

Package ‘changepoint.mv’

August 7, 2020

Type Package

Title Changepoint Analysis for Multivariate Time Series

Version 1.0.2

Date 2020-08-05

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 7.1.0

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-08-07 10:00:03 UTC

R topics documented:

costs	2
cpts.mr	3
cpts.mv	3
cpts.uv	4
data.set	5
MDL	5
mrc	6

mrceexample	8
plot	8
show	9
summary	10

Index	11
--------------	-----------

costs	<i>Uses the changepoint locations to determine the penalised cost of the segmented data.</i>
-------	--

Description

For results obtained using `mrc`, `costs` calculates the 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 $\log_2 n^p p^N$ 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

```
## S4 method for signature 'changepoint.mv.mrc.class'
costs(x)
```

Arguments

`x` An S4 object as returned by `mrc`.

Value

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

References

Bardwell L, Fearnhead P, Eckley IA, Smith S, Spott M (2018). “Most Recent Changepoint Detection in Panel Data.” *Technometrics*, 1-11. doi: [10.1080/00401706.2018.1438926](https://doi.org/10.1080/00401706.2018.1438926).

Examples

```
library(changepoint.mv)
data(mrceexample)
res<-mrc(mrceexample, pmax=8)
costs(res)
```

cpts.mr	<i>Most recent changepoint locations.</i>
---------	---

Description

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

Usage

```
## 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

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

cpts.mv	<i>Multivariate changepoint locations.</i>
---------	--

Description

Returns a list of vectors containing the multivariate changepoint locations.

Usage

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

Arguments

x	An S4 object as returned by <code>mrc</code> .
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

```
library(changepoint.mv)
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

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

Arguments

x	An S4 object as returned by <code>mrc</code> .
---	--

Value

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

Examples

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

data.set	<i>Recovers the data from the results of a changepoint analysis.</i>
----------	--

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

```
## 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

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

MDL	<i>Calculates the Minimum Description Length.</i>
-----	---

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

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

Arguments

x An S4 object as returned by [mrc](#).

Value

The Minimum Description Length (MDL).

References

Bardwell L, Fearnhead P, Eckley IA, Smith S, Spott M (2018). “Most Recent Changepoint Detection in Panel Data.” *Technometrics*, 1-11. doi: [10.1080/00401706.2018.1438926](https://doi.org/10.1080/00401706.2018.1438926).

See Also

[mrc](#)

Examples

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

mrc

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 [tb.raw](#) (Teitz and Bart, 1968).

Usage

```
mrc(
  data,
  cost = "mean",
  alpha = function(n) 1.5 * log(n),
  pmax = 5,
  indexed = FALSE,
  mad = FALSE,
  phi = 0
)
```

Arguments

data	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.
cost	A string indicating which cost function to use. Possible choices are "mean" (change in mean) or "var" (change in variance). The default value is cost="mean".
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=function(n) 1.5*log(n).
pmax	Maximum number of most recent changepts to search for. Default value pmax=5.
indexed	Boolean value indicating that the first column of data is an index variable. If indexed=FALSE an index variable will automatically be generated. Default value is indexed=FALSE.
mad	Boolean value indicating if the variates should be scaled by an estimate of the their standard deviation obtained using mean absolute deviation (Zhang, Nancy, Siegmund and David 2007). This is useful in conjunction with cost="mean" for which unit variance is assumed. Default value is mad=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

- Bardwell L, Fearnhead P, Eckley IA, Smith S, Spott M (2018). "Most Recent Change-point Detection in Panel Data." *Technometrics*, 1-11. doi: [10.1080/00401706.2018.1438926](https://doi.org/10.1080/00401706.2018.1438926).
- Teitz, M. B., Bart P. (1968). "Heuristic methods for estimating generalized vertex median of a weighted graph." *Operations Research*, **16**, 995–961.
- Killick R, Fearnhead P, Eckley IA (2012). "Optimal Detection of Changepts With a Linear Computational Cost." *Journal of the American Statistical Association*, **107**(500), 1590-1598. doi: [10.1080/01621459.2012.737745](https://doi.org/10.1080/01621459.2012.737745).

Examples

```
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 changepts (corresponding to pmax = 5)
unique(cpts.mr(res,p=5)[,1])
summary(res) # summary of result
```

 mrcexample

Simulated data.

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

Bardwell L, Fearnhead P, Eckley IA, Smith S, Spott M (2018). “Most Recent Changepoint Detection in Panel Data.” *Technometrics*, 1-11. doi: [10.1080/00401706.2018.1438926](https://doi.org/10.1080/00401706.2018.1438926).

 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

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


Arguments

<code>x</code>	An S4 class returned by <code>mrc</code> .
<code>p</code>	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 <code>pmax</code> used in the call to <code>mrc</code> . Default value is the <code>p=MDL(x)</code> (see <code>MDL</code> for further details).
<code>group</code>	Logical value used to indicate if the variates that share a most recent changepoint should be grouped together in the plot. Default is <code>group=FALSE</code> .
<code>display.variable.names</code>	Logical value. If <code>display.variable.names=TRUE</code> 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 to disable variable names by setting <code>display.variable.names=FALSE</code> . Default is <code>display.variable.names=TRUE</code> .
<code>show</code>	Logical value used to indicate if the plot should be displayed (the <code>ggplot</code> object produced by <code>plot</code> is always invisibly returned). Default value is <code>show=TRUE</code> .

Value

Invisibly returns a `ggplot` object.

Examples

```
## 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

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

Arguments

object An S4 object produced by `mrc`.

Examples

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

summary

Summary

Description

Produces and prints a summary of the results contained in an S4 class instance produced by `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

```
## 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 `mrc`.

Index

* datasets

- mrceexample, 8

- 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
- mrceexample, 8

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

- tb.raw, 6