Package 'netReg'

February 9, 2020

Type Package

Title Network-Regularized Regression Models

Version 1.11.0

Date 2018-04-28

Maintainer Simon Dirmeier <simon.dirmeier@web.de>

Description netReg fits linear regression models using network-penalization. Graph prior knowledge, in the form of biological networks, is being incorporated into the loss function of the linear model. The networks describe biological relationships such as co-regulation or dependency of the same transcription factors/metabolites/etc. yielding a part sparse and part smooth solution for coefficient profiles.

URL https://github.com/dirmeier/netReg

BugReports https://github.com/dirmeier/netReg/issues

Depends R(>= 3.4)

biocViews Software, StatisticalMethod, Regression, FeatureExtraction, Network, GraphAndNetwork

License GPL-3 | BSL-1.0 + file LICENSE

Encoding UTF-8

Suggests testthat, knitr, rmarkdown, BiocStyle, lintr, lassoshooting

VignetteBuilder knitr

RoxygenNote 6.0.1

SystemRequirements C++11

LinkingTo Rcpp, RcppArmadillo

Imports Rcpp, stats

NeedsCompilation yes

git_url https://git.bioconductor.org/packages/netReg

git_branch master

git_last_commit a30652f

git_last_commit_date 2019-10-29

Date/Publication 2020-02-08

Author Simon Dirmeier [aut, cre]

R topics documented:

`netReg-package` *netReg*

Description

netReg is a package for generalized linear regression that includes prior graphs in the models objective function.

Details

netReg uses *Armadillo*, *OpenBLAS*, *BLAS* and *LAPACK* for fast matrix computations and *Dlib* for constrained derivate-free optimization.

Author(s)

Simon Dirmeier | <simon.dirmeier@gmx.de>

References

Dirmeier, Simon and Fuchs, Christiane and Mueller, Nikola S and Theis, Fabian J (2018), netReg: Network-regularized linear models for biological association studies.
Bioinformatics

Friedman J., Hastie T., Hoefling H. and Tibshirani R. (2007), Pathwise coordinate optimization.
The Annals of Applied Statistics

Friedman J., Hastie T. and Tibshirani R. (2010), Regularization Paths for Generalized Linear Models via Coordinate Descent.
Journal of Statistical Software

Fu W. J. (1998), Penalized Regression: The Bridge Versus the Lasso.
Journal of Computational and Graphical Statistics

Cheng W. and Wang W. (2014), Graph-regularized dual Lasso for robust eQTL mapping.
Bioinformatics

Powell M.J.D. (2009), The BOBYQA algorithm for bound constrained optimization without derivatives.
http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf

cv.edgenet *Find the optimal shrinkage parameters for edgenet*

Description

Finds the optimal shrinkage parameters using cross-validation for edgenet. We use the BOBYQA algorithm to find the optimial regularization parameters and coordinate descent in order to minimize the objective function of the linear model.

Usage

```
cv.edgenet(X, Y, G.X = NULL, G.Y = NULL, lambda = NULL, psigx = NULL,
  psigy = NULL, thresh = 1e-05, maxit = 1e+05, family = c("gaussian"),
  optim.epsilon = 0.001, optim.maxit = 10000, nfolds = 10)
```

Arguments

X	input matrix, of dimension (n x p) where n is the number of observations and p is the number of covariables. Each row is an observation vector.
Y	output matrix, of dimension (n x q) where n is the number of observations and q is the number of response variables Each row is an observation vector.
G.X	non-negativ affinity matrix for n, of dimensions (p x p) where p is the number of covariables X. Providing a graph G.X will optimize the regularization parameter psi.gx. If this is not desired just set G.X to NULL.
G.Y	non-negativ affinity matrix for n, of dimensions (q x q) where q is the number of responses Y. Providing a graph G.Y will optimize the regularization parameter psi.gy. If this is not desired just set G.Y to NULL.
lambda	numerical shrinkage parameter for LASSO. Per default this parameter is set to NULL which means that lambda is going to be estimated using cross-validation. If any numerical value for lambda is set, estimation of the optimal parameter will *not* be conducted.
psigx	numerical shrinkage parameter for graph-regularization of G.X. Per default this parameter is set to NULL which means that psigx is going to be estimated in the cross-validation. If any numerical value for psigx is set, estimation of the optimal parameter will *not* be conducted.
psigy	numerical shrinkage parameter for graph-regularization of G.Y. Per default this parameter is set to NULL which means that psigy is going to be estimated in the cross-validation. If any numerical value for psigy is set, estimation of the optimal parameter will *not* be conducted.
thresh	numerical threshold for coordinate descent
maxit	maximum number of iterations for the coordinate descent (integer)
family	family of response, e.g. *gaussian*
optim.epsilon	numerical threshold criterion for the optimization to stop. Usually 1e-3 is a good choice.
optim.maxit	the maximum number of iterations for the optimization (integer). Usually 1e4 is a good choice.
nfolds	the number of folds to be used - default is 10 (minimum 3, maximum nrow(X)).

Value

An object of class `cv.edgenet`

`call`	the call that produced the object
`lambda`	the estimated (p x q)-dimensional coefficient matrix B.hat
`psigx`	the estimated (q x 1)-dimensional vector of intercepts
`psigy`	the estimated (q x 1)-dimensional vector of intercepts

References

Dirmeier, Simon and Fuchs, Christiane and Mueller, Nikola S and Theis, Fabian J (2018), netReg: Network-regularized linear models for biological association studies.
Friedman J., Hastie T., Hoefling H. and Tibshirani R. (2007), Pathwise coordinate optimization. *The Annals of Applied Statistics*

Friedman J., Hastie T. and Tibshirani R. (2010), Regularization Paths for Generalized Linear Models via Coordinate Descent.
Journal of Statistical Software

Fu W. J. (1998), Penalized Regression: The Bridge Versus the Lasso.
Journal of Computational and Graphical Statistics

Cheng W. and Wang W. (2014), Graph-regularized dual Lasso for robust eQTL mapping.
Bioinformatics

Powell M.J.D. (2009), The BOBYQA algorithm for bound constrained optimization without derivatives.
http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf

Examples

```
X <- matrix(rnorm(100*10), 100, 10)
b <- rnorm(10)
G.X <- matrix(rpois(10*10,1),10)
G.X <- t(G.X) + G.X
diag(G.X) <- 0

# fit a Gaussian model
Y <- X%*%b + rnorm(100)
cv.edge <- cv.edgenet(X=X, Y=Y, G.X=G.X, family="gaussian")
```

---

edgenet *Fit a graph-regularized linear regression model using edge-based regularization.*

Description

Fit a graph-regularized linear regression model using edge-penalization. The coefficients are computed using graph-prior knowledge in the form of one/two affinity matrices. Graph-regularization is an extension to previously introduced regularization techniques, such as the LASSO. For that reason we are also using coordinate descent for minimization of the objective function of the linear model.

Usage

```
edgenet(X, Y, G.X = NULL, G.Y = NULL, lambda = 1, psigx = 1,
  psigy = 1, thresh = 1e-05, maxit = 1e+05, family = c("gaussian"))
```

Arguments

X	input matrix, of dimension (n x p) where n is the number of observations and p is the number of covariables. Each row is an observation vector.
Y	output matrix, of dimension (n x q) where n is the number of observations and q is the number of response variables. Each row is an observation vector.
G.X	non-negativ affinity matrix for n, of dimensions (p x p) where p is the number of covariables X
G.Y	non-negativ affinity matrix for n, of dimensions (q x q) where q is the number of responses Y
lambda	numerical shrinkage parameter for LASSO.
psigx	numerical shrinkage parameter for graph-regularization of G.X
psigy	numerical shrinkage parameter for graph-regularization of G.Y
thresh	numerical threshold for coordinate descent
maxit	maximum number of iterations for coordinate descent (integer)
family	family of response, e.g. *gaussian*

Value

An object of class edgenet

coefficients	the estimated (p x q)-dimensional coefficient matrix B.hat
intercept	the estimated (q x 1)-dimensional vector of intercepts
call	the call that produced the object
family	the family of the response

References

Dirmeier, Simon and Fuchs, Christiane and Mueller, Nikola S and Theis, Fabian J (2018), netReg: Network-regularized linear models for biological association studies.
Bioinformatics

Friedman J., Hastie T., Hoefling H. and Tibshirani R. (2007), Pathwise coordinate optimization.
The Annals of Applied Statistics

Friedman J., Hastie T. and Tibshirani R. (2010), Regularization Paths for Generalized Linear Models via Coordinate Descent.
Journal of Statistical Software

Fu W. J. (1998), Penalized Regression: The Bridge Versus the Lasso.
Journal of Computational and Graphical Statistics

Cheng W. and Wang W. (2014), Graph-regularized dual Lasso for robust eQTL mapping.
Bioinformatics

Examples

```
X <- matrix(rnorm(100*10), 100, 10)
b <- rnorm(10)
G.X <- matrix(rpois(100,1), 10)
G.X <- t(G.X) + G.X
diag(G.X) <- 0

# fit a Gaussian model
Y <- X%*%b + rnorm(100)
fit <- edgenet(X=X, Y=Y, G.X=G.X, family="gaussian")
```

`predict.gaussian.edgenet` *Predict method for gaussian edgenet fits*

Description

Predicts the estimated Y.hat values for a newdata design matrix X similar to the other predict methods, e.g. from glm and glmnet

Usage

```
## S3 method for class 'gaussian.edgenet'
predict(object, newdata = NULL, ...)
```

Arguments

`object`	a fitted object of class *gaussian.edgenet*
`newdata`	a new (`m` x `p`)-dimensional design matrix with a variable number of observations `m`, but a constant number of co-variables `p`
`...`	further arguments

Value

A (`m` x `q`)-dimensional matrix

Examples

```
## Not run:
X <- matrix(rnorm(100*10),100,10)
G.X <- matrix(rpois(10*10,1),10)
G.X <- t(G.X) + G.X
diag(G.X) <- 0

Y <- matrix(rnorm(100*10),100,10)
fit <- edgenet(X=X, Y=Y, G.X=G.X, family="gaussian")
pred <- predict(fit, X)

## End(Not run)
```

yeast *A sample yeast data set for regression*

Description

The yeast data set is a `list` containing three matrices that can be used as an example for using `netReg`. The data have been taken from the references listed below.

Usage

```
data(yeast)
```

Format

A `list` containing three matrices

Details

- `X` (112 x 500)-dimensional binary matrix of 500 genetic markers for 112 yeast samples
- `Y` (112 x 231)-dimensional double matrix of 231 gene expression values for 112 yeast samples
- `GY` (231 x 231)-dimensional adjaceny matrix representing protein-protein interactions for 231 yeast genes

References

Brem, Rachel B., et al. (2005), Genetic interactions between polymorphisms that affect gene expression in yeast.
Nature

Storey, John D., Joshua M. Akey, and Leonid Kruglyak (2005), Multiple locus linkage analysis of genomewide expression in yeast.
PLoS Biology

Index