Package 'restlos'

July 2, 2014

Type Package

Title Robust estimation of location and scatter

Version 0.1-2

Date 2013-06-18

Author Steffen Liebscher and Thomas Kirschstein

Maintainer Steffen Liebscher <steffen.liebscher@wiwi.uni-halle.de>

Description The package implements some recently introduced
methodology for robust estimation of location and scatter as well as outlier identification.

Depends som, rgl, ade4, geometry, nnclust

License GPL (>= 2)

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2013-06-18 16:16:56

R topics documented:

restlos-package *Robust estimation of location and scatter*

Description

The package implements some recently introduced methodology for robust estimation of location and scatter as well as outlier identification.

Details

Package:	restlos
Type:	Package
Version:	0.1-2
Date:	2013-06-18
License:	GPL (>= 2)
LazyLoad:	yes

Author(s)

Steffen Liebscher and Thomas Kirschstein

Maintainer: Steffen Liebscher <steffen.liebscher@wiwi.uni-halle.de>

References

Kirschstein, T., Liebscher, S., and Becker, C. (2013): Robust estimation of location and scatter by pruning the minimum spanning tree, *Journal of Multivariate Analysis*, 120, 173-184, DOI: 10.1016/j.jmva.2013.05.004.

Liebscher, S., Kirschstein, T., and Becker, C. (2012): The Flood Algorithm - A Multivariate, Self-Organizing-Map-Based, Robust Location and Covariance Estimator, *Statistics and Computing*, 22(1), 325-336, DOI: 10.1007/s11222-011-9250-3.

Liebscher, S., Kirschstein, T., and Becker, C. (2013): RDELA - A Delaunay-Triangulation-based, Location and Covariance Estimator with High Breakdown Point, *Statistics and Computing*, DOI: 10.1007/s11222-012-9337-5.

flood *The Flood Algorithm*

Description

The function determines a robust subsample utilizing self-organizing maps (SOM).

Usage

```
flood(data, Nx=10, Ny=10, rlen=2000)
```

Arguments

data	At least a two-dimensional data matrix is required. Number of observations needs to be greater than number of dimensions.
Nx	Size of the SOM-net in x direction. Default is 10.
Ny	Size of the SOM-net in y direction. Default is 10.
rlen	Number of iterations during SOM learn process. Default is 2000.

Details

The function first calls the som function within the **som**-package. The results are subsequently used to determine a robust subsample. Arguments Nx, Ny and rlen are passed to som. These arguments should be selected depending on the size of the data set (number of observations/dimensions). The larger the data set the larger the net size and the number of iterations should be. Note: At the moment only rectangular and quadratic SOM nets are supported.

Value

som.results	SOM results as delivered by som.
som.neigh	A matrix showing for every neuron (first column) the index off the neighboring neurons (columns 2-5).
umatrix	The U-matrix shows the U-value for every neuron.
winneuron	Vector of length n giving the index of the nearest neuron (Euclidean distance).
lib	List of all basins found. Index of neurons. Smallest subsample of size (n+d+1)/2.
lin	List of all neighboring neurons per basin. Index of neurons. Smallest subsample of size (n+d+1)/2.
geb	Number of associated data points per basin. Smallest subsample of size (n+d+1)/2.
l	Internal value necessary for plotting.
fafh	Data for plotting the flood area flood height curve.
fafh.lib	Internal data necessary for plotting extented flooding.
fafh.drin	Internal data necessary for plotting extented flooding.
drin	Robust subsample of minimal size.

Author(s)

Steffen Liebscher <steffen.liebscher@wiwi.uni-halle.de>

References

Liebscher, S., Kirschstein, T., and Becker, C. (2012): The Flood Algorithm - A Multivariate, Self-Organizing-Map-Based, Robust Location and Covariance Estimator, *Statistics and Computing*, 22(1), 325-336, DOI: 10.1007/s11222-011-9250-3.

Examples

```
#flood(halle)
```

`halle` *halle data set*

Description

Artificial data set containing a total of 1600 observations in two dimensions (in three groups with 100, 1000, 500 obs. from top to bottom). Central cluster is quadratically transformed leading to a U-shaped main part of the data.

Usage

```
halle
```

Examples

```
#plot(halle)
```

`plot.flood` *Plot function for objects of class* `flood`

Description

Function to plot the results obtained by function `flood`

Usage

```
## S3 method for class 'flood'
plot(x,..., level = 0)
```

Arguments

`x`	Object of class `flood`.
`level`	Flood level. Numeric value between 0 and 1. Default is 0 (i.e. all plots are based on the smallest robust subsample).
`...`	Further graphical parameters.

Details

The resulting plots depend on the dimensionality of the data set. For `d=2` and `d=3` the data set and the superimposed SOM net are plotted. For `d>3` a Mahalanobis distance plot is generated instead. The U-landscape and the Flood-Area-Flood-Height-curve are always plotted.

Note

At the moment no additional graphical parameters can be passed.

Author(s)

Steffen Liebscher <steffen.liebscher@wiwi.uni-halle.de>

References

Liebscher, S., Kirschstein, T., and Becker, C. (2012): The Flood Algorithm - A Multivariate, Self-Organizing-Map-Based, Robust Location and Covariance Estimator, *Statistics and Computing*, 22(1), 325-336, DOI: 10.1007/s11222-011-9250-3.

Examples

```
#plot(flood(halle))
```

---

plot.pMST *Plot function for objects of class* pMST

Description

Function to plot the results obtained by the functions pMST.

Usage

```
## S3 method for class 'pMST'
plot(x, ...)
```

Arguments

x	Object of class pMST.
...	Further graphical parameters.

Details

The resulting plots display the LC- and the AL-plot to support the decision on the size of the robust subsample, see references. Moreover, if data set has dimension 2 or 3 the data set is plotted where the chosen robust subset is superimposed as red points.

Author(s)

Thomas Kirschstein <thomas.kirschstein@wiwi.uni-halle.de>

References

Kirschstein, T., Liebscher, S., and Becker, C. (2013): Robust estimation of location and scatter by pruning the minimum spanning tree, *Journal of Multivariate Analysis*, 120, 173-184, DOI: 10.1016/j.jmva.2013.05.004.

Examples

```
#plot(pMST(halle))
```

`plot.rdela` *Plot function for objects of class* `rdela`

Description

Function to plot the results obtained by function `rdela`

Usage

```
## S3 method for class 'rdela'
plot(x,...)
```

Arguments

`x`	Object of class `rdela`.
`...`	Further graphical parameters.

Details

The resulting plots depend on the dimensionality of the data set. For `d=2` and `d=3` the data set and the selected robust subsample are plotted. For `d>3` a Mahalanobis distance plot is generated instead.

Note

At the moment no additional graphical parameters can be passed.

Author(s)

Steffen Liebscher <steffen.liebscher@wiwi.uni-halle.de>

References

Liebscher, S., Kirschstein, T., and Becker, C. (2013): RDELA - A Delaunay-Triangulation-based, Location and Covariance Estimator with High Breakdown Point, *Statistics and Computing*, DOI: 10.1007/s11222-012-9337-5.

Examples

```
#plot(rdela(halle))
```

pMST *pMST (minimum spanning tree pruning)*

Description

The function determines a robust subsample and computes estimates of location and scatter on the subset.

Usage

```
pMST(data, N = floor((dim(data)[1] + dim(data)[2] + 1)/2), lmax = dim(data)[1] * 100)
```

Arguments

`data`	data set to be analyzed, at least a 2-dimensional matrix whose number of rows (i.e. observations `n`) is greater then the number of columns (i.e. dimension `d`)
`N`	Size of the (robust) subsample to be determined. Default is `(n+d+1)/2`
`lmax`	Numerical option: determines the maximal number pruning steps, see deteils.

Details

The function uses the `mstree` function from the **ade4**-package to determine the minimum spanning tree (MST) of the data. The resulting MST is iteratively pruned by deleting edges (starting with the longest edge in the MST) until a connected subset with sufficient size (`N`) remains. Based on the robust subsample location and scatter are estimated.

Value

`loc`	Location estimate based on the robust subsample
`cov`	Covariance estimate based on the robust subsample
`sample`	Index of the observations in the robust subsample
`data`	The input data set

Author(s)

Thomas Kirschstein <thomas.kirschstein@wiwi.uni-halle.de>

References

Kirschstein, T., Liebscher, S., and Becker, C. (2013): Robust estimation of location and scatter by pruning the minimum spanning tree, *Journal of Multivariate Analysis*, 120, 173-184, DOI: 10.1016/j.jmva.2013.05.004.

Examples

```
## Determine subsample of minimal size
#sub<-pMST(halle)
## Determine subsample of size=900
#extsub<-pMST(halle, N=900)
```

---

rdela *The RDELA Algorithm*

Description

The function determines a robust subsample utilizing the Delaunay triangulation.

Usage

```
rdela(data, N)
```

Arguments

data	At least a two-dimensional data matrix is required. Number of observations need to be greater than number of dimensions. No degenerated (i.e. collinear) data sets allowed.
N	Size of the identified subsample. Default is (n+d+1)/2.

Details

The function first calls the delaunayn function within the **geometry**-package. The results are subsequently used to determine a robust subsample.

Value

data	The input data set.
tri	Vertices of all simplices of the Delaunay triangulation. Each row represents a simplex.
neigh	Lists for every simplex the adjacent/neighboring simplices. Each list entry represents a simplex.
radii	Circum-(hypersphere-)radius of each simplex.
center	Center coordinates of all simplices.
LiB	List of all basins found. Index of simplices. Smallest subsample of size (n+d+1)/2.
LiN	List of all neighboring simplices per basin. Index of simplices. Smallest subsample of size (n+d+1)/2.
GeB	Number of associated data points per basin. Smallest subsample of size (n+d+1)/2.
drin	Robust subsample of minimal size.

Author(s)

Steffen Liebscher <steffen.liebscher@wiwi.uni-halle.de>

References

Liebscher, S., Kirschstein, T., and Becker, C. (2013): RDELA - A Delaunay-Triangulation-based, Location and Covariance Estimator with High Breakdown Point, *Statistics and Computing*, DOI: 10.1007/s11222-012-9337-5.

Examples

```
#rdela(halle)
```

Index