Package ‘dobin’

February 24, 2020

Type Package

Title Dimension Reduction for Outlier Detection

Version 1.0.2

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 2019) <doi:10.13140/RG.2.2.15437.18403>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports pracma, RANN

RoxygenNote 6.1.1

Suggests knitr, rmarkdown, OutliersO3, ggplot2, FNN

VignetteBuilder knitr

Depends R (>= 3.4.0)

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

NeedsCompilation no

Author Sevvandi Kandanaarachchi [aut, cre]

Repository CRAN

Date/Publication 2020-02-24 05:20:02 UTC

R topics documented:

   dobin ................................................................. 2

Index 4
**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 Median-IQR, which normalizes each column of median 0 and IQR 1.
- `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:

- `vec`: The basis vectors suitable for outlier detection.
- `coords`: The dobin coordinates of the data `xx`.
- `Y`: 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.
s.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)
```
```r
labs <- c(rep(0,400), rep(1,5), rep(0,400))
out <- dobin(X)
plot(out$coords[, 1:2], col=as.factor(labs), pch=20)
```
Index

dobin, 2