Package 'SMLoutliers'

January 20, 2025

Type Package

Title Outlier Detection Using Statistical and Machine Learning Methods

Version 0.1

Date 2017-06-06

Author Siddharth Jain and Prabhanjan Tattar

Maintainer Siddharth Jain <siddharthjain242@gmail.com>

Description Local Correlation Integral (LOCI) method for outlier identification is implemented here. The LOCI method developed here is invented in Breunig, et al. (2000), see <doi:10.1145/342009.335388>.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2017-02-06 17:38:16

Contents

`SMLoutliers-package` *An R package for identifying outliers using statistical and machine learning methods*

Description

We intend to provide host of methods for identifying outliers. This will cut across statistical and machine learning methods.

References

M.M. Breunig, H.P. Kriegel, R.T. Ng, and J. Sander. Lof: Identifying density-based local outliers. In Proc. SIGMOD Conf., pages 93-104, 2000.

Examples

```
data(stiff)
summary(stiff)
```

LOCI *Local Correlation Integral*

Description

We provide an R implementation of the Local Correlation Integral method for detecting outliers as developed by Breunig, et al. (2000), and we follow its description given in Papadimitriou, et al. (2002).

Usage

```
LOCI(data, alpha)
```

Arguments

data	Any R data.frame which consists of numeric values only
alpha	a number in the unit interval for the fractional circle search

Details

A simple implementation is provided here. The core function is the distance function. For each observation, a search is made for nearest neighbors within r distance of it, and then for each of these neighbors, we find the number of observations in the fractional circle. Calculations based on multi-granularity deviation factor, MDEF, help in determining the outlier.

Author(s)

Siddharth Jain and Prabhanjan Tattar

References

M.M. Breunig, H.P. Kriegel, R.T. Ng, and J. Sander. Lof: Identifying density-based local outliers. In Proc. SIGMOD Conf., pages 93-104, 2000. Papadimitriou, S., Kitagawa, H., Gibbons, P.B. and Faloutsos, C., 2003, March. Loci: Fast outlier detection using the local correlation integral. In Data Engineering, 2003. Proceedings. 19th International Conference on (pp. 315-326). IEEE.

Examples

```
data(stiff)
OM <- LOCI(stiff,0.5)
OM
```

---

stiff *The Board Stiffness Dataset*

Description

Four measures of stiffness of 30 boards are available. The first measure of stiffness is obtained by sending a shock wave down the board, the second measure is obtained by vibrating the board, and remaining are obtained from static tests.

Usage

```
data(stiff)
```

Format

A data frame with 30 observations on the following 4 variables.

`x1` first measure of stiffness is obtained by sending a shock wave down the board

`x2` second measure is obtained by vibrating the board

`x3` third measure is obtained by a static test

`x4` fourth measure is obtained by a static test

References

Johnson, R.A., and Wichern, D.W. (1982-2007). Applied Multivariate Statistical Analysis, 6e. Pearson Education. Tattar, et al. (2016). A Course in Statistics with R. J. Wiley.

Examples

```
data(stiff)
summary(stiff)
```

Index