Package 'PPCDT'

July 21, 2025

Type Package

Title An Optimal Subset Selection for Distributed Hypothesis Testing

Version 0.2.0

Maintainer Guangbao Guo <ggb11111111@163.com>

Description In the era of big data, data redundancy and distributed characteristics present novel challenges to data analysis. This package introduces a method for estimating optimal subsets of redundant distributed data, based on PPCDT (Conjunction of Power and P-value in Distributed Settings). Leveraging PPC technology, this approach can efficiently extract valuable information from redundant distributed data and determine the optimal subset. Experimental results demonstrate that this method not only enhances data quality and utilization efficiency but also assesses its performance effectively. The philosophy of the package is described in Guo G. (2020) <doi:10.1007/s00180-020-00974-4>.

License Apache License (== 2.0)

Depends R (>= 3.5.0)

Encoding UTF-8

Imports MASS,stats

NeedsCompilation no

RoxygenNote 7.3.1

Author Guangbao Guo [aut, cre, cph],
Jiarui Li [ctb]

Repository CRAN

Date/Publication 2024-07-08 11:00:06 UTC

Contents

PPCDT   *An Optimal Subset Selection for Distributed Hypothesis Testing*

Description

We introduce an optimal subset selection for distributed hypothesis testing called as PPCDT.

Usage

```
PPCDT(X,Y,alpha,K)
```

Arguments

X	A independent variable
Y	The response variable
alpha	Significance level
K	The number of blocks into which variable X is divided

Value

Xopt	optimal subset of selected independent variables
Yopt	optimal subset of selected response variables
Bopt	Regression coefficients
Eopt	The Mean Squared Error of optimal subset
Aopt	The Mean Absolute Error of optimal subset

Author(s)

Guangbao Guo, Jiarui Li

Examples

```
alpha=0.05
t=5;K=10;n=1000;p=5
X=matrix(rnorm(n*p,0,1),ncol=p)
beta=matrix(runif(p),nrow = p)
esp=matrix(rnorm(n),nrow = n)
Y=X%*%beta+esp
PPCDT(X=X,Y=Y,alpha=alpha,K=K)
```

Index