Package ‘FastRCS’

March 6, 2024

Type   Package
Title  Fits the FastRCS Robust Multivariable Linear Regression Model
Version 0.0.9
Date   2024-03-13
Depends R (>= 3.1.1), matrixStats
Suggests mvtnorm
LinkingTo Rcpp, RcppEigen
Description The FastRCS algorithm of Vakili and Schmitt (2014) for robust fit of the multivariable linear regression model and outliers detection.
License GPL (>= 2)
LazyLoad yes
LazyData yes
Author Kaveh Vakili [aut, cre]
Maintainer Kaveh Vakili <vakili.kaveh.email@gmail.com>
NeedsCompilation yes
Repository CRAN
Date/Publication 2024-03-06 15:30:06 UTC

R topics documented:

FastRCS-package .................................................. 2
FastRCS ............................................................. 2
FRCSnumStarts ...................................................... 5
Lemons .............................................................. 6
plot.FastRCS ......................................................... 7
quanf ............................................................... 8

Index 9
FastRCS-package

Description

Uses the FastRCS algorithm to compute the RCS outlyingness index of regression.

Details

<table>
<thead>
<tr>
<th>Package: FastRCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type: Package</td>
</tr>
<tr>
<td>Version: 0.0.9</td>
</tr>
<tr>
<td>Date: 2024-02-13</td>
</tr>
<tr>
<td>Suggests: mvtnorm</td>
</tr>
<tr>
<td>License: GPL (&gt;= 2)</td>
</tr>
<tr>
<td>LazyLoad: yes</td>
</tr>
</tbody>
</table>

Index:

<table>
<thead>
<tr>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>FastRCS</td>
</tr>
<tr>
<td>FRCNumStarts</td>
</tr>
<tr>
<td>plot.FastRCS</td>
</tr>
<tr>
<td>quanf</td>
</tr>
</tbody>
</table>

Author(s)

Kaveh Vakili [aut, cre], Maintainer: Kaveh Vakili <vakili.kaveh.email@gmail.com>

References


Description

Computes a fast and robust regression model for a n by p matrix of multivariate continuous regressors and a single dependent variable.
FastRCS

Usage

FastRCS(x,y,nSamp,alpha=0.5,seed=1,intercept=1)

Arguments

x
A numeric n (n>5*p) by p (p>1) matrix or data frame. Should not contain an intercept.

y
A numeric nvector.

nSamp
A positive integer giving the number of resamples required; "nSamp" may not be reached if too many of the p-subsamples, chosen out of the observed vectors, are in a hyperplane. If "nSamp" is omitted, it is calculated so that the probability of getting at least one uncontaminated starting point is always at least 99 percent when there are n/2 outliers.

alpha
numeric parameter controlling the size of the active subsets, i.e., "h=quanf(alpha,n,p)". Allowed values are between 0.5 and 1 and the default is 0.5.

seed
starting value for random generator. A positive integer. Default is seed = 1

intercept
If true, a model with constant term will be estimated; otherwise no constant term will be included. Default is intercept=TRUE.

Details

The current version of FastRCS includes the use of a C-step procedure to improve efficiency (Rousseeuw and van Driessen (1999)). C-steps are taken after the raw subset is found and before reweighting. In experiments, we found that carrying C-Steps starting from the members of $rawBest improves the speed of convergence without increasing the bias of the final estimates. FastRCS is regression and affine equivariant and thus consistent at the elliptical model (Grubel and Rock (1990)).

Value

nSamp
The value of nSamp used.

alpha
The value of alpha used.

obj
The value of the FastRCS objective function (the I-index) obtained for H*.

rawBest
The index of the h observation with smallest outlyingness indexes.

rawDist
The distances of the observations to the model defined by rawBest.

best
The index of the J observation with outlyingness smaller than the rejection threshold.

coefficients
The vector of coefficients of the hyperplane fitted to the members of $rew$best.

fitted.values
the fitted mean values: cbind(1,x)%*%rew$coefficients.

residuals
the residuals, that is response minus fitted values.

rank
the numeric rank of the fitted linear model.

weights
(only for weighted fits) the specified weights.

df.residual
the residual degrees of freedom.

scale
(robust) scale estimate of the reweighted residuals.
Author(s)

Kaveh Vakili

References


Examples

```r
## testing outlier detection
set.seed(123)
n<-100
p<-3
x0<-matrix(rnorm(n*p),nc=p)
y0<-rnorm(n)
z<-c(rep(0,30),rep(1,70))
x0[1:30,]<-matrix(rnorm(30*p,5,1/100),nc=p)
y0[1:30]<-rnorm(30,10,1/100)
ns<-FRCSnumStarts(p=p,eps=0.4);
results<-FastRCS(x=x0,y=y0,alpha=0.5,nSamp=ns)
z[results$best]

## testing outlier detection, different value of alpha
set.seed(123)
n<-100
p<-3
x0<-matrix(rnorm(n*p),nc=p)
y0<-rnorm(n)
z<-c(rep(0,20),rep(1,80))
x0[1:20,]<-matrix(rnorm(20*p,5,1/100),nc=p)
y0[1:20]<-rnorm(20,10,1/100)
ns<-FRCSnumStarts(p=p,eps=0.25);
results<-FastRCS(x=x0,y=y0,alpha=0.75,nSamp=ns)
z[results$best]

#testing exact fit
set.seed(123)
n<-100
p<-3
x0<-matrix(rnorm(n*p),nc=p)
y0<-rep(1,n)
z<-c(rep(0,30),rep(1,70))
x0[1:30,]<-matrix(rnorm(30*p,5,1/100),nc=p)
y0[1:30]<-rnorm(30,10,1/100)
ns<-FRCSnumStarts(p=p,eps=0.4);
```
results<-FastRCS(x=x0,y=y0,alpha=0.5,nSamp=ns,seed=1)
z[results$rawBest]
results$obj

#testing regression equivariance
n<-100
p<-3
x0<-matrix(rnorm(n*(p-1)),nc=p-1)
y0<-rnorm(n)
ns<-FRCSnumStarts(p=p,eps=0.4);
y1<-y0+cbind(1,x0)%*%rep(-1,p)
results1<-FastRCS(y=y0,x=x0,nSamp=ns,seed=1)$coefficients
results2<-FastRCS(y=y1,x=x0,nSamp=ns,seed=1)$coefficients
results1+rep(-1,p)
#should be the same:
results2

FRCSnumStarts

Computes the number of starting p-subsets

Description

Computes the number of starting p-subsets so that the desired probability of selecting at least one clean one is achieved. This is an internal function not intended to be called by the user.

Usage

FRCSnumStarts(p,gamma=0.99,eps=0.5)

Arguments

p 
number of dimensions of the data matrix X.

gamma 
desired probability of having at least one clean starting p-subset.

eps 
suspected contamination rate of the sample.

Value

An integer number of starting p-subsets.

Author(s)

Kaveh Vakili

Examples

FRCSnumStarts(p=3,gamma=0.99,eps=0.4)
**Description**

Sales data for the Chrysler Town & Country.

**Usage**

Lemons

**Format**

- **VehBCost**  Acquisition cost paid for the vehicle at time of purchase.
- **MMRAcquisitionAuctionAveragePrice** Acquisition price for this vehicle in average condition at time of purchase.
- **MMRAcquisitionRetailCleanPrice** Acquisition price for this vehicle in the above Average condition at time of purchase.
- **MMRAcquisitionRetailAveragePrice** Acquisition price for this vehicle in the retail market in average condition at time of purchase.
- **MMRAcquisitionRetailCleanPrice** Acquisition price for this vehicle in the retail market in above average condition at time of purchase.
- **MMRCurrentAuctionAveragePrice** Acquisition price for this vehicle in average condition as of current day.
- **MMRCurrentAuctionCleanPrice** Acquisition price for this vehicle in above condition as of current day.
- **MMRCurrentRetailAveragePrice** Acquisition price for this vehicle on the retail market in average condition as of current day.
- **MMRCurrentRetailCleanPrice** Acquisition price for this vehicle on the retail market in above average condition as of current day.
- **WarrantyCost**  Warranty price (term=36 month and millage=36K).
- **VehOdo**  The vehicle’s odometer reading.

**Examples**

```r
data(Lemons)
alpha<-0.5
p<-ncol(Lemons)
sns<-FRCSnumStarts(p=p,eps=(1-alpha)*4/5)
Fit<-FastRCS(x=Lemons[,-1],y=Lemons[,1],nSamp=ns,seed=1)
plot(Fit)
```
plot.FastRCS

**Robust Diagnostic Plots For FastRCS**

**Description**

Shows the robust Score distances versus robust Orthogonal distances and their respective cutoffs, for the an object of class FastRCS.

**Usage**

```r
## S3 method for class 'FastRCS'
plot(x,col="black",pch=16,...)
```

**Arguments**

- `x` For the `plot()` method, a `FastRCS` object, typically result of `FastRCS`
- `col` A specification for the default plotting color. Vector of values are recycled.
- `pch` Either an integer specifying a symbol or a single character to be used as the default in plotting points. Note that only integers and single-character strings can be set as a graphics parameter. Vector of values are recycled.
- `...` Further arguments passed to the plot function.

**Details**

This function produces the robust standardized, residuals as well as an indicative cut-off (under normal model). This tool is a diagnostic plot for robust regression and can be used used to reveal the outliers.

**See Also**

`FastRCS`

**Examples**

```r
set.seed(123)
n<-100
p<-3
x0<-matrix(rnorm(n*p),nc=p)
y0<-rnorm(n)
z<-c(rep(0,30),rep(1,70))
x0[1:30,]<-matrix(rnorm(30*p,5,1/100),nc=p)
y0[1:30]<-rnorm(30,10,1/100)
ns<-FRCSnumStarts(p=p,eps=0.4);
results<-FastRCS(x=x0,y=y0,alpha=0.5,nSamp=ns)
plot(results)
```
\textit{quanf} \hspace{1cm} \textit{Converts alpha values to h-values}

\textbf{Description}

FastRCS selects the subset of size h that minimizes the I-index criterion. The function \textit{quanf} determines the size of h based on the rate of contamination the user expects is present in the data. This is an internal function not intended to be called by the user.

\textbf{Usage}

\texttt{quanf(n,p,alpha)}

\textbf{Arguments}

- \texttt{n} \hspace{1cm} Number of rows of the data matrix.
- \texttt{p} \hspace{1cm} Number of columns of the data matrix.
- \texttt{alpha} \hspace{1cm} Numeric parameter controlling the size of the active subsets, i.e., "h=quanf(alpha,n,p)". Allowed values are between 0.5 and 1 and the default is 0.5.

\textbf{Value}

An integer number of the size of the starting p-subsets.

\textbf{Author(s)}

Kaveh Vakili

\textbf{Examples}

\texttt{quanf(p=3,n=500,alpha=0.5)}
Index

* datasets
  Lemons, 6
* hplot
  plot.FastRCS, 7
* package
  FastRCS-package, 2
* regression
  FastRCS, 2
  FRCSnumStarts, 5
  plot.FastRCS, 7
  quanf, 8
* robust
  FastRCS, 2
  FRCSnumStarts, 5
  plot.FastRCS, 7
  quanf, 8

FastRCS, 2, 7
FastRCS-package, 2
FRCSnumStarts, 5

Lemons, 6

plot.FastRCS, 7

quanf, 8