Package 'popsom'

July 23, 2015

Version 3.0.1

Date 2015-07-22

Title Routines for Constructing and Evaluating Self-Organizing Maps

Maintainer Lutz Hamel <hamel@cs.uri.edu>

Imports som, fields, graphics, stats, grDevices

Description A set of routines which are useful in constructing
and evaluating self-organizing maps (SOMs).

License GPL

URL http://homepage.cs.uri.edu/faculty/hamel/

NeedsCompilation no

Author Lutz Hamel [aut, cre],
Benjamin Ott [aut],
Gregory Breard [aut]

Repository CRAN

Date/Publication 2015-07-23 15:25:01

R topics documented:

`map.accuracy` *Estimated Topographical Accuracy*

Description

Evaluate the topological quality of a SOM using the estimated topographical accuracy.

Usage

```
map.accuracy(map, k=50, conf.int = 0.95, verb=FALSE)
```

Arguments

`map`	an object of type 'map'.
`k`	number of samples to use in the computation of the estimated topographical accuracy (default=50)
`conf.int`	the confidence interval of the estimated topographical accuracy (default 95 percent).
`verb`	a switch controlling the structure of the output value (default=FALSE)

Value

A list with three components: 1) the value of the estimated topographical accuracy 2) the low value of the confidence interval 'conf.int' 3) the high value of the confidence interval. If verb=TRUE then map.accuracy will return a vector with the accuracies of the individual k samples.

Author(s)

Lutz Hamel

References

"SOM Quality Measures: A Statistical Approach," Lutz Hamel, in preparation, 2015.

Examples

```
data(iris)

## set data frame and labels
df <- subset(iris,select=-Species)
labels <- subset(iris,select=Species)

## build a map
m <- map.build(df, labels, xdim=15, ydim=10, train=1000)

## display estimated topographical accuracy of the map
map.accuracy(m)
```

`map.build` *Build Map*

Description

Constructs a SOM, returns an object of class 'map'.

Usage

```
map.build(data, labels=NULL, xdim = 10, ydim = 5, alpha = 0.6, train = 1000)
```

Arguments

`data`	a dataframe where each row contains an unlabeled training instance.
`labels`	a vector or dataframe with one label for each observation in data.
`xdim`	the x-dimension of the map.
`ydim`	the y-dimension of the map.
`alpha`	the learning rate, should be a positive non-zero real number.
`train`	the number of training iterations.

Value

object of type 'map'.

Note

If your training data does not have any labels you can construct a simple label vector as follows: labels <- 1:nrow(training.data). If you let the labels default to the NULL value then no labels will be shown in the map visualization.

Author(s)

Lutz Hamel, Benjamin Ott, Gregory Breard

Examples

```
data(iris)

## set data frame and labels
df <- subset(iris,select=-Species)
labels <- subset(iris,select=Species)

## build a map
m <- map.build(df, labels, xdim=15, ydim=10, train=1000)
```

`map.convergence` *Evaluate Map Convergence*

Description

Evaluates the convergence of a map using a two-sample statistical test.

Usage

```
map.convergence(map, conf.int = 0.95, verb=FALSE)
```

Arguments

`map`	an object of type 'map'.
`conf.int`	the confidence interval of the convergence test (default 95 percent).
`verb`	a switch controlling the structure of the output value (default=FALSE)

Value

The convergence index of the map. If the switch verb=TRUE then a vector of the individual feature convergence indices is returned.

Note

The convergence index is the variance of the trainig data captured by the map. Maps with low convergence index are typically not trustworthy. However, the precise cut-off depends on the noise level in your training data.

Author(s)

Lutz Hamel, Benjamin Ott, Gregory Breard

References

"A Population Based Convergence Criterion for Self-Organizing Maps," Lutz Hamel and Benjamin Ott. Proceeding of the 2012 International Conference on Data Mining (DMIN'12), pp98-104, July 16-19, 2012, Las Vegas Nevada, USA.

Examples

```
data(iris)

## set data frame and labels
df <- subset(iris,select=-Species)
labels <- subset(iris,select=Species)

## build a map
m <- map.build(df, labels, xdim=15, ydim=10, train=1000)
```

```
## display the convergence index of the map
map.convergence(m)

## display the convergence indices of the individual features
data.frame(names(df),map.convergence(m,verb=TRUE))
```

`map.feature` *Generate Enhanced Unified Distance Matrix For Feature*

Description

Computes and displays the enhanced unified distance matrix for a feature of the training data.

Usage

```
map.feature(map, feature, explicit = FALSE, smoothing = 2)
```

Arguments

`map`	an object of type 'map'.
`feature`	an integer as the index of the feature.
`explicit`	controls the shape of the connected components.
`smoothing`	controls the smoothing level of the umat (NULL, 0, >0).

Note

This essentially constructs a horizontal slice of the UMAT.

Author(s)

Lutz Hamel, Benjamin Ott, Gregory Breard

Examples

```
data(iris)

## set data frame and labels
df <- subset(iris,select=-Species)
labels <- subset(iris,select=Species)

## build a map
m <- map.build(df, labels, xdim=15, ydim=10, train=1000)

## display the umat for the first feature of the map
map.feature(m, 1)
```

map.projection *Map Projection*

Description

Prints the association of labels with map elements.

Usage

```
map.projection(map)
```

Arguments

map	an object of type 'map'.

Value

a dataframe containing the projection onto the map for each observation.

Author(s)

Lutz Hamel, Benjamin Ott, Gregory Breard

Examples

```
data(iris)

## set data frame and labels
df <- subset(iris,select=-Species)
labels <- subset(iris,select=Species)

## build a map
m <- map.build(df, labels, xdim=15, ydim=10, train=1000)

## display the label association for the map
map.projection(m)
```

map.quality *SOM Quality Assessment*

Description

Evaluate the quality of a SOM using the embedding and estimated topographical accuracy.

Usage

```
map.quality(map,conf.int=.95,k=50)
```

Arguments

map	an object of type 'map'.
conf.int	is the confidence interval of the quality assessment (default 95%)
k	number of samples to use in the computation of the estimated topographical accuracy (default=50)

Value

A pair of values: 1) embedding accuracy 2) estimated topographic accuracy

Author(s)

Lutz Hamel

References

"SOM Quality Measures: A Statistical Approach," Lutz Hamel, in preparation, 2015.

Examples

```
data(iris)

## set data frame and labels
df <- subset(iris,select=-Species)
labels <- subset(iris,select=Species)

## build a map
m <- map.build(df, labels, xdim=15, ydim=10, train=1000)

## map quality
map.quality(m)
```

---

map.significance *Compute Significance Of Features*

Description

Computes the relative significance of each feature and plots it.

Usage

```
map.significance(map, graphics = TRUE, feature.labels = TRUE)
```

Arguments

map	an object of type 'map'.
graphics	a switch that controls whether a plot is generated or not.
feature.labels	a switch to allow the plotting of feature names vs feature indices.

Value

if graphics=FALSE a vector containing the significance for each feature is returned.

Note

We use a Bayesian approach to compute the relative significance of features based on variance.

Author(s)

Lutz Hamel, Benjamin Ott, Gregory Breard

References

"Bayesian Probability Approach to Feature Significance for Infrared Spectra of Bacteria," Lutz Hamel, Chris W. Brown, Applied Spectroscopy, Volume 66, Number 1, 2012.

Examples

```
data(iris)

## set data frame and labels
df <- subset(iris,select=-Species)
labels <- subset(iris,select=Species)

## build a map
m <- map.build(df, labels, xdim=15, ydim=10, train=1000)

## show the relative feature significance for each feature
data.frame(names(df),map.significance(m,graphics=FALSE))

## display the relative feature significance graphically
map.significance(m)
```

---

map.starburst *Generate Starburst For Map*

Description

Computes and displays the starburst representation of the clusters for the map.

Usage

```
map.starburst(map, explicit = FALSE, smoothing = 2)
```

Arguments

map	an object of type 'map'.
explicit	controls the shape of the connected components.
smoothing	controls the smoothing level of the umat (NULL, 0, >0).

Note

We apply a gradient approach to the UMAT in order to compute the starburst visualization of the clusters.

Author(s)

Lutz Hamel, Benjamin Ott, Gregory Breard

References

"Improved Interpretability of the Unified Distance Matrix with Connected Components," Lutz Hamel and Chris W. Brown. Proceeding of the 7th International Conference on Data Mining (DMIN'11), July 18-21, 2011, Las Vegas Nevada, USA, ISBN: 1-60132-168-6, pp338-343, CSREA Press, 2011.

Examples

```
data(iris)

## set data frame and labels
df <- subset(iris,select=-Species)
labels <- subset(iris,select=Species)

## build a map
m <- map.build(df, labels, xdim=15, ydim=10, train=1000)

## display the starburst for the map
map.starburst(m)
```

---

map.umat *Generate Unified Distance Matrix For Map*

Description

Computes and displays the unified distance matrix for the map.

Usage

```
map.umat(map)
```

Arguments

map	an object of type 'map'.

Author(s)

Lutz Hamel, Benjamin Ott, Gregory Breard

Examples

```
data(iris)

## set data frame and labels
df <- subset(iris,select=-Species)
labels <- subset(iris,select=Species)

## build a map
m <- map.build(df, labels, xdim=15, ydim=10, train=1000)

## display the umat for the map
map.umat(m)
```

Index