Package ‘cwm’

February 19, 2015

Type Package
Title Cluster Weighted Models by EM algorithm
Version 0.0.3
Date 2013-03-26
Author Giorgio Spedicato, Simona C. Minotti
Depends R (>= 2.14), MASS
Imports methods, stats, matlab, permute
Maintainer Giorgio Spedicato <spedicato_giorgio@yahoo.it>
Description This package estimates gaussian cluster weighted linear regressions by EM algorithm.
License GPL (>= 2)
NeedsCompilation no
Repository CRAN
Date/Publication 2014-01-09 07:26:17

R topics documented:

cwm-package .......................................................... 2
bestPermutation ....................................................... 2
betaplasma ............................................................. 4
cwrEm ................................................................. 5
logLik.cwrObj ......................................................... 7
plot.cwrObj .......................................................... 8
predict.cwrObj ....................................................... 9
print.cwrObj ........................................................ 10
stepCwr ............................................................... 11
summary.cwrObj .................................................... 12

Index 14
**cwm-package**

*The package performs cluster weighed modelling assuming normal distribution*

**Description**

It is a R porting of Original Code from Murphy

**Details**

Package: cwm
Type: Package
Version: 0.0.1
Date: 2013-03-17
License: GPL

**Author(s)**

Giorgio A. Spedicato

**References**

Murphy

**Examples**

```r
library(MASS)
data(geyser)
x=geyser[,1]
y=geyser[,2]
cwmExample=cwrEm(x,y,nc=2)
print(cwmExample)
```

**bestPermutation**

*Function to obtain the best permutation for a classification problem*

**Description**

When a classifier is run on a set of 1,2,..., k groups it returns a possible classification schemes, but it does not know the correspondence of original groups and given groups. This function return the permutation of original group versus output group that maximizes the trace of the confusion matrix.
**Usage**

```r
bestPermutation(origClass, inizOutput)
```

**Arguments**

- `origClass` original group identification vector
- `inizOutput` classified group identification vector

**Details**

Program fails if number of original groups differs from identified groups as in `inizOutput`.

**Value**

An object of class `bestPermutation` containing:

- `permutation` Best permutation
- `groups` Classification with respect to best permutation

**Note**

Shall be improved

**Author(s)**

Giorgio Spedicato

**References**

Giorgio Spedicato

**See Also**

`cwrEm`

**Examples**

```r
#non sense example
x=c(1,2,3)
y=c(1,2,3)
bestPermutation(x,y)
```
**Betaplasma dataset**

**Description**
Example dataset

**Usage**
data(betaplasma)

**Format**
A data frame with 315 observations on the following 15 variables.

- id a numeric vector
- age a numeric vector
- sex a factor with levels F M
- smokestat a factor with levels Never Former Current
- bmi a numeric vector
- vituse a factor with levels Often Not often No
- calories a numeric vector
- fat a numeric vector
- fiber a numeric vector
- alcohol a numeric vector
- chol a numeric vector
- betadiet a numeric vector
- retdiet a numeric vector
- betacaro a numeric vector
- retplasma a numeric vector

**Details**
Unknown

**Source**
Unknown

**References**
unknown

**Examples**
data(betaplasma)
**cwrEm**

*Function to estimate Cluster Weighted Regression (CWR) models*

**Description**

This function estimates CWR models via EM algorithms. An object of class cwrObj is returned containing posterior probabilities and group parameters.

**Usage**

```r
cwrEm(x, y, nc, max_iter = 1000, thresh = 0.01, cov_typeX = "full",
cov_typeY = "full", clamp_weights = FALSE, create_init_params = TRUE,
cwrStart = NULL, cov_priorX = NULL, cov_priorY = NULL, verbose = TRUE,
regress = TRUE, clamp_covX = FALSE, clamp_covY = FALSE)
```

**Arguments**

- **x**: X data matrix
- **y**: Y data matrix
- **nc**: Number of clusters
- **max_iter**: Max iterations. Default 1000
- **thresh**: threshold to assess numerical convergence. Default 0.01
- **cov_typeX**: Type of covariance of groups in X space. May be: "full" (default), "spherical", "diagonal"
- **cov_typeY**: Type of covariance of groups in Y space. May be: "full" (default), "spherical", "diagonal"
- **clamp_weights**: Fixed weights
- **create_init_params**: Creates initial parameters
- **cwrStart**: cwrObj to initialize. If autostart -> NULL
- **cov_priorX**: Prior X covariance if not autostart. See cov_typeX
- **cov_priorY**: Prior Y covariance if not autostart. See cov_typeY
- **verbose**: Prints details of estimation process
- **regress**: Regression model. Default TRUE
- **clamp_covX**: Fixed covX matrix.
- **clamp_covY**: Fixed covY matrix.

**Details**

This is the main function to estimate CWR models.
Value

A CWR object with the following component:

- **muX**: Means matrix of X component
- **muY**: Means matrix of X component
- **aic**: AIC of model
- **X**: X matrix
- **Y**: Y matrix
- **SigmaY**: Array containing Y Variances
- **SigmaX**: Array containing X Variances
- **weightsY**: Matrix containing posterior probabilities

Warning

Estimation can be slow. Convergence is not guaranteed.

Note

This is the main function. X and Y may be vectors or matrices. cwrObj objects containing parameters and posterior probabilities are returned.

Author(s)

Giorgio Spedicato

References

Murphy

See Also

stepCwr

Examples

```r
# using Geyser dataset from package MASS
library(MASS)
data(geyser)
x=geyser[,1]
y=geyser[,2]
cwrEmExample=cwrEm(x,y,nc=2)
print(cwrEmExample)
```
Description
This function extracts the logLikelihood.

Usage

```r
### S3 method for class 'cwrObj'
logLik(object, ...)
```

Arguments

- `object` A cwrObj
- `...` Additional data (not yet implemented)

Details
In the future this function will perform log-likelihood calculation directly.

Value
A numeric value

Author(s)
Giorgio A. Spedicato

References
Murphy

See Also
cwrEm

Examples

```r
### Not run:
library(MASS)
data(geyser, package="MASS")
x=geyser[,1]
y=geyser[,2]
ciao=stepCwr(x,y,nc=2)
loglik(ciao)
### End(Not run)
```
plot.cwrObj  

#### Description

Generic S3 plot method for CWR objects. It only works when data dimension is R2.

#### Usage

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

#### Arguments

- `x`: CWR object to plot
- `...`: Optional argument passed to plot method. Use of dots implemented yet.

#### Details

Only if data dimension lies in R2 it works.

#### Value

No value is returned.

#### Note

S3 method.

#### Author(s)

Giorgio Spedicato

#### References

Murphy

#### See Also

cwrEm
predict.cwrObj

Examples

```r
## Not run:
data(geyser, package="MASS")
x=geyser[,1]
y=geyser[,2]
ciao=cwrEm(x,y,nc=2)
plot(ciao)

## End(Not run)
```

---

### predict.cwrObj

S3 predict method for cwrObj

---

**Description**

Method to return predicted group membership

**Usage**

```r
## S3 method for class 'cwrObj'
predict(object, ...)
```

**Arguments**

- **object** A cwr obj
- **...** additional parameters

**Details**

Get the max column index of the matrix

**Value**

A numeric vector

**Author(s)**

Giorgio A. Spedicato

**References**

Murphy

**See Also**

cwrEm
print.cwrObj

S3 print method for CWR objects

Description
This method prints estimation summary values.

Usage

```r
## S3 method for class 'cwrObj'
print(x, ...)  
```

Arguments

- `x` CWR object to be printed
- `...` Further arguments. Not implemented yet.

Details
This is a short summary.

Value
This method returns no value.

Note
S3 method.

Author(s)
Giorgio Spedicato

References
Murphy

See Also
cwrEm, plot.cwrObj
stepCwr

Examples

data(geyser)
x = geyser[,1]
y = geyser[,2]
ciao = cwrEm(x, y, nc = 2)
print(ciao)

Description

This function iterates nIter times a single estimation of CWR models by cwrEm function. Then the one that has best logLikelihood is chosen.

Usage

stepCwr(X, Y, nc, prop = 0.1, nIter = 10, changeTrainingSet = FALSE)

Arguments

- X: X data vector
- Y: Y data vector
- nc: number of clusters.
- prop: Proportion of samples. Default 0.1.
- nIter: Number of iteration. Default 10.
- changeTrainingSet: Boolean. If TRUE the training set is changed.

Details

This function allows the estimation of models where the structure of the data set lies to probable convergence problems.

Value

A cwr object.

Note

Uses try.

Author(s)

Giorgio Spedicato
References

Murphy, Bettina.

See Also
cwrEm

Examples

data(geyser)
x=geyser[,1]
y=geyser[,2]
ciao=stepCwr(x,y,nc=2)

summary.cwrObj

Generic summary S3 method for CWR object.

Description

This function prints out a detailed summary of CWR object.

Usage

### S3 method for class 'cwrObj'
summary(object, ...)

Arguments

object cwrObj
... Further arguments to be passed. Not implemented yet

Details

This function expands output from summary method.

Value

This function returns no value.

Note

Will be converted in S4 method.

Author(s)

Giorgio Spedicato
References
Murphy

See Also
cwrEm, print.cwrObj, plot.cwrObj

Examples
data(geyser)
x=geyser[,1]
y=geyser[,2]
ciao=cwrEm(x,y,nc=2)
summary(ciao)
Index

*Topic datasets
  betaplasma, 4

*Topic methods
  bestPermutation, 2
cwrEm, 5
logLik.cwrObj, 7
plot.cwrObj, 8
predict.cwrObj, 9
print.cwrObj, 10
stepCwr, 11
summary.cwrObj, 12

*Topic package
  cwm-package, 2

bestPermutation, 2
betaplasma, 4

cwm (cwm-package), 2
cwm-package, 2
cwrEm, 3, 5, 7–10, 12, 13

logLik.cwrObj, 7
plot.cwrObj, 8, 10, 13
predict.cwrObj, 9
print.cwrObj, 10, 13

stepCwr, 6, 11
summary.cwrObj, 12