Package ‘EBEN’

Type Package
Title Empirical Bayesian Elastic Net
Version 5.1
Date 2023-5-28
Author Anhui Huang
Maintainer Anhui Huang <anhuihuang@gmail.com>
Description Provides the Empirical Bayesian Elastic Net for handling multicollinearity in generalized linear regression models. As a special case of the 'EBglmnet' package (also available on CRAN), this package encourages a grouping effects to select relevant variables and estimate the corresponding non-zero effects.
License GPL
Depends R (>= 2.10)
NeedsCompilation yes
Repository CRAN
Date/Publication 2023-05-29 16:40:02 UTC

R topics documented:

EBEN-package .......................................................... 2
BASIS ................................................................. 3
BASISbinomial ......................................................... 4
EBelasticNet.Binomial ............................................. 4
EBelasticNet.BinomialCV ......................................... 6
EBelasticNet.Gaussian ............................................... 7
EBelasticNet.GaussianCV ......................................... 8
EBlassoNEG.Binomial ............................................. 10
EBlassoNEG.BinomialCV ......................................... 11
EBlassoNEG.Gaussian ............................................ 13
EBlassoNEG.GaussianCV ......................................... 14
y ................................................................. 16
yBinomial ......................................................... 16

Index 18
Description

Fast EBEN algorithms.

EBEN implements a normal and generalized gamma hierarchical priors.

( **) Two parameters (alpha, lambda) are equivalent with elastic net priors.

( **) When parameter alpha = 1, it is equivalent with EBlasso-NE (normal + exponential)

Two models are available for both methods:

( **) General linear regression model.

( **) Logistic regression model.

Multi-collinearity:

( **) for group of high correlated or collinear variables: EBEN identifies the group of variables
estimates their effects together.

( **) group of variables can be selected together.

*Epistasis (two-way interactions) can be included for all models/priors

*model implemented with memory efficient c code.

*LAPACK/BLAS are used for most linear algebra computations.

Details

Package: EBEN
Type: Package
Version: 5.1
Date: 2015-10-06
License: gpl

Author(s)

Anhui Huang

References

key algorithms:

trait locus mapping. BMC Bioinformatics 12, 211.
trait locus mapping. BMC genetics 14(1):5.
locus mapping. Heredity 10.1038/hdy.2014.79

Other publications:

Huang, A., E. Martin, et al. (2014). "Detecting genetic interactions in pathway-based genome-wide

---

**BASIS**

An Example Data File for the Gauss Model

**Description**

This is a 1000x481 sample feature matrix

**Usage**

```r
data(BASIS)
```

**Format**

The format is: int [1:1000, 1:481] 0 -1 0 0 1 0 1 0 1 0 ...

**Details**

The data was simulated on a 2400cM chromosome, each column corresponded to an even spaced QTL

**Source**


**Examples**

```r
data(BASIS)
```
**EBelasticNet.Binomial**  

__Description__

This is a 500x481 sample feature matrix

__Usage__

```r
data(BASISbinomial)
```

__Format__

The format is: int [1:500, 1:481] 0 -1 0 0 0 0 -1 -1 0 1 ...  

__Details__

The data was simulated on a 2400cM chromosome, each column corresponded to an even spaced QTL

__Source__


__Examples__

```r
data(BASISbinomial)
```

---

**EBelasticNet.Binomial**  

_The EB Elastic Net Algorithm for Binomial Model with Normal-Gamma(NG) Prior Distribution_

__Description__

Generalized linear regression, normal-Gxponential (NG) hierarchical prior for regression coefficients

__Usage__

```r
EBelasticNet.Binomial(BASIS, Target, lambda, alpha, Epis = FALSE, verbose = 0)
```
Arguments

BASIS  sample matrix; rows correspond to samples, columns correspond to features
Target  Class label of each individual, TAKES VALUES OF 0 OR 1
lambda  Hyperparameter controls degree of shrinkage; can be obtained via Cross Validation; lambda>0
alpha   Hyperparameter controls degree of shrinkage; can be obtained via Cross Validation; 0<alpha<1
Epis    TRUE or FALSE for including two-way interactions
verbose 0 or 1; 1: display message; 0 no message

Details

If Epis=TRUE, the program adds two-way interaction of K*(K-1)/2 more columns to BASIS

Value

weight  the none-zero regression coefficients:
col1,col2 are the indices of the bases(main if equal);
col3: coefficient value;
col4: posterior variance;
col5: t-value;
col6: p-value
logLikelihood  log likelihood from the final regression coefficients
WaldScore  Wald Score
Intercept  Intercept
lambda  the hyperparameter; same as input lambda
alpha    thehyperparameter; same as input alpha

Author(s)

Anhui Huang; Dept of Electrical and Computer Engineering. Univ of Miami, Coral Gables, FL

References


Examples

library(EBEN)
data(BASISbinomial)
data(yBinomial)
#reduce sample size to speed up the running time
n = 50;
k = 100;
N = length(yBinomial);
set = sample(N,n);
BASIS = BASISbinomial[set,1:k];
y = yBinomial[set];
output = EBelasticNet.Binomial(BASIS, y, lambda = 0.1, alpha = 0.5, Epis = FALSE, verbose = 5)

---

**EBelasticNet.BinomialCV**

*Cross Validation (CV) Function to Determine Hyperparameter of the EB Elastic Net Algorithm for Binomial Model with Normal-Gamma (NG) Prior Distribution*

**Description**

Hyperparameter controls degree of shrinkage, and is obtained via Cross Validation (CV). This program calculates the maximum lambda that allows one non-zero basis; and performs a search down to 0.001*lambda_max at even steps. (20 steps)

**Usage**

`EBelasticNet.BinomialCV(BASIS, Target, nFolds, foldId, Epis = FALSE, verbose = 0)`

**Arguments**

- **BASIS**: sample matrix; rows correspond to samples, columns correspond to features
- **Target**: Class label of each individual, TAKES VALUES OF 0 OR 1
- **nFolds**: number of n-fold cv
- **Epis**: TRUE or FALSE for including two-way interactions
- **foldId**: random assign samples to different folds
- **verbose**: from 0 to 5; larger verbose displays more messages

**Details**

If Epis=TRUE, the program adds two-way interaction K*(K-1)/2 more columns to BASIS

**Value**

- **CrossValidation**: col1: hyperparameter; col2: loglikelihood mean; standard ERROR of nfold mean log likelihood
- **Lambda_optimal**: the optimal hyperparameter as computed
- **Alpha_optimal**: the optimal hyperparameter as computed

**Author(s)**

Anhui Huang; Dept of Electrical and Computer Engineering, Univ of Miami, Coral Gables, FL
EBelasticNet.Gaussian

References


Examples

```r
## not run
library(EBEN)
data(BASISbinomial)
data(yBinomial)
# reduce sample size to speed up the running time
n = 50;
k = 100;
N = length(yBinomial);
set.seed(1)
set = sample(N,n);
BASIS = BASISbinomial[set,1:k];
y = yBinomial[set];
nFolds = 3
## Not run:
CV = EBelasticNet.BinomialCV(BASIS, y, nFolds = 3,Epis = FALSE)
## End(Not run)
```

EBelasticNet.Gaussian  *The EB Elastic Net Algorithm for Gaussian Model*

Description

General linear regression, normal-Gamma (NG) hierarchical prior for regression coefficients

Usage

```r
EBelasticNet.Gaussian(BASIS, Target, lambda, alpha,Epis = FALSE,verbose = 0)
```

Arguments

- **BASIS**: sample matrix; rows correspond to samples, columns correspond to features
- **Target**: Response each individual
- **lambda**: Hyperparameter controls degree of shrinkage; can be obtained via Cross Validation; lambda>0
- **alpha**: Hyperparameter controls degree of shrinkage; can be obtained via Cross Validation; 0<alpha<1
- **Epis**: TRUE or FALSE for including two-way interactions
- **verbose**: 0 or 1; 1: display message; 0 no message
Details

If Epis=TRUE, the program adds two-way interaction of K*(K-1)/2 more columns to BASIS

Value

weight the none-zero regression coefficients:
   col1,col2 are the indices of the bases(main if equal);
   col3: coefficient value;
   col4: posterior variance;
   col5: t-value;
   col6: p-value

WaldScore Wald Score

Intercept Intercept

lambda the hyperparameter; same as input lambda

alpha the hyperparameter; same as input alpha

Author(s)

Anhui Huang; Dept of Electrical and Computer Engineering, Univ of Miami, Coral Gables, FL

References


Examples

library(EBEN)
data(BASIS)
data(y)
n = 50;
k = 100;
BASIS = BASIS[1:n,1:k];
y = y[1:n];
Blup = EBelasticNet.Gaussian(BASIS, y,lambda = 0.0072, alpha = 0.95, Epis = FALSE, verbose = 0)
betas = Blup$weight
betal

EBelastinNet.GaussianCV

Cross Validation (CV) Function to Determine Hyperparameters of the EBEN Algorithm for Gaussian Model
EBelasticNet.GaussianCV

Description

Hyperparameter controls degree of shrinkage, and is obtained via Cross Validation (CV). This program calculates the maximum lambda that allows one non-zero basis; and performs a search down to 0.0001*lambda_max at even steps. (20 steps)

Usage

EBelasticNet.GaussianCV(BASIS, Target, nFolds, foldId, Epis = FALSE, verbose = 0)

Arguments

BASIS sample matrix; rows correspond to samples, columns correspond to features
Target Response each individual
nFolds number of n-fold cv
Epis TRUE or FALSE for including two-way interactions
foldId random assign samples to different folds
verbose from 0 to 5; larger verbose displays more messages

Details

If Epis = TRUE, the program adds two-way interaction K*(K-1)/2 more columns to BASIS

Value

CrossValidation
  col1: hyperparameter; col2: loglikelihood mean; standard ERROR of nfold mean log likelihood
  Lambda_optimal the optimal hyperparameter as computed
  Alpha_optimal the optimal hyperparameter as computed

Author(s)

Anhui Huang; Dept of Electrical and Computer Engineering, Univ of Miami, Coral Gables, FL

References


Examples

library(EBEN)
data(BASIS)
data(y)
# reduce sample size to speed up the running time
n = 50;
k = 100;
BASIS = BASIS[1:n,1:k];
The EBlasso Algorithm for Binomial Model with Normal-Exponential-Gamma (NEG) Prior Distribution

Description

Generalized linear regression, normal-exponential-gamma (NEG) hierarchical prior for regression coefficients

Usage

EBlassoNEG.Binomial(BASIS, Target, a_gamma, b_gamma, Epis, verbose, group)

Arguments

- **BASIS**: sample matrix; rows correspond to samples, columns correspond to features
- **Target**: Class label of each individual, TAKES VALUES OF 0 OR 1
- **a_gamma**: Hyperparameters control degree of shrinkage; can be obtained via Cross Validation; \(a_{\gamma} \geq -1\)
- **b_gamma**: Hyperparameters control degree of shrinkage; can be obtained via Cross Validation; \(b_{\gamma} > 0\)
- **Epis**: TRUE or FALSE for including two-way interactions
- **verbose**: 0 or 1; 1: display message; 0 no message
- **group**: 0 or 1; 0: No group effect; 1 two-way interaction grouped. Only valid when Epis = TRUE

Details

If Epis=TRUE, the program adds two-way interaction \(K^*(K-1)/2\) more columns to BASIS

Value

- **weight**: the none-zero regression coefficients:
  - col1, col2 are the indices of the bases (main if equal);
  - col3: coefficient value;
  - col4: posterior variance;
  - col5: t-value;
  - col6: p-value
- **logLikelihood**: log likelihood with the final regression coefficients
EBlassoNEG.BinomialCV

WaldScore       Wald Score
Intercept       Intercept
a_gamma         the hyperparameter; same as input
b_gamma         the hyperparameter; same as input

Author(s)

Anhui Huang; Dept of Electrical and Computer Engineering, Univ of Miami, Coral Gables, FL

References


Examples

```r
library(EBEN)
data(BASISbinomial)
data(yBinomial)

# reduce sample size to speed up the running time
n = 50;
k = 100;
BASIS = BASISbinomial[1:n,1:k];
y = yBinomial[1:n];
output = EBlassoNEG.Binomial(BASIS,y,0.1,0.1,Epis = FALSE)
```

EBlassoNEG.BinomialCV  Cross Validation (CV) Function to Determine Hyperparameters of the EBlasso Algorithm for Binomial Model with Normal-Exponential-Gamma (NEG) Prior Distribution

Description

Hyperparameters control degree of shrinkage, and are obtained via Cross Validation. This program performs three steps of CV:
1st: $a = b = 0.001, 0.01, 0.1, 1$
2nd: fix $b=b_1$; $a=[-0.5, -0.4, -0.3, -0.2, -0.1, -0.01, 0.01, 0.05, 0.1, 0.5, 1]$
3rd: fix $a = a_2$; $b=0.01$ to 10 with a step size of one for $b > 1$ and a step size of one on the logarithmic scale for $b < 1$

In the 2nd step, $a$ can take value from -1 and values in $[-1, -0.5]$ can be added to the set in line 13 of this function (The smaller $a$ is, the less shrinkage.)

Usage

```r
EBlassoNEG.BinomialCV(BASIS, Target, nFolds, foldId, Epis, verbose, group)
```
Arguments

BASIS sample matrix; rows correspond to samples, columns correspond to features
Target Class label of each individual, TAKES VALUES OF 0 OR 1
nFolds number of n-fold cv
foldId random assign samples to different folds
Epis TRUE or FALSE for including two-way interactions
verbose from 0 to 5; larger verbose displays more messages
group TRUE or FALSE; FALSE: No group effect; TRUE two-way interaction grouped. Only valid when Epis = TRUE

Details

If Epis=TRUE, the program adds two-way interaction \( K \times (K-1)/2 \) more columns to BASIS
Note: Given the fact that degree of shrinkage is a monotonic function of \((a,b)\),
The function implemented a 3-step search as described in Huang, A. 2014, for full
grid search, user needs to modify the function accordingly.

Value

CrossValidation
  col1: hyperparameters; col2: loglikelihood mean; standard ERROR of nfold mean log likelihood
  a_optimal the optimal hyperparameter as computed
  b_optimal the optimal hyperparameter as computed

Author(s)

Anhui Huang; Dept of Electrical and Computer Engineering, Univ of Miami, Coral Gables, FL

References


Examples

library(EBEN)
data(BASISbinomial)
data(yBinomial)
#reduce sample size to speed up the running time
n = 50;
k = 100;
BASIS = BASISbinomial[1:n,1:k];
y = yBinomial[1:n];
## Not run:
CV = EBlassoNEG.BinomialCV(BASIS, y, nFolds = 3,Epis = FALSE, verbose = 0)
EBlassoNEG.Gaussian

The EBlasso Algorithm for Gaussian Model with Normal-Exponential-Gamma (NEG) Prior Distribution

Description

General linear regression, normal-exponential-gamma (NEG) hierarchical prior for regression coefficients

Usage

EBlassoNEG.Gaussian(BASIS, Target, a_gamma, b_gamma, Epis, verbose, group)

Arguments

BASIS
sample matrix; rows correspond to samples, columns correspond to features

Target
Response each individual

a_gamma
Hyperparameters control degree of shrinkage; can be obtained via Cross Validation

b_gamma
Hyperparameters control degree of shrinkage; can be obtained via Cross Validation

Epis
TRUE or FALSE for including two-way interactions

verbose
from 0 to 5; larger verbose displays more messages

group
0 or 1; 0: No group effect; 1 two-way interaction grouped. Only valid when Epis = TRUE

Details

If Epis=TRUE, the program adds two-way interaction $K*(K-1)/2$ more columns to BASIS for memory efficient, the function pass n_effect to C. n_effect > n_true effects, which is a rough guess on how many variables will be selected by the function by providing a relative 'small' n_effect, the function will not allocate a large trunk of memory during computation.

Value

weight
the none-zero regression coefficients:
col1,col2 are the indices of the bases(main if equal);
col3: coefficient value;
col4: posterior variance;
col5: t-value;
col6: p-value
EBlassoNEG.GaussianCV

Cross Validation (CV) Function to Determine Hyperparameters of the EBlasso Algorithm for Gaussian Model with Normal-Exponential-Gamma (NEG) Prior Distribution

Description

Hyperparameters control degree of shrinkage, and are obtained via Cross Validation. This program performs three steps of CV.
1st: \( a = b = 0.001, 0.01, 0.1, 1; \)
2nd: fix \( b = b1; \) \( a = [-0.5, -0.4, -0.3, -0.2, -0.1, -0.01, 0.01, 0.05, 0.1, 0.5, 1]; \)
3rd: fix \( a = a2; \) \( b = 0.01 \) to \( 10 \) with a step size of one for \( b > 1 \) and a step size of one on the logarithmic scale for \( b < 1 \)

In the 2nd step, \( a \) can take value from \(-1\) and values in \([-1,-0.5]\) can be added to the set in line 13 of this function (The smaller \( a \) is, the less shrinkage.)

Usage

\[
\text{EBlassoNEG.GaussianCV(BASIS, Target, nFolds, foldId, Epis, verbose, group)}
\]
Arguments

BASIS sample matrix; rows correspond to samples, columns correspond to features
Target Class label of each individual, TAKES VALUES OF 0 OR 1
nFolds number of n-fold cv
foldId random assign samples to different folds
Epis TRUE or FALSE for including two-way interactions
verbose from 0 to 5; larger verbose displays more messages
group TRUE or FALSE; FALSE: No group effect; TRUE two-way interaction grouped. Only valid when Epis = TRUE

Details

If Epis= TRUE, the program adds two-way interaction K*(K-1)/2 more columns to BASIS
Note: Given the fact that degree of shrinkage is a monotonic function of (a,b), The function implemented a 3-step search as described in Huang, A. 2014, for full grid search, user needs to modify the function accordingly.

Value

CrossValidation
  col1: hyperparameters; col2: loglikelihood mean; standard ERROR of nfold mean log likelihood
  a_optimal the optimal hyperparameter as computed
  b_optimal the optimal hyperparameter as computed

Author(s)

Anhui Huang; Dept of Electrical and Computer Engineering, Univ of Miami, Coral Gables, FL

References


Examples

library(EBEN)
data(BASIS)
data(y)
#reduce sample size to speed up the running time
n = 50;
k = 100;
BASIS = BASIS[1:n,1:k];
y = y[1:n];
## Not run:
CV = EBlassoNEG.GaussianCV(BASIS, y, nFolds = 3,Epis = FALSE)
Sample Response Data for Gaussian Model

Description
Corresponding to the response of BASIS

Usage
`data(y)`

Format
The format is: num [1:1000, 1] 113.5 97.1 116.6 96.7 105.5 ...

Source

Examples
`data(y)`

Sample Variable Data for Binomial Model

Description
Corresponding to the class label of BASISbinomial

Usage
`data(yBinomial)`

Format
The format is: int [1:500, 1] 1 1 1 1 1 1 1 1 1 1 ...

Source
yBinomial

Examples

data(BASISbinomial)
Index

* datasets
  BASIS, 3
  BASISbinomial, 4
  y, 16
  yBinomial, 16

* package
  EBEN-package, 2

BASIS, 3
BASISbinomial, 4

EBelasticNet.Binomial, 4
EBelasticNet.BinomialCV, 6
EBelasticNet.Gaussian, 7
EBelasticNet.GaussianCV, 8
EBEN (EBEN-package), 2
EBEN-package, 2
EBlassoNEG.Binomial, 10
EBlassoNEG.BinomialCV, 11
EBlassoNEG.Gaussian, 13
EBlassoNEG.GaussianCV, 14

y, 16
yBinomial, 16