Package 'Compositionalridge'

November 26, 2026

Type Package

Title Ridge Regression for Compositional Data

Version 1.1

Date 2026-11-26

Maintainer Michail Tsagris <mtsagris@uoc.gr>

Depends R (>= 4.0)

Imports Compositional, glmnet, minpack.lm, Rfast, stats, Rcpp

Suggests Rfast2

Description A few ridge regression models for compositional data with zero values present. These are the zero adjusted Dirichlet regression, the multivariate zero inflated logistic normal, the Kull-
back-Leibler divergence and the alpha-regression. Some relevant papers include Tsagris and Stewart (2018) <doi:10.1134/S1995080218030198>, Tsagris et al. (2026) <doi:10.48550/arXiv.2510.12663>, and Li et al. (2018 018-9219-2>.

License GPL (>= 2)

LinkingTo Rcpp, RcppArmadillo

Encoding UTF-8

Config/roxygen2/version 8.1.1

NeedsCompilation yes

Author Michail Tsagris [aut, cre],
Christos Adam [aut] (ORCID: <https://orcid.org/0009-0003-3244-7034>)

Repository CRAN

Date/Publication 2026-11-26 09:10:08 UTC

Contents

Compositionalridge-package 2
alfa.ridge 3
areg.ridge 4
cv.alfaridge 6

cv.aregridge 7
cv.zadrridge 9
zadr.ridge 10

Index **13**

`Compositionalridge-package`

Ridge Regression for Compositional Data

Description

A few ridge regression models for compositional data with zero values present. These are the zero adjusted Dirichlet regression, the multivariate zero inflated logistic normal, the Kullback-Leibler divergence and the alpha-regression.

Details

Package:	Compositionalridge
Type:	Package
Version:	1.1
Date:	2026-11-26

Maintainers

Michail Tsagris <mtsagris@uoc.gr>.

Author(s)

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.

`alfa.ridge` *Ridge regression with compositional predictors using the $alpha$-transformation*

Description

Ridge regression with compositional predictors using the $alpha$-transformation.

Usage

```
alfa.ridge(y, x, a = seq(-1, 1, by = 0.1), h = TRUE, 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.
`h`	Should the power transfrmed data be multipled by the Helmert matrix? The default value is TRUE.
`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 ridge 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.alfaridge

Examples

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

---

areg.ridge *Ridge α-regression for compositional data*

Description

Ridge α-regression for compositional data

Usage

```
areg.ridge(y, x, a, lambda = NULL, nlambda = 100, yb = NULL, xnew = NULL)
```

Arguments

y	A matrix with the compositional data.
x	A matrix with the continuous predictor variables or a data frame including categorical predictor variables.
a	The values of α for the α-transformation. If zero values are present only the positive values of α are used.
lambda	A vector with a sequence of the ridge λ values. If you do not know it leave it NULL and the function will compute it internally.

nlambda	The number of λ values to compute.
yb	If you have already α-transformed the data using α as given in the argument "a", put it here. Othewrise leave it NULL.
xnew	If you have new data whose response values you want to compute, pass them here.

Details

The ridge α-regression is applied.

Value

A list including:

runtime	The runtime of the ridge α-regression.
info	A matrix with two columns containing the lambda, and the Kullback-Leibler divergence between the observed and the fitted compositions.
be	A list with the ridge beta coefficients. **Note**: The predictors are standardised internally and then ridge regression is applied. The returned coefficients are transformed back to their original scale, just like the package glmnet does.
est	A list with the ridge estimated compositions, if xnew is not NULL.

Author(s)

Michail Tsagris.

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

References

Tsagris M. and Pantazis Y. (2026). The α–regression for compositional data: a unified framework for standard, temporal and spatial regression models including compositional predictors. https://arxiv.org/pdf/2510.12663

Tsagris M. (2015). Regression analysis with compositional data containing zero values. Chilean Journal of Statistics, 6(2): 47-57. https://arxiv.org/pdf/1508.01913v1.pdf

Tsagris M.T., Preston S. and Wood A.T.A. (2011). A data-based power transformation for compositional data. In Proceedings of the 4th Compositional Data Analysis Workshop, Girona, Spain. https://arxiv.org/pdf/1106.1451.pdf

See Also

cv.aregridge, zadr.ridge

Examples

```
x <- matrix( rnorm(150 * 10), ncol = 10 )
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
mod <- areg.ridge(y, x, 0.5)
```

`cv.alfaridge` *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.alfaridge(y, x, a = seq(-1, 1, by = 0.1), h = TRUE, 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.
`h`	Should the power transfrmed data be multipled by the Helmert matrix? The default value is TRUE.
`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 ridge 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 ridge regression 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.ridge

Examples

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

---

cv.aregridge *K-fold cross-validation for the ridge α-regression*

Description

K-fold cross-validation for the ridge α-regression.

Usage

```
cv.aregridge(y, x, a = seq(0.1, by = 0.1), lambda = NULL, nlambda = 100,
folds = NULL, nfolds = 10)
```

Arguments

y	A matrix with the compositional data (dependent variable). Only for the Dirichlet regression, zeros are not allowed.
x	The predictor variable(s), they can only be continnuous variables.
a	A vector with values of α for the α-transformation. If zero values are present only the positive values of α are used.
lambda	A vector with a sequence of the ridge λ values. If you do not know it leave it NULL and the function will compute it internally.
nlambda	The number of λ values to compute.
folds	If you have the list with the folds supply it here. You can also leave it NULL and it will create folds.
nfolds	The number of folds to produce.

Details

K-fold cross-validation for the ridge α-regression.

Value

A list including:

runtime	The runtime of the ridge regression model.
info	A matrix with two columns containing the lambda, the penalized log-likelihood (for dr and zadr), and the Kullback-Leibler divergence between the observed and te fitted compositions.
be	A list with the ridge beta coefficients.
sest	A list with the ridge estimated compositions.

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.

See Also

cv.zadrridge, areg.ridge

Examples

```
x <- matrix( rnorm(150 * 10), ncol = 10 )
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
mod <- kld.ridge(y, x)
```

`cv.zadrridge` *Cross-validation for ridge regression for compositional data*

Description

Cross-validation for ridge regression for compositional data.

Usage

```
cv.zadrridge(y, x, lambda = NULL, nlambda = 100, tol = 1e-06, maxit = 100,
folds = NULL, nfolds = 10)
cv.drridge(y, x, lambda = NULL, nlambda = 100, tol = 1e-06, maxit = 100,
folds = NULL, nfolds = 10)
cv.kldridge(y, x, lambda = NULL, nlambda = 100,
folds = NULL,  nfolds = 10)
cv.mzilnridge(y, x, lambda = NULL, nlambda = 100,
folds = NULL, nfolds = 10)
```

Arguments

`y`	A matrix with the compositional data (dependent variable). Only for the Dirichlet regression, zeros are not allowed.
`x`	The predictor variable(s), they can only be continnuous variables.
`lambda`	A vector with a sequence of the ridge λ values. If you do not know it leave it NULL and the function will compute it internally.
`nlambda`	The number of λ values to compute.
`tol`	The tolerance value to terminate the Fisher scoring algorithm.
`maxit`	The maximum number of iterations the Fisher scoring algorithm will perform.
`folds`	If you have the list with the folds supply it here. You can also leave it NULL and it will create folds.
`nfolds`	The number of folds to produce.

Details

K-fold cross-validation for the ridge regression using the Dirichlet regression, the zero adjuster Dirichlet regression, the Kullback-Leilber divergence regression or the multivariate zero inflated logistic normal regression.

Value

A list including:

runtime	The runtime of the ridge regression model.
info	A matrix with two or three columns containing the lambda, the penalized log-likelihood (for dr and zadr), and the Kullback-Leibler divergence between the observed and te fitted compositions.
be	A list with the ridge beta coefficients.
est	A list with the ridge estimated compositions.

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.

See Also

cv.zadrridge, areg.ridge

Examples

```
x <- matrix( rnorm(150 * 10), ncol = 10 )
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
mod <- kld.ridge(y, x)
```

---

zadr.ridge *Ridge regression for compositional data*

Description

Ridge regression for compositional data.

Usage

```
zadr.ridge(y, x, lambda = NULL, nlambda = 100, xnew = NULL, tol = 1e-6, maxit = 100)
dr.ridge(y, x, lambda = NULL, nlambda = 100, xnew = NULL, tol = 1e-6, maxit = 100)
kld.ridge(y, x, lambda = NULL, nlambda = 100, xnew = NULL)
mziln.ridge(y, x, lambda = NULL, nlambda = 100, xnew = NULL)
```

Arguments

y	A matrix with the compositional data (dependent variable). Only for the Dirichlet regression, zeros are not allowed.
x	The predictor variable(s), they can only be continnuous variables.
lambda	A vector with a sequence of the ridge λ values. If you do not know it leave it NULL and the function will compute it internally.
nlambda	The number of λ values to compute.
xnew	If you have new data whose response values you want to compute, pass them here.
tol	The tolerance value to terminate the Fisher scoring algorithm.
maxit	The maximum number of iterations the Fisher scoring algorithm will perform.

Details

Ridge regression using the Dirichlet regression, the zero adjuster Dirichlet regression, the Kullback-Leilber divergence regression or the multivariate zero inflated logistic normal regression.

Value

A list including:

runtime	The runtime of the ridge regression model.
info	A matrix with two or three columns containing the lambda, the penalized log-likelihood (for dr and zadr), and the Kullback-Leibler divergence between the observed and the fitted compositions.
be	A list with the ridge beta coefficients. Note that for the mziln.ridge() the additive log-ratio transformation is implied as well, but the reference category is the last component, unlike all other regression models that imply the first component. So, be careful with the interpretation. Also, for the kld.ridge() the coefficients are based on the inverse of the centered log-ratio transformation. **Note**: For all regression models, the predictors are standardised internally and then ridge regression is applied. The returned coefficients are transformed back to their original scale, just like the package glmnet does.
est	A list with the ridge estimated compositions, if xnew is not NULL.

Author(s)

Michail Tsagris and Christos Adam.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr> and the C++ part of zadr.ridge() by Christos Adam <pada4m4@gmail.com>.

References

Tsagris M. and Pantazis Y. (2026). The α–regression for compositional data: a unified framework for standard, temporal and spatial regression models including compositional predictors. https://arxiv.org/pdf/2510.12663

Tsagris M. (2015). Regression analysis with compositional data containing zero values. Chilean Journal of Statistics, 6(2): 47-57. https://arxiv.org/pdf/1508.01913v1.pdf

Tsagris M.T., Preston S. and Wood A.T.A. (2011). A data-based power transformation for compositional data. In Proceedings of the 4th Compositional Data Analysis Workshop, Girona, Spain. https://arxiv.org/pdf/1106.1451.pdf

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

Tsagris M. and Stewart C. (2018). A Dirichlet regression model for compositional data with zeros. Lobachevskii Journal of Mathematics,39(3): 398–412.

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.

See Also

cv.zadrridge, areg.ridge

Examples

```
x <- matrix( rnorm(150 * 10), ncol = 10 )
y <- as.matrix(iris[, 1:3])
y <- y / rowSums(y)
ind <- sample(150, 10)
for ( k in ind )  y[k, sample(3, 1)] <- 0
y <- y / rowSums(y)
mod <- kld.ridge(y, x)
```

Index

alfa.ridge, 3, *7*
areg.ridge, 4, *8*, *10*, *12*

Compositionalridge-package, 2
cv.alfaridge, *4*, 6
cv.aregridge, *5*, 7
cv.drridge *(*cv.zadrridge*)*, 9
cv.kldridge *(*cv.zadrridge*)*, 9
cv.mzilnridge *(*cv.zadrridge*)*, 9
cv.zadrridge, *8*, 9, *10*, *12*

dr.ridge *(*zadr.ridge*)*, 10

kld.ridge *(*zadr.ridge*)*, 10

mziln.ridge *(*zadr.ridge*)*, 10

zadr.ridge, *5*, 10