Package 'CompositionalLASSO'

November 20, 2026

Type Package

Title LASSO Compositional Data Regression

Version 1.1

Date 2026-11-20

Author Michail Tsagris [aut, cre]

Maintainer Michail Tsagris <mtsagris@uoc.gr>

Depends R (>= 4.0)

Imports Compositional, glmnet, Rfast, stats

Suggests Rfast2

Description LASSO (Least Absolute Shrinkage and Selection Operator) regression for compositional responses and Euclidean predictors, or Euclidean responses and compositional predictors with and without zero values present. The choice of the penalty parameter, lambda, takes place using K-fold cross-validation. If the compositional data appear on the predictors side the alpha-transformation is employed first. Two relevant papers are Li Z. et al. (2018) <doi:10.1007/s12561-018-9219-2> and Alenazi A.A. (2022) <doi:10.18187/pjsor.v18i4.3969>.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2026-11-20 12:00:15 UTC

Contents

CompositionalLASSO-package 2
alfa.lasso 3
cv.alfalasso 4
cv.lasso.compreg 6
cv.lasso.klcompreg 7
cv.lasso.mziln 9
lasso.compreg 10
lasso.klcompreg 11
lasso.mziln 13

Index **15**

`CompositionalLASSO-package`
Regression for Compositional Data with Zero Values

Description

LASSO (Least Absolute Shrinkage and Selection Operator) regression for compositional responses and Euclidean predictors, or Euclidean responses and compositional predictors with and without zero values present.

Details

Package:	CompositionalLASSO
Type:	Package
Version:	1.1
Date:	2026-11-20

Maintainers

Michail Tsagris <mtsagris@uoc.gr>.

Author(s)

Michail Tsagris <mtsagris@uoc.gr>

References

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

Li Z., Lee K., Karagas M. R., Madan J. C., Hoen A. G., O'Malley A. J. and Li H. (2018). Conditional regression based on a multivariate zero-inflated logistic-normal model for microbiome relative abundance data. Statistics in Biosciences, 10(3): 587–608.

Alenazi A. A. (2022). f-divergence regression models for compositional data. Pakistan Journal of Statistics and Operation Research, 18(4): 867–882.

`alfa.lasso` *LASSO with compositional predictors using the $alpha$-transformation*

Description

LASSO with compositional predictors using the $alpha$-transformation.

Usage

```
alfa.lasso(y, x, a = seq(-1, 1, by = 0.1), model = "gaussian", lambda = NULL,
xnew = NULL)
```

Arguments

`y`	A numerical vector or a matrix for multinomial logistic regression.
`x`	A numerical matrix containing the predictor variables, compositional data, where zero values are allowed..
`a`	A vector with a grid of values of the power transformation, it has to be between -1 and 1. If zero values are present it has to be greater than 0. If $\alpha = 0$ the isometric log-ratio transformation is applied.
`model`	The type of the regression model, "gaussian", "binomial", "poisson", "multinomial", or "mgaussian".
`lambda`	**This information is copied from the package glmnet.** A user supplied lambda sequence. Typical usage is to have the program compute its own lambda sequence based on nlambda and lambda.min.ratio. Supplying a value of lambda overrides this. WARNING: use with care. Avoid supplying a single value for lambda (for predictions after CV use predict() instead). Supply instead a decreasing sequence of lambda values. glmnet relies on its warms starts for speed, and its often faster to fit a whole path than compute a single fit.
`xnew`	If you have new data use it, otherwise leave it NULL.

Details

The function uses the glmnet package to perform LASSO penalised regression. For more details see the function in that package.

Value

A list including sublists for each value of α:

`mod`	We decided to keep the same list that is returned by glmnet. So, see the function in that package for more information.
`est`	If you supply a matrix in the "xnew" argument this will return an array of many matrices with the fitted values, where each matrix corresponds to each value of λ.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, Vol. 33(1), 1–22.

See Also

cv.alfa.lasso

Examples

```
y <- as.matrix(iris[, 1])
x <- matrix( rgamma(150 * 20, 3, 1), ncol = 20 )
x <- x / Rfast::rowsums(x)
mod <- alfa.lasso(y, x, a = c(0, 0.5, 1))
```

---

cv.alfalasso *Cross-validation for LASSO with compositional predictors using the $alpha$-transformation*

Description

Cross-validation for LASSO with compositional predictors using the $alpha$-transformation.

Usage

```
cv.alfa.lasso(y, x, a = seq(-1, 1, by = 0.1), model = "gaussian", lambda = NULL,
type.measure = "mse", nfolds = 10, folds = NULL, stratified = FALSE)
```

Arguments

y	A numerical vector or a matrix for multinomial logistic regression.
x	A numerical matrix containing the predictor variables, compositional data, where zero values are allowed.
a	A vector with a grid of values of the power transformation, it has to be between -1 and 1. If zero values are present it has to be greater than 0. If $\alpha = 0$ the isometric log-ratio transformation is applied.
model	The type of the regression model, "gaussian", "binomial", "poisson", "multinomial", or "mgaussian".

`lambda`	**This information is copied from the package glmnet.** A user supplied lambda sequence. Typical usage is to have the program compute its own lambda sequence based on nlambda and lambda.min.ratio. Supplying a value of lambda overrides this. WARNING: use with care. Avoid supplying a single value for lambda (for predictions after CV use predict() instead). Supply instead a decreasing sequence of lambda values. glmnet relies on its warms starts for speed, and its often faster to fit a whole path than compute a single fit.
`type.measure`	**This information is taken from the package glmnet.** The loss function to use for cross-validation. For gaussian models this can be "mse", "deviance" for logistic and poisson regression, "class" applies to binomial and multinomial logistic regression only, and gives misclassification error. "auc" is for two-class logistic regression only, and gives The area under the ROC curve. "mse" or "mae" (mean absolute error) can be used by all models.
`nfolds`	The number of folds. Set to 10 by default.
`folds`	If you have the list with the folds supply it here. You can also leave it NULL and it will create folds.
`stratified`	Do you want the folds to be created in a stratified way? TRUE or FALSE.

Details

The function uses the glmnet package to perform LASSO penalised regression. For more details see the function in that package.

Value

A matrix with two columns and number of rows equal to the number of α values used. Each row contains, the optimal value of the λ penalty parameter for the LASSO and the optimal value of the loss function, for each value of α.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, Vol. 33(1), 1–22.

See Also

`alfa.lasso`

Examples

```
y <- iris[, 1]
x <- matrix( rgamma(150 * 20, 3, 1), ncol = 20 )
x <- x / Rfast::rowsums(x)
mod <- cv.alfa.lasso( y, x, a = c(0.2, 0.5, 1) )
```

`cv.lasso.compreg` *Cross-validation for the LASSO log-ratio regression with compositional response*

Description

Cross-validation for the LASSO log-ratio regression with compositional response.

Usage

```
cv.lasso.compreg(y, x, lambda = NULL, nlambda = 100, nfolds = 10,
folds = NULL, seed = NULL)
```

Arguments

`y`	A numerical matrix with compositional data. Zero values are not allowed as the additive log-ratio transformation is applied to the compositional response prior to implementing the LASSO algortihm.
`x`	A matrix with the predictor variables.
`lambda`	A user supplied lambda sequence.
`nlambda`	**This information is copied from the package glmnet.** The number of $lambda$ values, default is 100.
`nfolds`	The number of folds for the K-fold cross-validation, set to 10 by default.
`folds`	If you have the list with the folds supply it here. You can also leave it NULL and it will create folds.
`seed`	You can specify your own seed number here or leave it NULL.

Details

The K-fold cross-validation is performed in order to select the optimal value for λ, the penalty parameter in LASSO.

Value

A list including:

`runtime`	The runtime of the process.
`kld`	The cross-validated Kullback-Leibler divergence for each value of λ.
`opt`	The optimal value of λ and its correesponding KLD.
`be`	The regression coefficients for the optimal value of λ.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, Vol. 33(1), 1-22.

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

See Also

lasso.compreg, lasso.klcompreg, cv.lasso.klcompreg

Examples

```
y <- as.matrix(iris[, 1:3])
y <- y/rowSums(y)
x <- matrix(rnorm(150 * 50), ncol = 50)
mod <- cv.lasso.compreg(y, x)
```

---

cv.lasso.klcompreg *Cross-validation for the LASSO Kullback-Leibler divergence based regression*

Description

Cross-validation for the LASSO Kullback-Leibler divergence based regression.

Usage

```
cv.lasso.klcompreg(y, x, lambda = NULL, nlambda = 100, type = "grouped",
nfolds = 10, folds = NULL, seed = NULL)
```

Arguments

y	A numerical matrix with compositional data with or without zeros.
x	A matrix with the predictor variables.
lambda	A user supplied lambda sequence.
nlambda	**This information is copied from the package glmnet.** The number of $lambda$ values, default is 100.
type	**This information is copied from the package glmnet.**. If "grouped" then a grouped lasso penalty is used on the multinomial coefficients for a variable. This ensures they are all in our out together. The default in our case is "grouped".
nfolds	The number of folds for the K-fold cross-validation, set to 10 by default.

folds	If you have the list with the folds supply it here. You can also leave it NULL and it will create folds.
seed	You can specify your own seed number here or leave it NULL.

Details

The K-fold cross-validation is performed in order to select the optimal value for λ, the penalty parameter in LASSO.

Value

A list including:

runtime	The runtime of the process.
kld	The cross-validated Kullback-Leibler divergence for each value of λ.
opt	The optimal value of λ and its correesponding KLD.
be	The regression coefficients for the optimal value of λ.

Author(s)

Michail Tsagris and Abdulaziz Alenazi.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr> and Abdulaziz Alenazi <a.alenazi@nbu.edu.sa>.

References

Alenazi, A. A. (2022). f-divergence regression models for compositional data. Pakistan Journal of Statistics and Operation Research, 18(4): 867–882.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, Vol. 33(1), 1-22.

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

See Also

lasso.klcompreg, lasso.compreg, cv.lasso.compreg

Examples

```
y <- as.matrix(iris[, 1:3])
y <- y/rowSums(y)
x <- matrix(rnorm(150 * 50), ncol = 50)
mod <- cv.lasso.klcompreg(y, x)
```

`cv.lasso.mziln` *Cross-validation for the LASSO log-ratio regression with compositional response*

Description

Cross-validation for the LASSO log-ratio regression with compositional response.

Usage

```
cv.lasso.mziln(y, x, lambda = NULL, nlambda = 100, nfolds = 10,
folds = NULL, seed = NULL)
```

Arguments

`y`	A numerical matrix with compositional data. Zero values are not allowed as the additive log-ratio transformation is applied to the compositional response prior to implementing the LASSO algortihm.
`x`	A matrix with the predictor variables.
`lambda`	A user supplied lambda sequence.
`nlambda`	**This information is copied from the package glmnet.** The number of $lambda$ values, default is 100.
`nfolds`	The number of folds for the K-fold cross-validation, set to 10 by default.
`folds`	If you have the list with the folds supply it here. You can also leave it NULL and it will create folds.
`seed`	You can specify your own seed number here or leave it NULL.

Details

The K-fold cross-validation is performed in order to select the optimal value for λ, the penalty parameter in LASSO.

Value

A list including:

`runtime`	The runtime of the process.
`kld`	The cross-validated Kullback-Leibler divergence for each value of λ.
`opt`	The optimal value of λ and its correesponding KLD.
`be`	The regression coefficients for the optimal value of λ.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Li Z., Lee K., Karagas M. R., Madan J. C., Hoen A. G., O'Malley A. J. and Li H. (2018). Conditional regression based on a multivariate zero-inflated logistic-normal model for microbiome relative abundance data. Statistics in Biosciences, 10(3): 587–608.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, Vol. 33(1), 1-22.

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

See Also

lasso.compreg, lasso.klcompreg, lasso.mziln

Examples

```
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
y[sample(150, 10), 3] <- 0
y <- y / rowSums(y)
x <- matrix(rnorm(150 * 50), ncol = 50)
mod <- cv.lasso.mziln(y, x)
```

---

lasso.compreg *LASSO log-ratio regression*

Description

LASSO log-ratio regression.

Usage

```
lasso.compreg(y, x, lambda = NULL, nlambda = 100, xnew = NULL)
```

Arguments

y	A numerical matrix with compositional data. Zero values are not allowed as the additive log-ratio transformation is applied to the compositional response prior to implementing the LASSO algortihm.
x	A numerical matrix containing the predictor variables.
lambda	**This information is copied from the package glmnet.** A user supplied lambda sequence. Typical usage is to have the program compute its own lambda sequence based on nlambda and lambda.min.ratio. Supplying a value of lambda overrides this. WARNING: use with care. Avoid supplying a single value for lambda (for predictions after CV use predict() instead). Supply instead a decreasing sequence of lambda values. glmnet relies on its warms starts for speed, and its often faster to fit a whole path than compute a single fit.
nlambda	**This information is copied from the package glmnet.** The number of $lambda$ values, default is 100.
xnew	If you have new data use it, otherwise leave it NULL.

Details

The function uses the glmnet package to perform LASSO penalised regression. For more details see the function in that package.

Value

A list including:

be	A list with the estimated regression coefficients for each value of λ.
info	A matrix with two columns, the values of lambda used and their corresponding Kullback-Leibler divergence between the observed and the fitted values.
est	If you supply a matrix in the "xnew" argument this will return an array of many matrices with the fitted values, where each matrix corresponds to each value of λ.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, Vol. 33(1), 1-22.

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

See Also

cv.lasso.compreg, lasso.klcompreg, cv.lasso.klcompreg, lasso.mziln

Examples

```
y <- as.matrix(iris[, 1:3])
y <- y/rowSums(y)
x <- matrix(rnorm(150 * 50), ncol = 50)
mod <- lasso.compreg(y, x)
```

---

lasso.klcompreg *LASSO Kullback-Leibler divergence based regression*

Description

LASSO Kullback-Leibler divergence based regression.

Usage

```
lasso.klcompreg(y, x, lambda = NULL, nlambda = 100, type = "grouped", xnew = NULL)
```

Arguments

y	A numerical matrix with compositional data. Zero values are allowed.
x	A numerical matrix containing the predictor variables.
lambda	**This information is copied from the package glmnet.** A user supplied lambda sequence. Typical usage is to have the program compute its own lambda sequence based on nlambda and lambda.min.ratio. Supplying a value of lambda overrides this. WARNING: use with care. Avoid supplying a single value for lambda (for predictions after CV use predict() instead). Supply instead a decreasing sequence of lambda values. glmnet relies on its warms starts for speed, and its often faster to fit a whole path than compute a single fit.
nlambda	**This information is copied from the package glmnet.** The number of $lambda$ values, default is 100.
type	**This information is copied from the package glmnet.**. If "grouped" then a grouped lasso penalty is used on the multinomial coefficients for a variable. This ensures they are all in our out together. The default in our case is "grouped".
xnew	If you have new data use it, otherwise leave it NULL.

Details

The function uses the glmnet package to perform LASSO penalised regression. For more details see the function in that package.

Value

A list including:

be	A list with the estimated regression coefficients for each value of λ.
info	A matrix with two columns, the values of lambda used and their corresponding Kullback-Leibler divergence between the observed and the fitted values.
est	If you supply a matrix in the "xnew" argument this will return an array of many matrices with the fitted values, where each matrix corresponds to each value of λ.

Author(s)

Michail Tsagris and Abdulaziz Alenazi.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr> and Abdulaziz Alenazi <a.alenazi@nbu.edu.sa>.

References

Alenazi A. A. (2022). f-divergence regression models for compositional data. Pakistan Journal of Statistics and Operation Research, 18(4): 867–882.

Friedman J., Hastie T. and Tibshirani R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, Vol. 33(1), 1–22.

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

See Also

cv.lasso.klcompreg, lasso.compreg, lasso.mziln

Examples

```
y <- as.matrix(iris[, 1:3])
y <- y/rowSums(y)
x <- matrix(rnorm(150 * 50), ncol = 50)
mod <- lasso.klcompreg(y, x)
```

---

lasso.mziln *LASSO using the multivariate zero inflated logistic normal regression*

Description

LASSO using the multivariate zero inflated logistic normal regression.

Usage

```
lasso.mziln(y, x, lambda = NULL, nlambda = 100, xnew = NULL)
```

Arguments

y	A numerical matrix with compositional data with zero values.
x	A numerical matrix containing the predictor variables.
lambda	**This information is copied from the package glmnet.** A user supplied lambda sequence. Typical usage is to have the program compute its own lambda sequence based on nlambda and lambda.min.ratio. Supplying a value of lambda overrides this. WARNING: use with care. Avoid supplying a single value for lambda (for predictions after CV use predict() instead). Supply instead a decreasing sequence of lambda values. glmnet relies on its warms starts for speed, and its often faster to fit a whole path than compute a single fit.
nlambda	**This information is copied from the package glmnet.** The number of $lambda$ values, default is 100.
xnew	If you have new data use it, otherwise leave it NULL.

Details

The function uses the glmnet package to perform LASSO penalised regression.

Value

A list including:

be	A list with the estimated regression coefficients for each value of λ.
info	A matrix with two columns, the values of lambda used and their corresponding Kullback-Leibler divergence between the observed and the fitted values.
est	If you supply a matrix in the "xnew" argument this will return an array of many matrices with the fitted values, where each matrix corresponds to each value of λ.

Author(s)

Michail Tsagris.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr>.

References

Li Z., Lee K., Karagas M. R., Madan J. C., Hoen A. G., O'Malley A. J. and Li H. (2018). Conditional regression based on a multivariate zero-inflated logistic-normal model for microbiome relative abundance data. Statistics in Biosciences, 10(3): 587–608.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, Vol. 33(1), 1-22.

Aitchison J. (1986). The statistical analysis of compositional data. Chapman & Hall.

See Also

cv.lasso.mziln, cv.lasso.compreg, cv.lasso.klcompreg

Examples

```
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
y[sample(150, 10), 3] <- 0
y <- y / rowSums(y)
x <- matrix(rnorm(150 * 50), ncol = 50)
mod <- lasso.mziln(y, x)
```

Index

alfa.lasso, 3, *5*

CompositionalLASSO-package, 2
cv.alfa.lasso, *4*
cv.alfa.lasso *(*cv.alfalasso*)*, 4
cv.alfalasso, 4
cv.lasso.compreg, 6, *8*, *11*, *14*
cv.lasso.klcompreg, *7*, 7, *11*, *13, 14*
cv.lasso.mziln, 9, *14*

lasso.compreg, *7, 8*, *10*, 10, *13*
lasso.klcompreg, *7, 8*, *10, 11*, 11
lasso.mziln, *10, 11*, *13*, 13