

Package ‘cord’

October 16, 2025

Title Community Estimation in G-Models via CORD

Version 0.2.0

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

Author Xi (Rossi) LUO [aut, cre],
Florentina Bunea [aut],
Christophe Giraud [aut]

Description Partitions 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 semiparametric distributions. For more details see Bunea et al. (2020), Annals of Statistics <doi:10.1214/18-AOS1794>.

License GPL-3

URL <https://doi.org/10.1214/18-AOS1794>

Encoding UTF-8

Suggests pcaPP

Imports Rcpp

LinkingTo Rcpp, RcppArmadillo

NeedsCompilation yes

RoxygenNote 7.3.3

Repository CRAN

Date/Publication 2025-10-15 22:30:08 UTC

Contents

cord	2
Index	3

cord

*Community estimation in G-models via CORD***Description**

Partition data points (variables) into clusters/communities. Reference: Bunea et al (2020). Model assisted variable clustering: Minimax-optimal recovery and algorithms, Annals of Statistics, [doi:10.1214/18AOS1794](https://doi.org/10.1214/18AOS1794).

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 <code>X</code> 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 <code>X</code> , 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 <code>X</code> . It should be set to "data" when <code>X</code> is an n (samples) by p (variables) matrix. If <code>X</code> 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

- * **cluster**
 - cord, [2](#)
- * **multivariate**
 - cord, [2](#)

cord, [2](#)