Package ‘dobin’

Type Package

Title Dimension Reduction for Outlier Detection

Version 1.0.4

Maintainer Sevvandi Kandanaarachchi <sevvandik@gmail.com>

Description A dimension reduction technique for outlier detection. DOBIN: a Distance based Outlier BasIs using Neighbours, constructs a set of basis vectors for outlier detection. This is not an outlier detection method; rather it is a pre-processing method for outlier detection. It brings outliers to the fore-front using fewer basis vectors (Kandanaarachchi, Hyndman 2020) <doi:10.1080/10618600.2020.1807353>.

License MIT + file LICENSE

Encoding UTF-8

Imports dbscan, ggplot2, pracma

RoxygenNote 7.2.1

Suggests knitr, rmarkdown, OutliersO3, FNN

VignetteBuilder knitr

Depends R (>= 3.4.0)

URL https://sevvandi.github.io/dobin/

NeedsCompilation no

Author Sevvandi Kandanaarachchi [aut, cre]

(https://orcid.org/0000-0002-0337-0395)

Repository CRAN

Date/Publication 2022-08-25 22:52:33 UTC

R topics documented:

autoplot.dobin ................................................................. 2
dobin ............................................................................. 2

Index 4
**autoplot.dobin**

*Plots the first two components of the dobin space.*

**Description**

Scatterplot of the first two columns in the dobin space.

**Usage**

```r
## S3 method for class 'dobin'
autoplot(object, ...)
```

**Arguments**

- `object` The output of the function ‘dobin’.
- `...` Other arguments currently ignored.

**Value**

A ggplot object.

**Examples**

```r
X <- rbind(  
data.frame(x = rnorm(500), y = rnorm(500), z = rnorm(500)),  
data.frame(x = rnorm(5, mean = 10, sd = 0.2), y = rnorm(5, mean = 10, sd = 0.2), z = rnorm(5, mean = 10, sd = 0.2))
)
dob <- dobin(X)
autoplot(dob)
```

**dobin**

*Computes a set of basis vectors for outlier detection.*

**Description**

This function computes a set of basis vectors suitable for outlier detection.

**Usage**

```r
dobin(xx, frac = 0.95, norm = 1, k = NULL)
```
**Arguments**

- **xx** The input data in a dataframe, matrix or tibble format.
- **frac** The cut-off quantile for \( Y \) space. Default is 0.95.
- **norm** The normalization technique. Default is Min-Max, which normalizes each column to values between 0 and 1. \( \text{norm} = 0 \) skips normalization. Other values of norm defaults to Median-IQR normalization.
- **k** Parameter \( k \) for \( k \) nearest neighbours with a default value of 5\% of the number of observations with a cap of 20.

**Value**

A list with the following components:

- **rotation** The basis vectors suitable for outlier detection.
- **coords** The dobin coordinates of the data xx.
- **Yspace** The associated \( Y \) space.
- **Ypairs** The pairs in xx used to construct the \( Y \) space.
- **zerosdcols** Columns in xx with zero standard deviation. This is computed only if the number of columns are greater than the number of rows.

**Examples**

```r
# A bimodal distribution in six dimensions, with 5 outliers in the middle.
set.seed(1)
x2 <- rnorm(405)
x3 <- rnorm(405)
x4 <- rnorm(405)
x5 <- rnorm(405)
x6 <- rnorm(405)
x1_1 <- rnorm(mean = 5, 400)
mu2 <- 0
x1_2 <- rnorm(5, mean=mu2, sd=0.2)
x1 <- c(x1_1, x1_2)
X1 <- cbind(x1,x2,x3,x4,x5,x6)
X2 <- cbind(-1*x1_1,x2[1:400],x3[1:400],x4[1:400],x5[1:400],x6[1:400])
X <- rbind(X1, X2)
labs <- c(rep(0,400), rep(1,5), rep(0,400))
dob <- dobin(X)
autoplot(dob)
```
Index

autoplot.dobin, 2

dobin, 2