Package ‘changepoint.mv’

November 29, 2018

Type Package
Title Changepoint Analysis for Multivariate Time Series
Version 1.0.1
Date 2018-11-29
Author Lawrence Bardwell <l.bardwell@lancaster.ac.uk> [aut], Idris Eckley <i.eckley@lancaster.ac.uk> [ths,ctb], Paul Fearnhead <p.fearnhead@lancaster.ac.uk> [ths,ctb], Daniel Grose <dan.grose@lancaster.ac.uk> [aut,cre]
Maintainer Daniel Grose <dan.grose@lancaster.ac.uk>
Description Detects the Most Recent Changepoints (mrc) for panel data consisting of many related univariate timeseries using the method developed by Bardwell, Fearnhead, Eckley, Smith and Spott (2018) <doi:10.1080/00401706.2018.1438926>.
License GPL (>= 2)
Imports Rcpp (>= 0.12.12), methods, zoo, tbart, Rdpack, ggplot2, reshape2, assertive
LinkingTo Rcpp
Suggests gridExtra
RdMacros Rdpack
RoxygenNote 6.1.0
NeedsCompilation yes
Repository CRAN
Date/Publication 2018-11-29 14:30:06 UTC

R topics documented:

costs .................................................. 2
cpts.mr ................................................. 3
cpts.mv ................................................. 3
cpts.uv ................................................. 4
data.set .............................................. 5
MDL ................................................... 5
mrc ..................................................... 6
Uses the changepoint locations to determine the penalised cost of the segmented data.

Description

For results obtained using \texttt{mrc}, \texttt{costs} calculates the total penalised cost for all segments across all variates for different numbers of most recent change point values (\(p\)). It also calculates the code length \(\log_2 n^p N^p\) and the total of the penalised cost and code length. The result is a data frame containing \(p\), the penalised cost, the code length and the sum of penalised cost and code length. The row in the data frame with the smallest total corresponds to the minimum description length (MDL). See Bardwell, Eckley, Fearnhead and Smith, (2016) for more details about the cost and code length.

Usage

\begin{verbatim}
## S4 method for signature 'changepoint.mv.mrc.class'

\texttt{costs(x)}
\end{verbatim}

Arguments

\begin{itemize}
  \item \texttt{x} An S4 object as returned by \texttt{mrc}.
\end{itemize}

Value

Data frame containing cost information as described in the description section (above).

References


Examples

\begin{verbatim}
library(changepoint.mv)
data(mrcexample)
res<-mrc(mrcexample,pmax=8)
costs(res)
\end{verbatim}
**cpts.mr**

**Most recent changepoint locations.**

**Description**

Returns a list of vectors containing the most recent changepoint locations.

**Usage**

```r
## S4 method for signature 'changepoint.mv.mrc.class'
cpts.mr(x, p = x$pmax)
```

**Arguments**

- `x` An S4 object as returned by `mrc`.
- `p` The number of most recent changepoints locations to be considered. Default value is `p=x$pmax` where `pmax` is the value specified when `mrc` was called.

**Value**

A data frame containing the most recent changepoint locations and the the variates corresponding to those locations.

**Examples**

```r
library(changepoint.mv)
data(mrcexample)
res <- mrc(mrcexample)
cpts.mr(res)
cpts.mr(res, p=2)
```

---

**cpts.mv**

**Multivariate changepoint locations.**

**Description**

Returns a list of vectors containing the multivariate changepoint locations.

**Usage**

```r
## S4 method for signature 'changepoint.mv.mrc.class'
cpts.mv(x, p = NULL)
```

**Examples**

```r
library(changepoint.mv)
data(mrcexample)
res <- mrc(mrcexample)
cpts.mv(res)
cpts.mv(res, p=2)
```
Arguments

- **x**: An S4 object as returned by `mrc`.
- **p**: The number of most recent changepoints locations to be considered. Default value is \( p=x@pmax \) where \( pmax \)

Value

A list of \( N \) vectors containing the multivariate changepoint locations. Each vector corresponds to an individual variate in the data.

Examples

```r
library(changepointNmv)
data(mrcexample)
res <- mrc(mrcexample)
cpts.mv(res)
cpts.mv(res, p = 3)
```

---

cpts.uv

**Univariate changepoint locations.**

Description

Returns a list of vectors containing the univariate changepoint locations.

Usage

```r
## S4 method for signature 'changepoint.mv.mrc.class'
cpts.uv(x)
```

Arguments

- **x**: An S4 object as returned by `mrc`.

Value

A list of \( N \) vectors containing the univariate changepoint locations. Each vector corresponds to an individual variate in the data.

Examples

```r
library(changepoint.mv)
data(mrcexample)
res <- mrc(mrcexample)
cpts.uv(res)
```
data.set

Recovers the data from the results of a changepoint analysis.

Description

Recovers the data from the S4 class returned by mrc. The data is stored in a data frame and, unless indexed = TRUE was specified when the data was analysed, the first column will contain an index variable. If the original data did not have column names, default ones of the form V.n, where n is the column number, will be added.

Usage

```r
## S4 method for signature 'changepoint.mv.mrc.class'
data.set(x)
```

Arguments

- `x` An S4 class instance obtained from mrc.

Value

A data frame containing the original data and an index variable (if indexed = TRUE was used for the analysis).

Examples

```r
library(changepoint.mv)
data(mrcexample)
res <- mrc(mrcexample)
head(data.set(res))
```

MDL

Calculates the Minimum Description Length.

Description

Calculates the Minimum Description Length (MDL) using the result obtained from mrc. The MDL indicates how many most recent changepoints there are for the data. For a full definition of the MDL and a description of its calculation see Bardwell, Eckley, Fearnhead and Smith, (2016).

Usage

```r
## S4 method for signature 'changepoint.mv.mrc.class'
MDL(x)
```
Arguments

An S4 object as returned by \texttt{mrc}.

Value

The Minimum Description Length (MDL).

References


See Also

\texttt{mrc}

Examples

library(changepoint.nmv)
data(mrcexample)
res<-mrc(mrcexample,pmax=8)
MDL(res)

\texttt{mrc} \hspace{1cm} \textit{Most Recent Changepoints.}

Description

Detects the Most Recent Changepoints (mrc) for panel data consisting of many related univariate timeseries (Bardwell, Eckley, Fearnhead and Smith, 2016). The method first determines the most recent univariate changepoints using PELT (Killick, Fearnhead and Eckley 2012) and then pools information across the time-series by solving the K-median problem using \texttt{tb.raw} (Teitz and Bart, 1968).

Usage

\texttt{mrc(data, cost = "mean", alpha = function(n) 1.5 * log(n), pmax = 5,}
\indent \hspace{1cm} \texttt{indexed = FALSE, mad = FALSE, phi = 0)}

Arguments

\begin{itemize}
\item \texttt{data} \hspace{1cm} An $n \times N$ matrix or data frame representing a length $n$ time series containing observations of $N$ variables. The data can contain missing values which are indicated by NA.
\item \texttt{cost} \hspace{1cm} A string indicating which cost function to use. Possible choices are "mean" (change in mean) or "var" (change in variance). The default value is \texttt{cost="mean"}.
\end{itemize}
alpha

The variable-specific penalty used to penalise the addition of a given change-point into a given variable. This can be specified as a real positive value or as a function of \( n \). The latter form is used when the data contains missing values which leads to time series of different lengths. Default value \( \alpha = \text{function}(n) \cdot 1.5 \cdot \log(n) \).

pmax

Maximum number of most recent changepoints to search for. Default value \( \text{pmax} = 5 \).

indexed

Boolean value indicating that the first column of data is an index variable. If \( \text{indexed} = \text{FALSE} \) an index variable will automatically be generated. Default value is \( \text{indexed} = \text{FALSE} \).

mad

Boolean value indicating if the variates should be scaled by an estimate of their standard deviation obtained using mean absolute deviation (Zhang, Nancy, Siegmund and David 2007). This is useful in conjunction with \( \text{cost} = \text{"mean"} \) for which unit variance is assumed. Default value is \( \text{mad} = \text{FALSE} \).

phi

Lag 1 autocorrelation to model the temporal dependence in the residuals of the time series assuming a MA(1) process. Default \( \phi = 0 \).

References


Examples

```r
library(changepoint.mv)
data(mrcexample)
res<-mrc(mrcexample,pmax=2)
MDL(res)  # MDL == pmax (possibly under-estimating MDL, retry)
res<-mrc(mrcexample,pmax=6)
MDL(res)  # MDL = 5 (< pmax)
# view the most recent changepoints (corresponding to pmax = 5)
unique(cpts.mr(res,p=5)[,1])
summary(res)  # summary of result
```
Description

A dataset containing time series for 100 variates with 500 data points for each variate. The data contains 5 most recent changepoints (mrc’s) with each mrc being shared independently by 20 variates. The observations prior to the mrcs are IID Gaussian distributed with unit variance and mean \( \mu \sim N(0, 2^2) \). The mean after the mrc is \( \mu \pm 0.5 \) with the sign of the change chosen uniformly at random for each variate. For more details regarding simulated testing of mrc see Bardwell, Eckley, Fearnhead and Smith 2016, pages 5–7.

Usage

data(mrcexample)

Format

A matrix with 100 rows and 500 columns.

References


plot

Visualisation of data and changepoint locations.

Description

Plot methods for S4 objects returned by mrc. The plot produced depends on the type of the S4 object.

For objects produced by mrc a heatmap of the data is displayed along with the location of the univariate changepoints (yellow), most recent univariate changepoints (green), and most recent multivariate changepoint locations (red).

A number of arguments with default values are provided to control aspects of how the data and changepoint locations are displayed. The plot methods return a ggplot2 object which can be modified if required.

Usage

```r
## S4 method for signature 'changepoint.mv.mrc.class,ANY'
plot(x, p = MDL(x), group = FALSE,
     display.variable.names = TRUE, show = TRUE)
```
show

Arguments

- **x**: An S4 class returned by `mrc`.
- **p**: Integer value indicating the number of most recent changepoint locations to consider. The minimum value is 1 and the maximum value is the value of pmax used in the call to `mrc`. Default value is the `p=MDL(x)` (see `MDL` for further details).
- **group**: Logical value used to indicate if the variates that share a most recent changepoint should be grouped together in the plot. Default is `group=FALSE`.
- **display.variable.names**: Logical value. If `display.variable.names=TRUE` then the variable names are displayed in the plot. If there are a large number of variates in the data, then it can be useful disable variable names by setting `display.variable.names=FALSE`. Default is `display.variable.names=TRUE`.
- **show**: Logical value used to indicate if the plot should be displayed (the ggplot object produced by plot is always invisibly returned). Default value is `show=TRUE`.

Value

Invisibly returns a ggplot object.

Examples

```r
## not run:
# visualising most recent changepoints
data(mrcexample)
res<-'mrc(mrcexample[,1:10])
p.1<-'plot(res,p=2)
p.2<-'plot(res,p=5)
p.3<-'plot(res,p=2,group=TRUE)
p.4<-'plot(res,p=5,group=TRUE)
if(require(gridExtra)){
  grid.arrange(p.1,p.2,p.3,p.4)
}
## End(not run)
```

---

**show**

Displays S4 objects produced by `changepoint.mv` methods

Description

Displays S4 object produced by `mrc`. The information produced is the same as that provided by the summary method. The method is used by the S4 system for automatic printing.

Usage

```r
## S4 method for signature 'changepoint.mv.mrc.class'
show(object)
```
Arguments

- **object**: An S4 object produced by \texttt{mrc}.

Examples

```r
library(changepointNmv)
data(mrcexample)
res<-mrc(mrcexample)
# the following lines all produce the same output
res
summary(res)
show(res)
print(res)
```

Description

Produces and prints a summary of the results contained in an S4 class instance produced by \texttt{mrc}. The information produced depends on the analysis type and options used, but typically includes the total penalised cost, penalty values and cost function type.

Usage

```r
## S4 method for signature 'changepoint.mv.class'
summary(object)

## S4 method for signature 'changepoint.mv.mrc.class'
summary(object)
```

Arguments

- **object**: An S4 class instance produced by \texttt{mrc}.
Index

*Topic datasets
  mrcexample, 7

  costs, 2
  costs, changepoint.mv.mrc.class-method
    (costs), 2
  cpts.mr, 3
  cpts.mr, changepoint.mv.mrc.class-method
    (cpts.mr), 3
  cpts.mv, 3
  cpts.mv, changepoint.mv.mrc.class-method
    (cpts.mv), 3
  cpts.uv, 4
  cpts.uv, changepoint.mv.mrc.class-method
    (cpts.uv), 4

  data.set, 5
  data.set, changepoint.mv.mrc.class-method
    (data.set), 5

  MDL, 5, 9
  MDL, changepoint.mv.mrc.class-method
    (MDL), 5
  mrc, 2–6, 6, 8–10
  mrcexample, 7

  plot, 8
  plot, changepoint.mv.mrc.class, ANY-method
    (plot), 8

  show, 9
  show, changepoint.mv.mrc.class-method
    (show), 9
  summary, 10
  summary, changepoint.mv.mrc.class-method
    (summary), 10
  summary, changepoint.mv.mrc.class-method
    (summary), 10

  tb.raw, 6