Package 'ELIC'

September 4, 2025

Type Package

Title LIC for Distributed Elliptical Model

Version 0.1.0

Date 2025-08-16

Description This comprehensive toolkit for Distributed Elliptical model is designated as ``ELIC'' (The LIC for Distributed Elliptical Model Analysis) analysis. It is predicated on the assumption that the error term adheres to a Elliptical distribution. The philosophy of the package is described in Guo G. (2020) <doi:10.1080/02664763.2022.2053949>.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.2

NeedsCompilation no

Author Guangbao Guo [aut, cre],
Xiyu Zhao [aut]

Maintainer Guangbao Guo <ggb11111111@163.com>

Repository CRAN

Config/testthat/edition 3

Imports distr, distrEllipse,MASS

Suggests testthat (>= 3.0.0), sn

Depends R (>= 4.4.0)

Date/Publication 2025-09-04 14:20:33 UTC

Contents

`beta_AD` *Caculate the estimators of beta on the A-opt and D-opt*

Description

Caculate the estimators of beta on the A-opt and D-opt

Usage

```
beta_AD(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

Arguments

`K`	is the number of subsets
`nk`	is the length of subsets
`alpha`	is the significance level
`X`	is the observation matrix
`y`	is the response vector

Value

A list containing:

`betaA`	The estimator of beta on the A-opt.
`betaD`	The estimator of beta on the D-opt.

References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). doi:10.1007/s1122202410471z

Guo, G., Sun, Y., Qian, G., & Wang, Q. (2022). LIC criterion for optimal subset selection in distributed interval estimation. *Journal of Applied Statistics*, 50(9), 1900-1920. doi:10.1080/02664763.2022.2053949.

Chang, D., Guo, G. (2024). LIC: An R package for optimal subset selection for distributed data. *SoftwareX*, 28, 101909.

Jing, G., & Guo, G. (2025). TLIC: An R package for the LIC for T distribution regression analysis. *SoftwareX*, 30, 102132.

Chang, D., & Guo, G. Research on Distributed Redundant Data Estimation Based on LIC. *IAENG International Journal of Applied Mathematics*, 55(1), 1-6 (2025).

Gao, H., & Guo, G. LIC for Distributed Skewed Regression. *IAENG International Journal of Applied Mathematics*, 55(9), 2925-2930 (2025).

Zhang, C., & Guo, G. (2025). The optimal subset estimation of distributed redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 270-277.

Jing, G., & Guo, G. (2025). Student LIC for distributed estimation. *IAENG International Journal of Applied Mathematics*, 55(3), 575-581.

Liu, Q., & Guo, G. (2025). Distributed estimation of redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 332-337.

Examples

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%*%beta+e;
beta_AD(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

`beta_cor` *Caculate the estimator of beta on the COR*

Description

Caculate the estimator of beta on the COR

Usage

```
beta_cor(K = K, nk = nk, alpha = alpha, X = X, y = y)
```

Arguments

`K`	is the number of subsets
`nk`	is the length of subsets
`alpha`	is the significance level
`X`	is the observation matrix
`y`	is the response vector

Value

A list containing:

`betaC`	The estimator of beta on the COR.

References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). doi:10.1007/s1122202410471z

Guo, G., Sun, Y., Qian, G., & Wang, Q. (2022). LIC criterion for optimal subset selection in distributed interval estimation. *Journal of Applied Statistics*, 50(9), 1900-1920. doi:10.1080/02664763.2022.2053949.

Chang, D., Guo, G. (2024). LIC: An R package for optimal subset selection for distributed data. *SoftwareX*, 28, 101909.

Jing, G., & Guo, G. (2025). TLIC: An R package for the LIC for T distribution regression analysis. *SoftwareX*, 30, 102132.

Chang, D., & Guo, G. Research on Distributed Redundant Data Estimation Based on LIC. *IAENG International Journal of Applied Mathematics*, 55(1), 1-6 (2025).

Gao, H., & Guo, G. LIC for Distributed Skewed Regression. *IAENG International Journal of Applied Mathematics*, 55(9), 2925-2930 (2025).

Zhang, C., & Guo, G. (2025). The optimal subset estimation of distributed redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 270-277.

Jing, G., & Guo, G. (2025). Student LIC for distributed estimation. *IAENG International Journal of Applied Mathematics*, 55(3), 575-581.

Liu, Q., & Guo, G. (2025). Distributed estimation of redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 332-337.

Examples

```
p=6;n=1000;K=2;nk=200;alpha=0.05;sigma=1
e=rnorm(n,0,sigma); beta=c(sort(c(runif(p,0,1))));
data=c(rnorm(n*p,5,10));X=matrix(data, ncol=p);
y=X%*%beta+e;
beta_cor(K=K,nk=nk,alpha=alpha,X=X,y=y)
```

`eerr` *Generate Data with Elliptically Distributed Covariates*

Description

This function generates a dataset for a linear model where the covariate matrix X follows an elliptical distribution.

Usage

```
eerr(n, p, dist_type)
```

Arguments

`n`	The number of observations (rows) to generate.
`p`	The number of predictors/dimensions (columns) for the covariate matrix X.
`dist_type`	A character string specifying the type of elliptical distribution for X. Must be one of "Elliptical-Normal", "Elliptical-t", or "Elliptical-cov".

Details

The function generates a response vector Y based on the linear model Y = X The covariate matrix X is generated from one of three types of elliptical distributions: 1. 'Elliptical-Normal': Based on a multivariate normal distribution structure. 2. 'Elliptical-t': Based on a multivariate t-distribution structure. 3. 'Elliptical-cov': Based on a custom covariance matrix adjusted via its eigenvalues. The error term 'e' is drawn from a standard normal distribution.

Value

A list containing the following components:

X	An n x p matrix of covariates from the specified elliptical distribution.
Y	A numeric vector of n responses.
e	A numeric vector of n error terms from a standard normal distribution.

References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). doi:10.1007/s1122202410471z

Guo, G., Sun, Y., Qian, G., & Wang, Q. (2022). LIC criterion for optimal subset selection in distributed interval estimation. *Journal of Applied Statistics*, 50(9), 1900-1920. doi:10.1080/02664763.2022.2053949.

Chang, D., Guo, G. (2024). LIC: An R package for optimal subset selection for distributed data. *SoftwareX*, 28, 101909.

Jing, G., & Guo, G. (2025). TLIC: An R package for the LIC for T distribution regression analysis. *SoftwareX*, 30, 102132.

Chang, D., & Guo, G. Research on Distributed Redundant Data Estimation Based on LIC. *IAENG International Journal of Applied Mathematics*, 55(1), 1-6 (2025).

Gao, H., & Guo, G. LIC for Distributed Skewed Regression. *IAENG International Journal of Applied Mathematics*, 55(9), 2925-2930 (2025).

Zhang, C., & Guo, G. (2025). The optimal subset estimation of distributed redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 270-277.

Jing, G., & Guo, G. (2025). Student LIC for distributed estimation. *IAENG International Journal of Applied Mathematics*, 55(3), 575-581.

Liu, Q., & Guo, G. (2025). Distributed estimation of redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 332-337.

Examples

```
# Generate 100 observations with 5 predictors from an Elliptical-Normal distribution
data_normal <- eerr(n = 100, p = 5, dist_type = "Elliptical-Normal")
str(data_normal)

# Generate 100 observations with 3 predictors from an Elliptical-cov distribution
data_cov <- eerr(n = 100, p = 3, dist_type = "Elliptical-cov")
pairs(data_cov$X) # Visualize the relationships between covariates
```

ELIC *A General Length and Information Criterion (LIC) Function*

Description

This function applies the LIC method to find an optimal data subset, supporting various error term distributions like T-distribution and skewed distributions.

Usage

```
ELIC(X, Y, alpha = 0.05, K = 10, nk = NULL, dist_type = "student_t")
```

Arguments

X	A numeric design matrix.
Y	A numeric response vector.
alpha	The significance level for criterion calculation, default is 0.05.
K	The number of subsets to sample, default is 10.
nk	The sample size of each subset. If NULL (default), it's calculated as n/K.
dist_type	A character string specifying the assumed error distribution. Accepts T-distribution types (e.g., "student_t") from the original TLIC, and skewed types ("skew_normal", "skew_t", "skew_laplace") from SLIC. Note: In this implementation, the core calculation is robust and does not change based on dist_type. The parameter is kept for consistency with the original functions.

Details

The function iteratively samples subsets from the data, calculates a length criterion (L1) and an information criterion (N), and finds an optimal subset based on the intersection of the best samples from both criteria. It is a general implementation combining the logic of TLIC and SLIC.

Value

A list containing the optimal model components:

MUopt	The predicted values for the optimal subset.
Bopt	The estimated coefficients for the optimal model.
MAEMUopt	The Mean Absolute Error of the optimal model.
MSEMUopt	The Mean Squared Error of the optimal model.
opt	The indices of the optimal data subset.
Yopt	The response values of the optimal subset.

References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). doi:10.1007/s1122202410471z

Guo, G., Sun, Y., Qian, G., & Wang, Q. (2022). LIC criterion for optimal subset selection in distributed interval estimation. *Journal of Applied Statistics*, 50(9), 1900-1920. doi:10.1080/02664763.2022.2053949.

Chang, D., Guo, G. (2024). LIC: An R package for optimal subset selection for distributed data. *SoftwareX*, 28, 101909.

Jing, G., & Guo, G. (2025). TLIC: An R package for the LIC for T distribution regression analysis. *SoftwareX*, 30, 102132.

Chang, D., & Guo, G. Research on Distributed Redundant Data Estimation Based on LIC. *IAENG International Journal of Applied Mathematics*, 55(1), 1-6 (2025).

Gao, H., & Guo, G. LIC for Distributed Skewed Regression. *IAENG International Journal of Applied Mathematics*, 55(9), 2925-2930 (2025).

Zhang, C., & Guo, G. (2025). The optimal subset estimation of distributed redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 270-277.

Jing, G., & Guo, G. (2025). Student LIC for distributed estimation. *IAENG International Journal of Applied Mathematics*, 55(3), 575-581.

Liu, Q., & Guo, G. (2025). Distributed estimation of redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 332-337.

Examples

```
# Example with T-distributed error data (like TLIC)
set.seed(12)
n <- 200
p <- 5
X_t <- matrix(stats::runif(n * p), ncol = p)
beta_t <- sort(stats::runif(p, 1, 5))
e_t <- stats::rt(n, df = 5)
Y_t <- X_t %*% beta_t + e_t
result_t <- ELIC(X_t, Y_t, dist_type = "student_t")
str(result_t)

# Example with Skew-Normal error data (like SLIC)
if (requireNamespace("sn", quietly = TRUE)) {
  set.seed(123)
  n <- 200
  p <- 5
  X_s <- matrix(stats::rnorm(n * p), ncol = p)
  beta_s <- stats::runif(p, 1, 2)
  e_s <- sn::rsn(n = n, xi = 0, omega = 1, alpha = 5)
  Y_s <- X_s %*% beta_s + e_s
  result_s <- ELIC(X_s, Y_s, K = 5, dist_type = "skew_normal")
  str(result_s)
}
```

`LICnew` *Calculate the LIC estimator based on A-optimal and D-optimal criterion*

Description

Calculate the LIC estimator based on A-optimal and D-optimal criterion

Usage

```
LICnew(X, Y, alpha, K, nk)
```

Arguments

`X`	A matrix of observations (design matrix) with size n x p
`Y`	A vector of responses with length n
`alpha`	The significance level for confidence intervals
`K`	The number of subsets to consider
`nk`	The size of each subset

Value

A list containing:

`E5`	The LIC estimator based on A-optimal and D-optimal criterion.

References

Guo, G., Song, H. & Zhu, L. The COR criterion for optimal subset selection in distributed estimation. *Statistics and Computing*, 34, 163 (2024). doi:10.1007/s1122202410471z

Guo, G., Sun, Y., Qian, G., & Wang, Q. (2022). LIC criterion for optimal subset selection in distributed interval estimation. *Journal of Applied Statistics*, 50(9), 1900-1920. doi:10.1080/02664763.2022.2053949.

Chang, D., Guo, G. (2024). LIC: An R package for optimal subset selection for distributed data. *SoftwareX*, 28, 101909.

Jing, G., & Guo, G. (2025). TLIC: An R package for the LIC for T distribution regression analysis. *SoftwareX*, 30, 102132.

Chang, D., & Guo, G. Research on Distributed Redundant Data Estimation Based on LIC. *IAENG International Journal of Applied Mathematics*, 55(1), 1-6 (2025).

Gao, H., & Guo, G. LIC for Distributed Skewed Regression. *IAENG International Journal of Applied Mathematics*, 55(9), 2925-2930 (2025).

Zhang, C., & Guo, G. (2025). The optimal subset estimation of distributed redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 270-277.

Jing, G., & Guo, G. (2025). Student LIC for distributed estimation. *IAENG International Journal of Applied Mathematics*, 55(3), 575-581.

Liu, Q., & Guo, G. (2025). Distributed estimation of redundant data. *IAENG International Journal of Applied Mathematics*, 55(2), 332-337.

Examples

```
p = 6; n = 1000; K = 2; nk = 200; alpha = 0.05; sigma = 1
e = rnorm(n, 0, sigma); beta = c(sort(c(runif(p, 0, 1))));
data = c(rnorm(n * p, 5, 10)); X = matrix(data, ncol = p);
Y = X %*% beta + e;
LICnew(X = X, Y = Y, alpha = alpha, K = K, nk = nk)
```

Index