Package 'Preservation'

September 24, 2026

Title Assessing Preservation for Dimension Reduction Algorithms

Version 0.2.1

Date 2026-09-24

Description Contains tools to compute metrics that can compare the performance of dimension reduction algorithms for the preservation of pairwise distances, neighborhoods, or paths. For distances and neighborhoods, see <doi:10.1101/2025.04.28.651069>.

Depends R (>= 3.5.0)

Imports methods, stats, DRquality, emdist, FNN, ProjectionBasedClustering, coRanking

Suggests knitr, rmarkdown

License Apache License (== 2.0)

LazyLoad yes

URL http://silicovore.com/OOMPA/index.html

VignetteBuilder knitr

NeedsCompilation no

Author Kevin R. Coombes [cre, aut], Polina Bombina [aut]

Maintainer Kevin R. Coombes <krc@silicovore.com>

Contents

CAVA DAta *Dimension Reduction Applied to CAVA Daat Subset*

Description

A subset of protein expression data from a CITE-Seq experiment performed in the laboratory of Klaus Ley.

Usage

```
data("prot5")
data("path-idex")
data("drs")
```

Format

The `prot5` data set contains two objects. one,also called `prot5` is a numerical matrix protein expression data having 51 columns (proteins) and 3096 rows (individual cells). The second, `colvec` is a character vector holding coor assignments dividing the `prot5` data into four subclusters.

The `paths-idex` set contains an integer vector, `idex` of length 301, indirectly defining a path by pointing, in order, to rows of the data marix.

The `drs` set contains one object, `drs`, which is a list of four matrices, each with 2 columns and 3096 rows, representing the image of the 51-dimensioanl manifold after applying differnt dimension reduction algorithms,

Details

The `prot5` object is a subset of the protein data from the CAVA data set [Saigusa et al.]. The subset was slected because it represent a single coherent cluster of the CD4 T-cells in the experiment. We used hierarchical clustering with Ward's linkage rule and Euclidean distance to subdivide this cluster into four parts. The `colvec` vector records these cluster assignments. We then constructed a path in high dimendisonal space by performing principal components analysis and slecting all points close to the first principal component axis, ordered by their proection onto the axis. The (ordered) indices of the rows of these cells is sotred in the `idex` object. We also applid four dimension reduction algorithms (MDS and TSNE, using the `Rdimtools` package; UMAP, using the `umap` package, and DM, using the `diffusionMap` package) to the full `prot5` data set. The coordinates of the various rductions are stored inthe `drs` list.

References

Saigusa et al., Sex Differences in Coronary Artery Disease and Diabetes Revealed by scRNA-Seq and CITE-Seq of Human CD4+ T Cells. Int J Mol Sci. 2022; 23(17):9875. doi: 10.3390/ijms23179875.

Examples

```
data(prot5)
dim(prot5)
data(drs)
plot(drs[["MDS"]], col =colvec, pch = 16)
```

More Path Preservation Metrics

More Metrics for Path Preservation

Description

Metrics to estimate how well or poorly a dimension reduction algorithm preserves paths

Usage

```
Smoothness(M1, M2)
RankSmoothness(M1, M2)
Coil(M1, M2)
Contact(M1, M2, q = 0.1, type = c("mean", "max", "count"))
EndpointDisplacement(M1, M2, k = 5)
CrossingNumber(M2, eps = 1e-8)
```

Arguments

M1	a data matrix whose rows are the pointsin the original space.
M2	a data matrix whose rows are the same points after dimension reduction.
q	a real number; the quantile of pairwise diatances used to define "close".
type	a character based enumeration. If type = mean, we report the average number of steps between two points of contact. If type = max, we report the maximum number of steps. If type = count, we report the number of pairs whose separation is greater than 12 percent of all distances.
k	an intefger; the number of points to average at each end of the path.
eps	A real fudge factor (defauly 1E-8) used when normalizing the crossing number.

Value

Each of the functions returns a real number representing the corresponding scores comparing the spaces before and after dimension reduction.

Author(s)

Kevin R. Coombes <krc@silicovore.com>

Examples

```
data(prot5) # original space
data(drs) # spaces after diomwnion reduction
Smoothness(prot5, drs[["TSNE"]])
```

Neighborhood Preservation

Metrics for Preservation of Pairwise Distances

Description

Metrics to estimate how well or poorly a dimension reduction algorithm preserves local neighborhoods.

Usage

```
AverageJaccardDistance(M1, M2, k = 50)
AvgContinuity(M1, M2, k = 300)
AvgTrustworthiness(M1, M2, k = 300)
QNX(D1, D2)
```

Arguments

M1	a data matrix whose rows are the pointsin the original space.
M2	a data matrix whose rows are the same points after dimension reduction.
D1	a dist object containing the distances between points in the orignial space.
D2	a dist object containing the distances between points after dimension reduction.
k	an integer; the number of neighbors to consider.

Value

The Average Jaccard Distance returns a real number representing the score comparing the spaces before and after dimension reduction. The otgher three functions return numeric vectors of length k.

Author(s)

Kevin R. Coombes <krc@silicovore.com>

Examples

```
data(prot5) # original space
data(drs) # spaces after diomwnion reduction
AverageJaccardDistance(prot5, drs[["UMAP"]], k = 50)
```

Pairwise Distance Preservation
Metrics for Preservation of Pairwise Distances

Description

Metrics to estimate how well or poorly a dimension reduction algorithm preserves pairwise distances between points.

Usage

```
MilnorDistortion(D1, D2)
SigmaDistortion(D1, D2)
Stress(D1, D2)
M1Distortion(M1, M2)
SpearmanRho(D1, D2)
EarthMover(D1, D2)
```

Arguments

M1	a data matrix whose rows are the pointsin the original space.
M2	a data matrix whose rows are the same points after dimension reduction.
D1	a dist object containing the distances between points in the orignial space.
D2	a dist object containing the distances between points after dimension reduction.

Value

Each of the functions returns a real number representing the correpsonding scores comparing the spaces before and after dimension reduction.

Author(s)

Kevin R. Coombes <krc@silicovore.com>

Examples

```
data(prot5) # original space
data(drs) # spaces after diomwnion reduction
M1Distortion(prot5, drs[["UMAP"]])
```

path Preservation *Metrics for Path Preservation*

Description

Metrics to estimate how well or poorly a dimension reduction algorithm preserves paths.

Usage

```
LengthDistortion(M1, M2)
SegmentVariance(M1, M2)
Curvature(M1, M2)
SpatialSimilarity(M1, M2)
```

Arguments

M1	a data matrix decribing the path in the original space.
M2	a data matrix describing the (same) path after dimension reduction.

Value

Each of the four functions returns a real number representing the log ratio of the scores between the two path representation.

Author(s)

Kevin R. Coombes <krc@silicovore.com>

Examples

```
data(prot5) # original space
data(drs) # spaces after diomwnion reduction
LengthDistortion(prot5, drs[["TSNE"]])
```

Path RankPreservation *Rank-Based Metrics for Path Preservation*

Description

Rank-based metrics to estimate how well or poorly a dimension reduction algorithm preserves pathss.

Usage

```
RankLengthDistortion(M1, M2)
RankCurvature(M1, M2)
RankSpatialSimilarity(M1, M2)
```

Arguments

M1	a data matrix decribing the path in the original space.
M2	a data matrix describing the (same) path after dimension reduction.

Value

Each of the functions returns a real number representing the log ratio of the scores between the two path representation.

Author(s)

Kevin R. Coombes <krc@silicovore.com>

Examples

```
data(prot5) # original space
data(drs) # spaces after diomwnion reduction
RankCurvature(prot5, drs[["DM"]])
```

Index