Package 'EBEN'

October 7, 2015

Type Package

Title Empirical Bayesian Elastic Net

Version 4.6

Date 2015-10-06

Author Anhui Huang

Maintainer Anhui Huang <a.huang1@umiami.edu>

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 2015-10-07 09:13:08

R topics documented:

EBEN-package *Empirical Bayesian Elastic Net (EBEN)*

Description

Fast EBEN algorithms.
EBEN implements a normal and generalized gamma hierearchical 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:	4.6
Date:	2015-10-06
License:	gpl

Author(s)

Anhui Huang

References

key algorithms:
Cai, X., Huang, A., and Xu, S. (2011). Fast empirical Bayesian LASSO for multiple quantitative trait locus mapping. BMC Bioinformatics 12, 211.
Huang A, Xu S, Cai X. (2013). Empirical Bayesian LASSO-logistic regression for multiple binary trait locus mapping. BMC genetics 14(1):5.
Huang, A., Xu, S., and Cai, X. (2014). Empirical Bayesian elastic net for multiple quantitative trait 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

association studies." Genet Epidemiol 38(4): 300-309.
Huang, A., S. Xu, et al. (2014). "Whole-genome quantitative trait locus mapping reveals major role of epistasis on yield of rice." PLoS ONE 9(1): e87330.
Huang, A. (2014). "Sparse model learning for inferring genotype and phenotype associations." Ph.D Dissertation. University of Miami(1186).

BASIS *An Example Data File for the Gauss Model*

Description

This is a 1000x481 sample feature matrix

Usage

```
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

Huang, A., Xu, S., and Cai, X. (2014). Empirical Bayesian elastic net for multiple quantitative trait locus mapping. Heredity 10.1038/hdy.2014.79

Examples

```
data(BASIS)
```

BASISbinomial *An Example Data File for the Binomial Model*

Description

This is a 500x481 sample feature matrix

Usage

```
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

Huang A, Xu S, Cai X: Empirical Bayesian LASSO-logistic regression for multiple binary trait locus mapping. BMC genetics 2013, 14(1):5.

Examples

```
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

```
EBelasticNet.Binomial(BASIS, Target, lambda, alpha,Epis = "no",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	"yes" or "no" for including two-way interactions
verbose	0 or 1; 1: display message; 0 no message

Details

If Epis="yes", 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: coefficent 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	the hyperparameter; same as input alpha

Author(s)

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

References

Huang A, Xu S, Cai X: Empirical Bayesian LASSO-logistic regression for multiple binary trait locus mapping. BMC genetics 2013, 14(1):5.

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 = "no",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, Epis = "no",foldId)
```

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	"yes" or "no" for including two-way interactions
foldId	random assign samples to different folds

Details

If Epis="yes", 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
Lmabda_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

Huang A, Xu S, Cai X: Empirical Bayesian LASSO-logistic regression for multiple binary trait locus mapping. BMC genetics 2013, 14(1):5.

Examples

```
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
CV = EBelasticNet.BinomialCV(BASIS, y, nFolds = 3,Epis = "no")
```

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

Description

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

Usage

```
EBelasticNet.Gaussian(BASIS, Target, lambda, alpha,Epis = "no",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	"yes" or "no" for including two-way interactions
verbose	0 or 1; 1: display message; 0 no message

Details

If Epis="yes", 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: coefficent 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

Huang, A., Xu, S., and Cai, X. (2014). Empirical Bayesian elastic net for multiple quantitative trait locus mapping. Heredity 10.1038/hdy.2014.79

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 = "no",verbose = 0)
betas  = Blup$weight
betas
```

EBelasticNet.GaussianCV

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

---

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, Epis = "no",foldId)
```

Arguments

BASIS	sample matrix; rows correspond to samples, columns correspond to features
Target	Response each individual
nFolds	number of n-fold cv
Epis	"yes" or "no" for including two-way interactions
foldId	random assign samples to different folds

Details

If Epis="yes", 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
Lmabda_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

Huang, A., Xu, S., and Cai, X. (2013). Empirical Bayesian elastic net for multiple quantitative trait locus mapping. submitted.

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];
CV = EBelasticNet.GaussianCV(BASIS, y, nFolds = 3,Epis = "no")
```

y — *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

Huang, A., Xu, S., and Cai, X. (2014). Empirical Bayesian elastic net for multiple quantitative trait locus mapping. Heredity 10.1038/hdy.2014.79

Examples

```
data(y)
```

yBinomial — *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

Huang A, Xu S, Cai X: Empirical Bayesian LASSO-logistic regression for multiple binary trait locus mapping. BMC genetics 2013, 14(1):5.

Examples

```
data(BASISbinomial)
```

Index