Package ‘qrmix’

May 3, 2017

Type Package
Title Quantile Regression Mixture Models
Version 0.9.0
Date 2017-04-24
Author Maria de los Angeles Resa, Birol Emir, Javier Cabrera
Maintainer Maria de los Angeles Resa <maria@stat.columbia.edu>
Description Implements the robust algorithm for fitting finite mixture models based on quantile regression proposed by Emir et al., 2017 (unpublished).
License LGPL
LazyData TRUE
Imports MASS, quantreg
RoxygenNote 6.0.1
NeedsCompilation no
Repository CRAN
Date/Publication 2017-05-03 21:49:24 UTC

R topics documented:

- Bisquare ......................................................... 2
- blood.pressure .................................................. 2
- Huber ............................................................. 3
- plot.qrmix ......................................................... 4
- predict.qrmix .................................................... 5
- qrmix ............................................................. 6
- summary.qrmix .................................................. 7

Index 9
Bisquare

Tukey's Bisquare Loss

Description

"Bisquare" evaluates Tukey's Bisquare function defined as

\[
f(r) = \begin{cases} 
1 - (1 - (\frac{r}{c})^2)^3 & |r| \leq c \\
1 & |r| > c 
\end{cases}
\]

Usage

Bisquare(r, c = 4.685)

Arguments

- \(r\) a real number or vector.
- \(c\) a positive number. If the value is negative, its absolute value will be used.

Examples

```r
set.seed(1)
x = rnorm(200, mean = 3)
y = Bisquare(x)
plot(x, y)
```

Blood Pressure Data for qrmix

Description

Simulated blood pressure data created for usage in qrmix examples.

Usage

blood.pressure
**Huber**

**Format**

A data frame with 500 observations on the following 7 variables.

bmi a numeric vector referring to body mass index
age a numeric vector
systolic a numeric vector referring to systolic blood pressure
diastolic a numeric vector referring to diastolic blood pressure
gender a factor with levels female and male
race a factor with levels white, black, and other
smoking a factor with levels yes and no

**Note**

This data does not include any real patient information.

<table>
<thead>
<tr>
<th>Huber</th>
<th>Huber Loss</th>
</tr>
</thead>
</table>

**Description**

Evaluates the Huber loss function defined as

\[ f(r) = \begin{cases} 
\frac{1}{2} |r|^2 & |r| \leq c \\
\frac{c}{2}(|r| - \frac{1}{2}c) & |r| > c 
\end{cases} \]

**Usage**

\[ \text{Huber}(r, c = 1.345) \]

**Arguments**

- \( r \) a real number or vector.
- \( c \) a positive number. If the value is negative, its absolute value will be used.

**Examples**

```
set.seed(1)
x = rnorm(200, mean = 1)
y = Huber(x)
plot(x, y)
abline(h = (1.345)^2/2)
```
plot.qrmix

Plot Method for a qrmix Object

Description

Three types of plots (chosen with type) are currently available: density of the response variable by cluster, plots of the response variable against each covariate included in the model (scatterplots with the k fitted lines for continues variables and boxplots by cluster for the categorical variables), and boxplots of the residuals by cluster.

Usage

## S3 method for class 'qrmix'
plot(x, data = NULL, type = c(1, 2, 3), lwd = 2, bw = "SJ", adjust = 2, ...)

Arguments

x
  a fitted object of class "qrmix".

data
  the data used to fit the model object. It is only necessary when the parameter xy was set to FALSE when fitting the qrmix model.

type
  a numeric vector with values chosen from 1:3 to specify a subset of types of plots required.

lwd
  the line width for the first type of plot (density plot), a positive number. If a negative number is given, lwd = 1 will be used instead. See par.

bw
  the smoothing bandwidth to be used to obtain the density for the first type of plot. See density for details.

adjust
  the bandwidth used is adjust*bw. See density for details.

...
  other arguments passed to other methods.

Examples

data(blood.pressure)

# qrmix model using default function values:
mod1 = qrmix(bmi ~ ., data = blood.pressure, k = 3)
plot(mod1)
plot(mod1, type = c(1, 3), lwd = 1)
predict.qrmix

**Predict Method for qrmix Fits**

**Description**

Obtains clusters, predictions, or residuals from a fitted qrmix object.

**Usage**

```r
## S3 method for class 'qrmix'
predict(object, newdata = NULL, type = "clusters", ...)
```

**Arguments**

- `object`: a fitted object of class 'qrmix'.
- `newdata`: optional data frame for which clusters, predictions, or residuals will be obtained from the qrmix fitted object. If omitted, the training values will be used.
- `type`: the type of prediction. `type = "clusters"` (default value) for predicted clusters, "yhat" for the response predicted value corresponding to the predicted cluster, "residuals" for the residuals corresponding to the response predicted values.
- `...`: other arguments passed to other methods.

**Value**

A vector with predicted clusters, responses, or residuals, depending on `type`.

**Examples**

```r
data(blood.pressure)
set.seed(8)
sampleInd = sort(sample(1:500, 400))
bpSample1 = blood.pressure[sampleInd,]
bpSample2 = blood.pressure[-sampleInd,]

mod1 = qrmix(bmi ~ ., data = bpSample1, k = 3)

# Cluster assigned to the training values
predict(mod1)

# Residuals corresponding to the response predicted values from mod1 for new data
predict(mod1, newdata = bpSample2, type = "residuals")
```
qrmix

Quantile Regression Classification

Description

qrmix estimates the components of a finite mixture model by using quantile regression to select a group of quantiles that satisfy an optimality criteria chosen by the user.

Usage

qrmix(formula, data, k, Ntau=50, alpha=0.03, lossFn="Squared", fitMethod="lm", xy=TRUE, ...)

Arguments

- **formula**: an object of class "formula".
- **data**: an optional data frame that contains the variables in formula.
- **k**: number of clusters.
- **Ntau**: an optional value that indicates the number of quantiles that will be considered for quantile regression comparison. Ntau should be greater or equal than 2k.
- **alpha**: an optional value that will determine the minimum separation between the k quantiles that represent each of the k clusters. alpha should be smaller than \(\frac{1}{2k}\).
- **lossFn**: the loss function to be used to select the best combination of k quantiles. The available functions are "Squared", "Absolute", "Bisquare", and "Huber".
- **fitMethod**: the method to be used for the final fitting. Use "lm" for OLS (default), "rlm" for robust regression, and "rq" to use fit from quantile regression.
- **xy**: logical. If TRUE (the default), the data will be saved in the qrmix object.
- **...**: additional arguments to be passed to the function determined in fitMethod.

Details

The optimality criteria is determined by the lossFn parameter. If, for example, the default value is used (lossFn = "Squared"), the k quantiles selected will minimize the sum of squared residuals. Use "Bisquare" or "Huber" to make the method less sensitive to outliers.

Value

qrmix returns an object of class "qrmix"

- **coefficients**: a matrix with k columns that represent the coefficients for each cluster.
- **clusters**: cluster assignment for each observation.
- **quantiles**: the set of k quantiles that minimize the mean loss.
- **residuals**: the residuals, response minus fitted values.
- **fitted.values**: the fitted values.
- **call**: the matched call.
- **xy**: the data used if xy is set to TRUE.
References


Examples

data(blood.pressure)

# qrmix model using default function values:
mod1 = qrmix(bmi ~ ., data = blood.pressure, k = 3)
summary(mod1)

# qrmix model using Bisquare loss function and refitted with robust regression:
mod2 = qrmix(bmi ~ age + systolic + diastolic + gender, data = blood.pressure, k = 3,
Ntau = 25, alpha = 0.1, lossFn = "Bisquare", fitMethod = "rlm")
summary(mod2)

summary.qrmix

Summarizing qrmix Fits

Description

summary method for class "qrmix"

Usage

## S3 method for class 'qrmix'
summary(object, fitMethod=NULL, data=NULL, ...)

Arguments

object an object of class "qrmix".
fitMethod an optional refitting method if the user wants a method different than the one used to obtain "object" Use "lm" for OLS, "rlm" for robust regression, and "rq" to use fit from quantile regression.
data data used to fit object if it is not contained in object.
... other arguments passed to other methods.

Value

residuals the residuals, response minus fitted values.
clusters cluster assignment for each observation.
call the matched call.
fitMethod    the fitting method used to obtain residuals and clusters.
quantiles the set of k quantiles that minimize the mean loss.
clusters# generic summary from function fitMethod for data in cluster #.

Examples

```r
data(blood.pressure)

# qrmix model using default function values:
mod1 = qrmix(bmi ~ ., data = blood.pressure, k = 3)

# summary using fitMethod = "rlm" instead of the one used when fitting the model mod1
summary1 = summary(mod1, fitMethod = "rlm")

# Are the quantiles selected in this case the same as in the original model?
summary1$quantiles
mod1$quantiles
```
Index

*Topic **datasets**
  - blood.pressure, 2

  Bisquare, 2
  blood.pressure, 2

  density, 4

  Huber, 3

  par, 4
  plot.qrmix, 4
  predict.qrmix, 5

  qrmix, 6

  summary.qrmix, 7