Package ‘DRquality’
August 11, 2023

Type Package
Title Quality Measurements for Dimensionality Reduction
Version 0.2.0
Date 2023-08-08
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
Encoding UTF-8
NeedsCompilation no
Author Quirin Stier [aut],
  Florian Lerch [ctb],
  Julian Mürte [aut],
  Hermann Tafo [ctb],
  Laukert Schlichting [ctb],
  Michael Thrun [aut, cph, cre] (<https://orcid.org/0000-0001-9542-5543>)
Repository CRAN
Date/Publication 2023-08-11 07:50:08 UTC

R topics documented:

  ClassificationError ........................................ 2
  Cmeasure ..................................................... 3
  GabrielClassificationError ................................. 4
  Kendall'sTau .................................................. 6
  SpearmanError ............................................... 7
  TopologicalCorrelation ..................................... 7
  ZrehenMeasure4All .......................................... 8
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
Accuracy Accuracy
KNNCls [1:n] cls of knn classifier

Note

Here, the Outputdistances of the Projected points are used.

Author(s)

Michael Thrun

References


**Examples**

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

---

<table>
<thead>
<tr>
<th>Cmeasure</th>
<th>C-Measure subtypes</th>
</tr>
</thead>
</table>

**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**

```r
if(requireNamespace("FCPS")){
  data(Hepta, package="FCPS")
  projection=cmdscale(dist(Hepta$Data), k=2)
  Cmeasure(Hepta$Data, projection)
}
```
**Gabriel Classification Error (GCE)**

**Description**

GCE searches for the k-nearest neighbors of the first gabriel neighbors weighted by the Euclidean Distances of the Input space [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**

```r
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
GabrielClassificationError

**Details**

Gabriel classification error (GCE) makes an unbiased evaluation of distance and density-based structure which may be even non-linear separable. 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
- GCEperPoint [1:n] unnormalized GCE of each point: $GCE = \text{mean}(GCE_{\text{perPoint}})$
- nn the number of points in a relevant neighborhood: $0.5 \times 85\% \text{ percentile}(\text{AnzNN})$
- 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] $HD = \text{HarmonicDecay}(nn)$ i.e weight function for the NNdists: $GCE_{\text{perPoint}} = HD \times NNdists$
- 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 chosen

**Author(s)**

Michael Thrun, Quirin Stier, Julian Märte

**References**


**Examples**

```r
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)
```
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)))
}
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

TopologicalCorrelation

Topological Correlation

Description
Calculates the Topological 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!
ZrehenMeasure4All

method Determines whether the selected projections method for a given set of n-Dim. points is a good choice. Therefore, a result of 1 means the selected projections method is good, and a result value of 0 means that the visualization of the given Data in the two dim. space doesn’t fit for the problem.

type How the paths in the adjacency matrix should be weighted, norm represents path lengths of 1 and euclidean represents the distance in the euclidean metric.

K\textsubscript{n} 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

```r
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

- \( V\text{\$zrehen} \) the raw zrehe measure
- \( V\text{\$normedzrehen} \) the zrehe measure normed by the number of neighbours
- \( v\text{\$neighbourcounter} \)
- the number of possible neighbours by which the zrehe measure is normed

**Author(s)**

Florian Lerch 07/2015

**Examples**

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

* Classification Error
  ClassificationError, 2
  GabrielClassificationError, 4
* ClassificationErrorRate
  ClassificationError, 2
* ClassificationError
  ClassificationError, 2
* Evaluation of projection methods
  GabrielClassificationError, 4
* GCE
  GabrielClassificationError, 4
* Gabriel Classification Error
  GabrielClassificationError, 4
* QM
  GabrielClassificationError, 4
* Quality measurement
  GabrielClassificationError, 4
* Quality measure
  GabrielClassificationError, 4

ClassificationError, 2
Cmeasure, 3
cor.fk, 6

GabrielClassificationError, 4
KendallsTau, 6
SpearmanError, 7
TopologicalCorrelation, 7
ZrehenMeasure4All, 8