

# Package ‘PPCDT’

May 25, 2024

**Type** Package

**Title** An Optimal Subset Selection for Distributed Hypothesis Testing

**Version** 0.1.0

**Maintainer** Guangbao Guo <ggb11111111@163.com>

**Description** In the era of big data, data redundancy and distributed characteristics pose new challenges to data analysis. This article proposes a method of optimal subset estimation for redundant distributed data based on PPC. Through PPC technology, this method can effectively extract useful information from redundant distributed data and estimate the optimal subset. Experimental results show that this method can improve data quality and utilization efficiency and evaluate its performance. The philosophy of the package is described in Guo G. (2020) <[doi:10.1007/s00180-020-00974-4](https://doi.org/10.1007/s00180-020-00974-4)>.

**License** MIT + file LICENSE

**Depends** R (>= 3.5.0)

**Encoding** UTF-8

**Imports** MASS,stats

**NeedsCompilation** no

**RoxygenNote** 7.3.1

**Suggests** testthat (>= 3.0.0)

**Config/testthat/edition** 3

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

**Repository** CRAN

**Date/Publication** 2024-05-25 17:20:09 UTC

## R topics documented:

PPCDT	2
Index	3

**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),n,col=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

PPCDT, [2](#)