

Package ‘cord’

September 20, 2015

Type Package

Title Community Estimation in G-Models via CORD

Version 0.1.1

Date 2015-09-18

Author Xi (Rossi) LUO, Florentina Bunea, Christophe Giraud

Maintainer Xi (Rossi) LUO <xi.rossi.luo@gmail.com>

Description Partition data points (variables) into communities/clusters, similar to clustering algorithms, such as k-means and hierarchical clustering. This package implements a clustering algorithm based on a new metric CORD, defined for high dimensional parametric or semi-parametric distributions. Read <http://arxiv.org/abs/1508.01939> for more details.

License GPL-3

Suggests pcaPP

Imports Rcpp

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

Repository CRAN

Date/Publication 2015-09-20 08:01:07

R topics documented:

cord	2
Index	3

`cord`*Community estimation in G-models via CORD*

Description

Partition data points (variables) into clusters/communities. Reference: Bunea, F., Giraud, C., & Luo, X. (2015). Community estimation in *G*-models via CORD. arXiv preprint arXiv:1508.01939. <http://arxiv.org/abs/1508.01939>.

Usage

```
cord(X, tau = 2 * sqrt(log(ncol(X))/nrow(X)), kendall = T,  
      input = c("data", "cor", "dist"))
```

Arguments

<code>X</code>	Input data matrix. It should be an n (samples) by p (variables) matrix when <code>input</code> is set to the value "data" by default. It can also be a p by p symmetric matrix when X is a correlation matrix or a distance matrix if <code>input</code> is set accordingly.
<code>tau</code>	Threshold to use at each iteration. A theoretical choice is about $2n^{-1/2} \log^{1/2} p$.
<code>kendall</code>	Whether to compute Kendall's tau correlation matrix from X , when <code>input</code> is set to "data". If FALSE, Pearson's correlation will be computed, usually faster for large p .
<code>input</code>	Type of input X . It should be set to "data" when X is an n (samples) by p (variables) matrix. If X is a correlation matrix or a distance matrix, it should be set to "cor" or "dist" respectively.

Value

list with one element: a vector of integers showing which cluster/community each point is assigned to.

Examples

```
set.seed(100)  
X <- 2*matrix(rnorm(200*2), 200, 10)+matrix(rnorm(200*10), 200, 10)  
cord(X)
```

Index

*Topic **cluster**

cord, [2](#)

*Topic **multivariate**

cord, [2](#)

cord, [2](#)