Package: DRquality (via r-universe)

October 9, 2024

Type Package

Title Quality Measurements for Dimensionality Reduction

Version 0.2.1

Date 2023-08-12

Maintainer Michael Thrun <m.thrun@gmx.net>

Description Several quality measurements for investigating the performance of dimensionality reduction methods are provided here. In addition a new quality measurement called Gabriel classification error is made accessible.

License GPL-3

Imports DatabionicSwarm

Suggests plotly, geometry, deldir, FCPS, ProjectionBasedClustering, DataVisualizations, FastKNN, ggplot2, pcaPP,pracma, spdep, grid, igraph, cccd, sf

LazyData TRUE

Encoding UTF-8

Repository https://mthrun.r-universe.dev

RemoteUrl https://github.com/mthrun/drquality

RemoteRef HEAD

RemoteSha f633c77789563234b9f2ad7c6c125c4049638d92

Contents

`ClassificationError` *Classification Error (rate)*

Description

Compares projected points to a given prior classification using knn classifier.

Usage

```
ClassificationError(OutputDistances,Cls,k=5)
```

Arguments

`OutputDistances`	[1:n,1:n]
`Cls`	[1:n]
`k`	number of k nearest neighbors, in Venna 2010 set to 5 (here default)

Details

Projected points are evaluated by k-nearest neighbor classification accuracy (with k = 5), that is, each sample in the visualization is classified by majority vote of its k nearest neighbors in the visualization, and the classification is compared to the ground truth label. [Venna 2010].

Value

`Error`	Classification Error: 1-Accuracy[1]
`Accuracy`	Accuracy
`KNNCls`	[1:n]] cls of knn classifier

Note

Here, the Outputdistances of the Projected points are used.

Author(s)

Michael Thrun

References

Venna, J., Peltonen, J., Nybo, K., Aidos, H., and Kaski, S. Information retrieval perspective to nonlinear dimensionality reduction for data visualization. The Journal of Machine Learning Research, 11, 451-490. (2010)

Gracia, A., Gonzalez, S., Robles, V., and Menasalvas, E. A methodology to compare Dimensionality Reduction algorithms in terms of loss of quality. Information Sciences, 270, 1-27. (2014)

Examples

```
if(requireNamespace("FCPS")){
data(Hepta,package="FCPS")
projection=cmdscale(dist(Hepta$Data), k=2)
ClassificationError(as.matrix(dist(projection)),Hepta$Cls)
}
```

`Cmeasure` *C-Measure subtypes*

Description

Calculate the C-Measure subtypes of minimal path length and minimal wiring

Arguments

`Data`	[1:n,1:d] numerical matrix of points in input space.
`Projection`	[1:n,1:2] numerical matrix of points in output space.
`k`	Number of nearest neighbors, both measures set it always to k=1.

Value

[[1:2] Numerical vector of MinimalPathlength and MinimalWiring values.

Author(s)

Michael Thrun

Examples

```
if(requireNamespace("FCPS")){
data(Hepta,package="FCPS")
projection=cmdscale(dist(Hepta$Data), k=2)
Cmeasure(Hepta$Data,projection)
}
```

GabrielClassificationError

Gabriel Classification Error (GCE)

Description

GCE searches for the k-nearest neighbors of the first gabriel neighbors weighted by the Euclidean Distances of the Inputspace [Thrun et al, 2023]. GCE evaluates these neighbors in the Output space. A low value indicates a better two-dimensional projection of the high-dimensional Input space.

Usage

```
GabrielClassificationError(Data,ProjectedPoints,Cls,LC,
PlotIt=FALSE,Plotter = "native", Colors = NULL,LineColor= 'grey',
main = "Name of Projection", mainSize = 24,xlab = "X", ylab = "Y", xlim, ylim,
pch,lwd,Margin=list(t=50,r=0,l=0,b=0))
```

Arguments

Data	[1:n,1:d] Numeric matrix with n cases and d variables
ProjectedPoints	[1:n,1:2] Numeric matrix with 2D points in cartesian coordinates
Cls	[1:n] Numeric vector with class labels
LC	Optional, Numeric vector of two values determining grid size of the underlying projection
PlotIt	Optional, Boolean: TRUE/FALSE => Plot/Do not plot (Default: FALSE)
Plotter	Optional, Character with plot technique (native or plotly)
Colors	Optional, Character vector of class colors for points
LineColor	Optional, Character of line color used for edges of graph
main	Optional, Character plot title
mainSize	Optional, Numeric size of plot title
xlab	Optional, Character name of x ax
ylab	Optional, Character name of y ax
xlim	Optional, Numeric vector with two values defining x ax range
ylim	Optional, Numeric vector with two values defining y ax range
pch	Optional, Numeric of point size (graphic parameter)
lwd	Optional, Numeric of linewidth (graphic parameter)
Margin	Optional, Margin of plotly plot

Details

Gabriek classification error (GCE) makes an unbiased evaluation of distance and densitiybased structure which ma be even non-linear seperable. First, GCE utilizes the information provided by a prior classification to assess projected structures. Second, GCE applies the insights drawn from graph theory. Details are described in [Thrun et al, 2023]

Value

list of

`GCE`	GabrielClassificationError NOTE the rest is just for development purposes
`GCEperPoint`	[1:n] unnormalized GCE of each point: GCE = mean(GCEperPoint)
`nn`	the number of points in a relevant neghborhood: 0.5 * 85percentile(AnzNN)
`AnzNN`	[1:n] the number of points with a gabriel graph neighborhood
`NNdists`	[1:n,1:nn] the distances within the relevant neighborhood, 0 for inner cluster distances
`HD`	[1:nn] HD = HarmonicDecay(nn) i.e weight function for the NNdists: GCEperPoint = HD*NNdists
`IsInterDistance`	Distances to the nn closest neighbors
`GabrielDists`	Distance matrix implied by high dimensional distances and the underlying gabriel (Gabriel) graph
`ProjectionGraphError`	Plotly object in case, plotly is chose

Author(s)

Michael Thrun, Quirin Stier, Julian Märte

References

[Thrun et al, 2023] Thrun, M.C, Märte, J., Stier, Q.: Analyzing Quality Measurements for Dimensionality Reduction, Machine Learning and Knowledge Extraction (MAKE), Vol 5., accepted, 2023.

Examples

```
if(requireNamespace("FCPS")){
data(Hepta,package="FCPS")
projection=cmdscale(dist(Hepta$Data), k=2)
GabrielClassificationError(Hepta$Data,projection,Hepta$Cls)$GCE
}


if(requireNamespace("FCPS")){
data(Hepta,package="FCPS")
projection=cmdscale(dist(Hepta$Data), k=2)
GabrielClassificationError(Hepta$Data,projection,Hepta$Cls)$GCE
```

```
}
```

`KendallsTau` *Statistical correlation by Kendall*

Description

Calculates the statistical correlation by Kendall. Basically a wrapper to pcaPP::cor.fk.

Usage

```
KendallsTau(InputDists, OutputDists)
```

Arguments

`InputDists`	Matrix containing the distances of the first dataset.
`OutputDists`	Matrix containing the distances of the second dataset.

Value

Equivalent to `cor.fk`

Author(s)

Michael Thrun

Examples

```
if(requireNamespace("FCPS")){
data(Hepta,package="FCPS")
InputDist=dist(Hepta$Data)
projection=cmdscale(InputDist, k=2)
KendallsTau(as.matrix(InputDist),as.matrix(dist(projection)))
}
```

plotMeasureRAAR *Computes rank-based smoothed precision and recall*

Description

Compares the projection in pData with the original data in Data and calculates trustworthiness and continuity of the projection for neighborhood sizes ranging from 1 to the size of the neighborhood.

Usage

```
plotMeasureRAAR(Raar, label = 'ProjectionMethod',
gPlotList = list(RAARplot = ggplot2::ggplot()), LineType="solid", Shape = 16,
PointsPerE = 10, fancy = FALSE)
```

Arguments

Raar	Output of RAAR() applied for a projection method.
label	Title of plot.
gPlotList	Settings for ggplot.
LineType	Character - graphic parameter: Line type of ggplot.
Shape	Integer: type of point
PointsPerE	Numeric graphic parameter: Distance between markers on plot line
fancy	Boolean graphic parameter: Some automatic settings for a more appealing plot.

Value

ggplot object

Author(s)

Michael Thrun

plotMeasureTundD *Computes rank-based smoothed precision and recall*

Description

Compares the projection in pData with the original data in Data and calculates trustworthiness and continuity of the projection for neighborhood sizes ranging from 1 to the size of the neighborhood.

Usage

```
plotMeasureTundD(TDmatrix, label = 'ProjectionMethod',
gPlotList = list(TW = ggplot2::ggplot(), DC = ggplot2::ggplot()), LineType = "solid",
Shape = 16, PointsPerE = 16)
```

Arguments

TDmatrix	Output of MeasureTundD() applied for a projection method.
label	Title of plot.
gPlotList	Settings for ggplot.
LineType	Character - graphic parameter: Line type of ggplot.
Shape	Integer: type of point
PointsPerE	Numeric graphic parameter: Distance between markers on plot line

Value

ggplot object

Author(s)

Michael Thrun

RAAR *Rescaled average agreement rate*

Description

Rescaled average agreement rate deduced by the co-ranking matrix from LCMC.

Usage

```
RAAR(Data, ProjectedPoints, kmax = nrow(Data) - 2, PlotIt = T)
```

Arguments

Data	Matrix containing n cases in rows, d variables in columns or a distance matrix which in this case has to be symmetric
ProjectedPoints	n by OutputDimension matrix containing coordinates of the Projection
kmax	maximum of intervall 1:kmax of k nearest neighbors
PlotIt	Optional: Should the output be plottet. Default: TRUE

Value

A list containing:

Raar	Rescaled average agreement rate
Aar	Average agreement rate

Author(s)

Michael Thrun

References

Lee, J. A., Peluffo-Ordonez, D. H., & Verleysen, M. Multiscale stochastic neighbor embedding: Towards parameter-free dimensionality reduction. Paper presented at the Proceedings of 22st European Symposium on Artificial Neural Networks, Computational Intelligence And Machine Learning (ESANN) (2014).

`SpearmanError` *Calculates the error of a projection with spearman's rank correlation coefficient*

Description

Calculates the error of a projection with spearman's rank correlation coefficient.

Arguments

`VectorOfInputDists(1:n2)`
: dissimilarities in Input Space between the n data points in vector form as produced by squareform(Dists(1:n,1:n))

`VectorOfOutputDists(1:n2)`
: dissimilarities in Input Space between the n data points in vector form as produced by squareform(Dists(1:n,1:n))

Value

rho rank correlation coefficient

Author(s)

Florian Lerch

`SpearmansRho` *Calculates the error of a projection with spearman's rank correlation coefficient*

Description

Calculates the error of a projection with spearman's rank correlation coefficient

Usage

```
SpearmansRho(InputDists, OutputDists)
```

Arguments

InputDists	[1:d,1:d] numeric matrix with input distances
OutputDists	[1:d,1:d] numeric matrix with output distances

Value

rho

Author(s)

Julian Märte

TopologicalCorrelation

Topological Correlation

Description

Calculates the Topololgical Correlation

Usage

```
TopologicalCorrelation(Data,ProjectedPoints,type='norm',method,Kn=0)
```

Arguments

Data	a matrix of the given n-dim. points: the rows represent the points and the columns represent the coordinates in the n-dim. space.
ProjectedPoints	matrix of Projected Points, if missing, method should be set!
method	Determines whether the selected projections method for a given set of n-Dim. points is a good choice. Therefor, a result of 1 means the seleceted projections method is good, and a result value of 0 means that the Visualization of the given Data in the two dim. space doesnt fit for the problem.
type	How the paths in the adjacencematrix should be weighted, norm representes path lenthgs of 1 and eucldidean represents the distance in the euclidean metric.
Kn	k nearest neighbours in the graph. only needed in method is isomap and LocallyLinearEmbedding

Value

TC value

Author(s)

Hermann Tafo, Laukert Schlichting 07/2015

Examples

```
if(requireNamespace("FCPS")){
data(Hepta,package="FCPS")
projection=cmdscale(dist(Hepta$Data), k=2)
TopologicalCorrelation(Hepta$Data,projection)
}
```

`ZrehenMeasure4All` *A generalized version of the zrehen-measure which defines the neighbourhood by gabrielgraph and is therefore not restricted to grid-based projections.*

Description

A generalized version of the zrehen-measure which defines the neighbourhood by gabrielgraph and is therefore not restricted to grid-based projections.

Arguments

`Data`	[1:n,1:d] points in input room with d attributes
`Projection`	[1:n,1:2] projected points in output room, with index,x,y or index,line,column
`width`	only necessary if toroid
`height`	only necessary if toroid
`isToroid`	are the points toroid?
`isGrid`	is the grid a toroid?
`plotGabriel`	plot the generated GabrielGraph

Value

List with

`V$zrehen`	the raw zrehen measure
`V$normedzrehen`	the zrehen measure normed by the number of neighbours
`v$neighbourcounter`	the number of possible neighbours by which the zrehen measure is normed

Author(s)

Florian Lerch 07/2015

Examples

```
if(requireNamespace("FCPS")){
data(Hepta,package="FCPS")
projection=cmdscale(dist(Hepta$Data), k=2)
ZrehenMeasure4All(Hepta$Data,projection)$zrehen
}
```

Index