package`

Examples

```
## Not run:
data(Bordeaux)
names(Bordeaux)
attach(Bordeaux)
m<-lm(Price~ParkerPoints+CoatesPoints+FirstGrowth+Pomerol+VintageSuperstar+CultWine,data=Bordeaux)
mmpslider(m,ParkerPoints)
detach(Bordeaux)

data(MichelinNY)
attach(MichelinNY)
m <- glm(InMichelin~Food+Decor+Service+Cost+log(Cost)+Food:Decor,family=binomial(),data=MichelinNY)
mmpslider(m)
mmpslider(m,Food,bw=c(1,3,10))
```

```
mmpslider(m,Decor,bw=c(1,3,10))
detach(MichelinNY)

## End(Not run)
```

Index