Package 'Bayenet'

April 28, 2023

Type Package

Title Bayesian Quantile Elastic Net for Genetic Study

Version 1.0

Date 2023-04-17

Author Xi Lu

Maintainer Xi Lu <xilu@ksu.edu>

Description As heavy-tailed error distribution and outliers in the response variable widely exist, models which are robust to data contamination are highly demanded. Here, We develop a novel robust Bayesian variable selection method with elastic net penalty for quantile regression in genetic analysis. In particular, the spike-and-slab priors have been incorporated to impose sparsity. An efficient Gibbs sampler has been developed to facilitate computation. The algorithms of the proposed and alternative methods are efficiently implemented in 'C++'.

Depends R (>= 3.5.0)

License GPL-2

Encoding UTF-8

LazyData true

LinkingTo Rcpp, RcppArmadillo

Imports Rcpp,
stats,
MCMCpack,
base,
gsl,
VGAM,
MASS,
hbmem,
SuppDists

RoxygenNote 7.2.3

NeedsCompilation yes

Repository CRAN

R topics documented:

`Bayenet-package` *Bayesian Quantile Elastic Net for Genetic Study*

Description

In this package, we provide a set of robust Bayesian quantile variable selection methods for genetic analysis. A Bayesian formulation of the quantile regression has been adopted to accommodate data contamination and heavy-tailed distributions in the response. The proposed method conducts a robust quantile variable selection by accounting for structural sparsity. In particular, the spike-and-slab priors are imposed to identify important genetic effects. In addition to the default method, users can also choose different structures (robust or non-robust) and penalty (lasso or elastic net) with or without spike-and-slab priors.

Details

The user friendly, integrated interface **Bayenet()** allows users to flexibly choose the fitting methods they prefer. There are three arguments in Bayenet() that control the fitting method: robust: whether to use robust methods; sparse: whether to use the spike-and-slab priors to create sparsity; penalty: use lasso or elastic net as penalty. The function Bayenet() returns a Bayenet object that contains the posterior estimates of each coefficients. predict.Bayenet() and print.Bayenet() are implemented for Bayenet objects. predict.Bayenet() takes a Bayenet object and returns the predicted values for new observations.

References

Lu, X. and Wu, C. (2023). Bayesian quantile elastic net with spike-and-slab priors.

Lu, X., Fan, K., Ren, J., and Wu, C. (2021). Identifying Gene–Environment Interactions With Robust Marginal Bayesian Variable Selection. *Frontiers in Genetics*, 12:667074 doi:10.3389/fgene.2021.667074

Zhou, F., Ren, J., Lu, X., Ma, S. and Wu, C. (2020). Gene–Environment Interaction: a Variable Selection Perspective. Epistasis. Methods in Molecular Biology. *Humana Press* (Accepted) https://arxiv.org/abs/2003.02930

Wu, C., Cui, Y., and Ma, S. (2014). Integrative analysis of gene–environment interactions under a multi–response partially linear varying coefficient model. *Statistics in Medicine*, 33(28), 4988–4998 doi:10.1002/sim.6287

Li, Q. and Lin, N. (2010). The Bayesian elastic net. *Bayesian Anal*, 5(1): 151-170 doi:10.1214/10BA506

Li, Q., Xi, R. and Lin, N. (2010). The Bayesian regularized quantile regression. *Bayesian Analysis*, 5(3): 533-556 doi::10.1214/10BA521

See Also

`Bayenet`

Bayenet *fit a robust Bayesian elastic net variable selection model for genetic study.*

Description

fit a robust Bayesian elastic net variable selection model for genetic study.

Usage

```
Bayenet(
  X,
  Y,
  clin,
  max.steps = 10000,
  robust = TRUE,
  sparse = TRUE,
  penalty = c("lasso", "elastic net"),
  debugging = FALSE
)
```

Arguments

X	the matrix of predictors (genetic factors). Each row should be an observation vector.
Y	the continuous response variable.
clin	a matrix of clinical variables. Clinical variables are not subject to penalty. Clinical variables will be centered and a column of 1 will be added to the Clinical matrix as the intercept.
max.steps	the number of MCMC iterations.
robust	logical flag. If TRUE, robust methods will be used.
sparse	logical flag. If TRUE, spike-and-slab priors will be used to shrink coefficients of irrelevant covariates to zero exactly.
penalty	two choices are available. "lasso" for lasso penalty. "elastic net" for elastic net penalty.
debugging	logical flag. If TRUE, progress will be output to the console and extra information will be returned.

Details

Consider the data model described in "dat":

$$Y_i = \alpha_0 + \sum_{k=1}^{q} \gamma_k C_{ik} + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i,$$

where α_0 is the intercept, γ_k's and β_j's are the regression coefficients corresponding to effects of clinical factors and genetic variants, respectively.

When penalty="elastic net" (default), the elastic net penalty is adopted. If penalty="lasso", the lasso penalty is used.

When sparse=TRUE (default), spike–and–slab priors are imposed to identify important main and interaction effects. If sparse=FALSE, Laplacian shrinkage will be used.

When robust=TRUE (default), the distribution of ϵ_i is defined as a Laplace distribution with density $f(\epsilon_i|\nu) = \frac{\nu}{2}\exp\{-\nu|\epsilon_i|\}$, $(i = 1, \ldots, n)$, which leads to a Bayesian formulation of LAD regression. If robust=FALSE, ϵ_i follows a normal distribution.

Both X and $clin$ will be standardized before the generation of interaction terms to avoid the multicollinearity between main effects and interaction terms.

Please check the references for more details about the prior distributions.

Value

an object of class 'Bayenet' is returned, which is a list with component:

`posterior`	the posterior samples of coefficients from the MCMC.
`coefficient`	the estimated value of coefficients.
`burn.in`	the total number of burn-ins.
`iterations`	the total number of iterations.
`design`	the design matrix of all effects.

References

Lu, X. and Wu, C. (2023). Bayesian quantile elastic net with spike-and-slab priors.

#' @examples data(dat)

max.steps=5000 fit= Bayenet(X, Y, clin, max.steps, penalty="lasso")

coefficients of parameters fit$coefficient

Estimated values of main G effects fit$coefficient$G

Estimated values of clincal effects fit$coefficient$clin

`dat` *simulated data for demonstrating the features of Bayenet.*

Description

Simulated gene expression data for demonstrating the features of Bayenet.

Usage

```
data("dat")
```

Format

dat consists of four components: X, Y, clin, coef.

Details

The data model for generating Y

Use subscript i to denote the ith subject. Let $(Y_i, X_i, clin_i)$ $(i = 1, \ldots, n)$ be independent and identically distributed random vectors. Y_i is a continuous response variable representing the cancer outcome and disease phenotype. X_i is the p–dimensional vector of genetic factors. The clinical factors is denoted as the q-dimensional vector $clin_i$. The ϵ follows some heavy-tailed distribution. Considering the following model:

$$Y_i = \alpha_0 + \sum_{k=1}^{q} \gamma_k C_{ik} + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i,$$

where α_0 is the intercept, γ_k's and β_j's are the regression coefficients corresponding to effects of clinical factors and genetic variants, respectively. Denote $\gamma = (\gamma_1, \ldots, \gamma_q)^T$, $\beta = (\beta_1, \ldots, \beta_p)^T$. Then model can be written as

$$Y_i = C_i\gamma + X_i\beta + \epsilon_i.$$

See Also

Bayenet

Examples

```
data(dat)
dim(X)
```

`predict.Bayenet` *make predictions from a Bayenet object*

Description

make predictions from a Bayenet object

Usage

```
## S3 method for class 'Bayenet'
predict(object, X.new, clin.new, Y.new, ...)
```

Arguments

`object`	Bayenet object.
`X.new`	a matrix of new values for X at which predictions are to be made.
`clin.new`	a vector or matrix of new values for clin at which predictions are to be made.
`Y.new`	a vector of the response of new observations. If provided, the prediction error will be computed based on Y.new.
`...`	other predict arguments

Details

X.new must have the same number of columns as X used for fitting the model. If clin was provided when fit the model, clin.new must not be NULL, and vice versa. The predictions are made based on the posterior estimates of coefficients in the Bayenet object. Note that the effects of clinical factors are not subject to selection.

If Y.new is provided, the prediction error will be computed. For robust methods, the prediction mean absolute deviations (PMAD) will be computed. For non-robust methods, the prediction mean squared error (PMSE) will be computed.

Value

an object of class 'Bayenet.pred' is returned, which is a list with components:

`error`	prediction error. error is NULL is Y.new=NULL.
`y.pred`	predicted values of the new observations.

See Also

`Bayenet`

`print.Bayenet` *print a Bayenet object*

Description

Print a summary of a Bayenet object

Usage

```
## S3 method for class 'Bayenet'
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

`x`	Bayenet object.
`digits`	significant digits in printout.
`...`	other print arguments.

Value

No return value, called for side effects.

See Also

`Bayenet`

`print.Bayenet.pred` *print a predict.Bayenet object*

Description

Print a summary of a predict.Bayenet object

Usage

```
## S3 method for class 'Bayenet.pred'
print(x, digits = max(3, getOption("digits") - 3), ...)
```

Arguments

`x`	predict.Bayenet object.
`digits`	significant digits in printout.
`...`	other print arguments.

Value

No return value, called for side effects.

See Also

`predict.Bayenet`

Index