Package: anomalous (via r-universe)

September 6, 2024

Type Package

Title What the package does (short line)

Version 0.0.4.2

Date 2022-11-07

Author Who wrote it

Maintainer Paul Smith <paul@waternumbers.co.uk>

Description More about what it does (maybe more than one line).

License GPL-3

Encoding UTF-8

RoxygenNote 7.3.1

Suggests tinytest, robustbase, knitr, markdown

Imports R6, Rcpp

LinkingTo Rcpp

VignetteBuilder knitr

Depends R (>= 4.0.0)

Repository https://waternumbers.r-universe.dev

RemoteUrl https://github.com/waternumbers/anomalous

RemoteRef performance

RemoteSha 543b90b3fdb90e0f88c6d0a44172cbe0f06526e0

Contents

---

capa   *An R implimentation of the segmented search algorithmpelt algorithm*

Description

An R implimentation of the segmented search algorithmpelt algorithm

Usage

```
capa(part, fCost, prune = TRUE, verbose = FALSE)
```

Arguments

part	the starting partition
fCost	the cost function
prune	logical, should pruning be used
verbose	logical, print out progress

Details

Basic R implimentation of pelt - not efficent

Value

the optimal partition

---

crops   *An implimentation of the CROPS algorithm in 1D*

Description

An implimentation of the CROPS algorithm in 1D

Usage

```
crops(
  betaMin,
  betaMax,
  fCost,
  alg = pelt,
  betaP = Inf,
  min_length = 2,
  prune = TRUE,
  verbose = FALSE,
  maxIter = 100
)
```

Arguments

`betaMin`	lower bound of penalisation window
`betaMax`	upper bound of penalisation window
`fCost`	the cost function
`alg`	algorithm either capa of pelt
`betaP`	penalty for adding a point anomaly - only for use with capa
`min_length`	minimum number of values in a collective anomaly
`prune`	logical, should pruning be used
`verbose`	logical, print out progress
`maxIter`	maximum number of algorithm evaluations to perform

Details

This will only work for cost functions where the beta is additive!!!

Value

something...

Lai2005fig4	*Normalized glioblastoma profile for an excerpt of chromosome 7, the EGFR locus.*

Description

This dataset is taken from Lai W, Johnson MJ, Kucherlapati R, Park PJ, Bioinformatics , 2005. The paper states that the original source of the data is from Bredel et al. (2005). The data is an excerpt of chromosome 7 in GBM29 from 40 to 65 Mb.

This version of the data is a copy of that in the changepoint package.

Usage

```
data(Lai2005fig4)
```

Format

A matrix of dimensions 193 x 5. The columns are Spot, CH, POS.start, POS.end, GBM31.

Source

http://compbio.med.harvard.edu/Supplements/Bioinformatics05b/Profiles/Chrom_7_from40_to65Mb_GBM29.xls

---

machinetemp *Machine temperature data.*

Description

Temperature sensor data of an internal component of a large, industrial machine. The data contains three known anomalies. The first anomaly is a planned shutdown of the machine. The second anomaly is difficult to detect and directly led to the third anomaly, a catastrophic failure of the machine. The data consists of 22695 observations of machine temperature recorded at 5 minute intervals along with the date and time of the measurement. The data was obtained from the Numenta Anomaly Benchmark, which can be found at https://github.com/numenta/NAB.

Usage

```
data(machinetemp)
```

Format

A dataframe with 22695 rows and 2 columns. The first column contains the date and time of the temperature measurement. The second column contains the machine temperature.

---

param *Get the parameters for a partitioning result*

Description

Get the parameters for a partitioning result

Usage

```
param(res, fCost)
```

Arguments

res   the result of a partitioning algorithm

fCost   the cost function

Details

Not yet implimented for all cost functions

Value

list of parameters

---

pelt   *An R implimentation of the segmented search algorithmpelt algorithm*

Description

An R implimentation of the segmented search algorithmpelt algorithm

Usage

```
pelt(part, fCost, prune = TRUE, verbose = FALSE)
```

Arguments

part   the starting partition

fCost   the cost function

prune   logical, should pruning be used

verbose   logical, print out progress

Details

Basic R implimentation of pelt - not efficent

Value

the optimal partition

---

sim.data   *Simulated data.*

Description

A simulated data set for use in the examples and vignettes. The data consists of 500 observations on 20 variates drawn from the standard normal distribution. Within the data there are three multivariate anomalies of length 15 located at t=100, t=200, and t=300 for which the mean changes from 0 to 2. The anomalies affect variates 1 to 8, 1 to 12 and 1 to 16 respectively.

Usage

```
data(simulated)
```

Format

A matrix with 500 rows and 40 columns.

---

wind   *Ireland wind data, 1961-1978*

Description

Daily average wind speeds for 1961-1978 at 12 synoptic meteorological stations in the Republic of Ireland (Haslett and raftery 1989). Wind speeds are in knots (1 knot = 0.5418 m/s), at each of the stations in the order given in Fig.4 of Haslett and Raftery (1989, see below)

This data is a copy of that contained within the gstat package

Usage

```
data(wind)
```

Format

data.frame `wind` contains the following columns:

year year, minus 1900

month month (number) of the year

day day

RPT average wind speed in knots at station RPT

VAL average wind speed in knots at station VAL

ROS average wind speed in knots at station ROS

KIL average wind speed in knots at station KIL

SHA average wind speed in knots at station SHA

BIR average wind speed in knots at station BIR

DUB average wind speed in knots at station DUB

CLA average wind speed in knots at station CLA

MUL average wind speed in knots at station MUL

CLO average wind speed in knots at station CLO

BEL average wind speed in knots at station BEL

MAL average wind speed in knots at station MAL

data.frame `wind.loc` contains the following columns:

Station Station name

Code Station code

Latitude Latitude, in DMS, see examples below

Longitude Longitude, in DMS, see examples below

MeanWind mean wind for each station, metres per second

Note

This data set comes with the following message: "Be aware that the dataset is 532494 bytes long (thats over half a Megabyte). Please be sure you want the data before you request it." The data were obtained on Oct 12, 2008, from: http://www.stat.washington.edu/raftery/software.html The data are also available from statlib. Locations of 11 of the stations (ROS, Rosslare has been thrown out because it fits poorly the spatial correlations of the other stations) were obtained from: http://www.stat.washington.edu/research/reports/2005/tr475.pdf Roslare lat/lon was obtained from google maps, location Roslare. The mean wind value for Roslare comes from Fig. 1 in the original paper. Haslett and Raftery proposed to use a sqrt-transform to stabilize the variance.

Author(s)

Adrian Raftery; imported to R by Edzer Pebesma

References

These data were analyzed in detail in the following article:

Haslett, J. and Raftery, A. E. (1989). Space-time Modelling with Long-memory Dependence: Assessing Ireland's Wind Power Resource (with Discussion). Applied Statistics 38, 1-50. and in many later papers on space-time analysis, for example: Tilmann Gneiting, Marc G. Genton, Peter Guttorp: Geostatistical Space-Time Models, Stationarity, Separability and Full symmetry. Ch. 4 in: B. Finkenstaedt, L. Held, V. Isham, Statistical Methods for Spatio-Temporal Systems.

Examples

```
data(wind)
summary(wind)
```

Index