Package 'cml'

May 5, 2022

Type Package

Title Conditional Manifold Learning

Version 0.0.5

Author Anh Tuan Bui [aut, cre]

Maintainer Anh Tuan Bui <atbui@u.northwestern.edu>

Imports vegan, MASS

Description
Find a low-dimensional embedding of high-dimensional data, conditioning on auxiliary manifold information. The current version supports conditional MDS and conditional ISOMAP.

License GPL-2

Encoding UTF-8

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2022-05-05 06:40:02 UTC

R topics documented:

---

cml-package   *Conditional Manifold Learning*

Description

Find a low-dimensional embedding of high-dimensional data, conditioning on auxiliary manifold information. The current version supports conditional MDS and conditional ISOMAP.

Please cite this package as follows: Bui, A.T. (2021). Dimension Reduction with Prior Information for Knowledge Discovery. arXiv:2111.13646. URL: https://arxiv.org/abs/2111.13646.

Details

Brief descriptions of the main functions of the package are provided below:

condMDS(): is the conditional MDS method, which uses conditional SMACOF to optimize its conditional stress function.

condIsomap(): is the conditional ISOMAP method, which is basically conditional MDS applying to graph distances (i.e., estimated geodesic distances) of the given distances/dissimilarities.

Author(s)

Anh Tuan Bui

Maintainer: Anh Tuan Bui <atbui@u.northwestern.edu>

References

Bui, A.T. (2021). Dimension Reduction with Prior Information for Knowledge Discovery. arXiv:2111.13646. URL: https://arxiv.org/abs/2111.13646.

Examples

```
# Generate car-brand perception data
factor.weights <- c(90, 88, 83, 82, 81, 70, 68)/562
N <- 100
set.seed(1)
data <- matrix(runif(N*7), N, 7)
colnames(data) <- c('Quality', 'Safety', 'Value','Performance', 'Eco', 'Design', 'Tech')
rownames(data) <- paste('Brand', 1:N)
data.hat <- data + matrix(rnorm(N*7), N, 7)*data*.05
data.weighted <- t(apply(data, 1, function(x) x*factor.weights))
d <- dist(data.weighted)
d.hat <- d + rnorm(length(d))*d*.05

# Conditional MDS, using the first 4 factors as known features
u.cmds = condMDS(d.hat, data.hat[,1:4], 3)
u.cmds$B # compare with diag(factor.weights[1:4])
ccor(data[,5:7], u.cmds$U)$cancor # canonical correlations
```

```
# Conditional ISOMAP, using the first 4 factors as known features
u.cisomap = condIsomap(d.hat, data.hat[,1:4], 3, k = 20)
u.cisomap$B # compare with diag(factor.weights[1:4])
ccor(data[,5:7], u.cisomap$U)$cancor # canonical correlations

# NOTE: Setting method='vector' in the above calls may slightly improve the
# results because the first 4 factors are uncorrelated.
```

---

ccor   *Canonical Correlation*

Description

Find the canonical correlations of two multivariate datasets.

Usage

```
ccor(x, y)
```

Arguments

x	the first dataset.
y	the second dataset.

Value

cancor	canonical correlations.
xcoef	the columns of this matrix contain the coefficients for the first dataset.
ycoef	the columns of this matrix contain the coefficients for the second dataset.

Author(s)

Anh Tuan Bui

Examples

```
ccor(iris[,1:2], iris[,3:4])
```

---

condDist   *Conditional Euclidean distance*

Description

Internal functions for condSmacof_matrix() and condSmacof_vector().

Usage

```
condDist(U, V.tilda, one_n_t=t(rep(1,nrow(U))))
condDist2(U, V.tilda2, one_n_t=t(rep(1,nrow(U))))
```

Arguments

U	the embedding U
V.tilda	= V %*% B
V.tilda2	= V %*% b^2*t(V)
one_n_t	= t(rep(1,nrow(U)))

Value

a dist object.

Author(s)

Anh Tuan Bui

References

Bui, A.T. (2021). Dimension Reduction with Prior Information for Knowledge Discovery. arXiv:2111.13646. URL: https://arxiv.org/abs/2111.13646.

---

condIsomap   *Conditional ISOMAP*

Description

Finds a low-dimensional manifold embedding of a given distance/dissimilarity matrix, conditioning on auxiliary manifold parameters. The method applies conditional MDS (see condMDS) to a graph distance matrix computed for the given distances/dissimilarities, using the isomap{vegan} function.

Usage

```
condIsomap(d, V, u.dim, epsilon = NULL, k, W = NULL,
           method = c('matrix', 'vector'),
           it.max = 1000, gamma = 1e-05,
           init = c('none', 'user'), U.start = NULL,
           B.start = NULL, b.start = NULL, ...)
```

Arguments

d	a distance/dissimilarity matrix of *N* entities (or a `dist` object).
V	an *N*x*q* matrix of *q* manifold auxiliary parameter values of the *N* entities.
u.dim	the embedding dimension.
epsilon	shortest dissimilarity retained.
k	Number of shortest dissimilarities retained for a point. If both `epsilon` and `k` are given, `epsilon` will be used.
W	an *N*x*N* symmetric weight matrix. If not given, a matrix of ones will be used.
method	if `matrix`, conditional SMACOF with a full B matrix will be used. If `vector`, conditional SMACOF with a diagonal B matrix will be used. The latter is more efficient for large *q*.
it.max	the max number of conditional SMACOF iterations.
gamma	conditional SMACOF stops early if the reduction of normalized conditional stress is less than `gamma`
init	method to initialize the starting values for the embedding and B (or b).
U.start	user-defined starting values for the embedding (when `U.init = 'user'`)
B.start	starting B matrix.
b.start	starting diagonal values of the B matrix (which is restricted to be diagonal).
...	other arguments for the `isomap{vegan}` function.

Value

U	the embedding result.
B	the estimated B matrix.
stress	Normalized conditional stress value.
sigma	the conditional stress value at each iteration.
init	method to initialize the starting values for the embedding and B (or b).
U.start	the starting values for the embedding.
B.start	starting values for the B matrix, if `method="matrix"`.
b.start	the starting values of the B matrix (which is restricted to be diagonal), if `method="vector"`.

Author(s)

Anh Tuan Bui

References

Bui, A.T. (2021). Dimension Reduction with Prior Information for Knowledge Discovery. arXiv:2111.13646.
URL: https://arxiv.org/abs/2111.13646.

See Also

[condMDS](#)

Examples

```
# see help(cml)
```

---

condMDS   *Conditional Multidimensional Scaling*

Description

Finds a low-dimensional embedding of a given distance/dissimilarity matrix, conditioning on auxil-
iary manifold parameters. The method uses conditional SMACOF to optimize its conditional stress
function.

Usage

```
condMDS(d, V, u.dim, W = NULL, method = c('matrix', 'vector'),
        it.max = 1000, gamma = 1e-05,
        init = c('none', 'user'), U.start = NULL,
        B.start = NULL, b.start = NULL)
```

Arguments

d	a distance/dissimilarity matrix of N entities (or a `dist` object).
V	an Nxq matrix of q manifold auxiliary parameter values of the N entities.
u.dim	the embedding dimension.
W	an NxN symmetric weight matrix. If not given, a matrix of ones will be used.
method	if `matrix`, conditional SMACOF with a full B matrix will be used. If `vector`, conditional SMACOF with a diagonal B matrix will be used. The latter is more efficient for large q.
it.max	the max number of conditional SMACOF iterations.
gamma	conditional SMACOF stops early if the reduction of normalized conditional stress is less than `gamma`
init	method to initialize the starting values for the embedding and B (or b).
U.start	user-defined starting values for the embedding (when `U.init = 'user'`)
B.start	starting B matrix.
b.start	starting diagonal values of the B matrix (which is restricted to be diagonal).

Value

U	the embedding result.
B	the estimated B matrix.
stress	Normalized conditional stress value.
sigma	the conditional stress value at each iteration.
init	method to initialize the starting values for the embedding and B (or b).
U.start	the starting values for the embedding.
B.start	starting values for the B matrix, if method="matrix".
b.start	the starting values of the B matrix (which is restricted to be diagonal), if method="vector".

Author(s)

Anh Tuan Bui

References

Bui, A.T. (2021). Dimension Reduction with Prior Information for Knowledge Discovery. arXiv:2111.13646.
URL: https://arxiv.org/abs/2111.13646.

See Also

[condIsomap](#)

Examples

```
# see help(cml)
```

---

condSmacof   *Conditional SMACOF*

Description

Conditional SMACOF algorithms. Intended for internal usage.

Usage

```
condSmacof_matrix(d, V, u.dim, W = NULL, it.max = 1000, gamma = 1e-05,
                  init = c('none', 'user'), U.start, B.start)
condSmacof_vector(d, V, u.dim, W = NULL, it.max = 1000, gamma = 1e-05,
                  init = c('none', 'user'), U.start, b.start)
```

Arguments

`d`	a `dist` object of *N* entities.
`V`	an *N*x*q* matrix of *q* manifold auxiliary parameter values of the *N* entities.
`u.dim`	the embedding dimension.
`W`	an *N*x*N* symmetric weight matrix. If not given, a matrix of ones will be used.
`it.max`	the max number of conditional SMACOF iterations.
`gamma`	conditional SMACOF stops early if the reduction of normalized conditional stress is less than `gamma`
`init`	method to initialize the starting values for the embedding and B (or b).
`U.start`	user-defined starting values for the embedding (when `U.init = 'user'`)
`B.start`	`condSmacof_matrix()` only; starting B matrix.
`b.start`	for `condSmacof_vector()` only; starting diagonal values of the B matrix (which is restricted to be diagonal).

Details

`condSmacof_matrix()` uses a full B matrix, whereas `condSmacof_vector()` restricts B to be a diagonal matrix and uses vector operations on the diagonal elements of B. The latter is therefore faster when *q* is large.

Value

`U`	the embedding result.
`B`	the estimated B matrix.
`stress`	Normalized conditional stress value.
`sigma`	the conditional stress value at each iteration.
`init`	method to initialize the starting values for the embedding and B (or b).
`U.start`	the starting values for the embedding.
`B.start`	the starting B matrix, if using `condSmacof_matrix()`
`b.start`	the starting diagonal values of the B matrix (which is restricted to be diagonal), if using `condSmacof_vector()`.

Author(s)

Anh Tuan Bui

References

Bui, A.T. (2021). Dimension Reduction with Prior Information for Knowledge Discovery. arXiv:2111.13646. URL: https://arxiv.org/abs/2111.13646.

Index