Package ‘CompositionalRF’

September 8, 2025

Type Package

Title Multivariate Random Forest with Compositional Responses

Version 1.4

Date 2025-09-07

Maintainer Michail Tsagris <mtsagris@uoc.gr>

Depends R (>= 4.0)

Imports Compositional, RcppParallel, Rcpp, Rfast, stats

LinkingTo Rcpp, RcppParallel

Suggests Rfast2

Description Multivariate random forests with compositional responses and Euclidean predictors is performed. The compositional data are first transformed using the additive log-ratio transformation, or the alpha-transformation of Tsagris, Preston and Wood (2011), <doi:10.48550/arXiv.1106.1451>, and then the multivariate random forest of Rahman R., Otridge J. and Pal R. (2017), <doi:10.1093/bioinformatics/btw765>, is applied.

License GPL (>= 2)

SystemRequirements GNU make

Repository CRAN

NeedsCompilation yes

Author Michail Tsagris [aut, cre],
Christos Adam [aut]

Date/Publication 2025-09-07 22:10:39 UTC

Contents

`CompositionalRF-package` *Multivariate Random Forests with Compositional Responses*

Description

Multivariate random forest with compositional response variables and continuous predictor variables. The data are first transformed using the additive log-ratio transformation and then the multivariate random forest of Rahman R., Otridge J. and Pal R. (2017), <doi:10.1093/bioinformatics/btw765>, is applied.

Details

Package:	CompositionalRF
Type:	Package
Version:	1.4
Date:	2025-09-07
License:	GPL-2

Maintainers

Michail Tsagris <mtsagris@uoc.gr>

Author(s)

Michail Tsagris <mtsagris@uoc.gr>.

References

Rahman R., Otridge J. and Pal R. (2017). IntegratedMRF: random forest-based framework for integrating prediction from different data types. Bioinformatics, 33(9): 1407–1410.

Segal M. and Xiao Y. (2011). Multivariate random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1): 80–87.

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. (2023). A review of compositional data analysis and recent advances. Communications in Statistics–Theory and Methods, 52(16): 5535–5567.

Friedman Jerome, Trevor Hastie and Robert Tibshirani (2009). The elements of statistical learning, 2nd edition. Springer, Berlin.

alfa.comp.rf *Compositional Random Forests using the alpha-transformation*

Description

Compositional Random Forests using the alpha-transformation.

Usage

```
alfa.comp.rf(xnew = x, y, x, a = seq(-1, 1, by = 0.1), ntrees,
nfeatures, minleaf, ncores = 1)
```

Arguments

xnew	A matrix with the new predictor variables whose compositional response values are to be predicted.
y	The response compositional data. Zero values are not allowed.
x	A matrix with the predictor variables data.
a	A vector of α values.
ntrees	The number of trees to construct in the random forest.
nfeatures	The number of randomly selected predictor variables considered for a split in each regression tree node, which must be less than the number of input precictors.
minleaf	Minimum number of observations in the leaf node. If a node has less than or equal to minleaf observations, there will be no splitting in that node and this node will be considered as a leaf node. The number evidently must be less than or equal to the sample size.
ncores	The number of cores to use. If greater than 1, parallel computing will take place. It is advisable to use it if you have many observations and or many variables, otherwise it will slow down the process. The default is 1, meaning that code is executed serially.

Details

The compositional data are first using the α-transformation and then the multivariate random forest algorithm of Rahman, Otridge and Pal (2017) is applied.

Value

A list with the estimated compositional response values, one matrix for each value of α.

Author(s)

Michail Tsagris.

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

References

Rahman R., Otridge J. and Pal R. (2017). IntegratedMRF: random forest-based framework for integrating prediction from different data types. Bioinformatics, 33(9): 1407–1410.

Segal M. and Xiao Y. (2011). Multivariate random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1): 80–87.

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.comprf

Examples

```
y <- as.matrix(iris[, 1:4])
y <- y/ rowSums(y)
x <- matrix( rnorm(150 * 10), ncol = 10 )
mod <- alfa.comp.rf(x[1:10, ], y, x, a = 0.5, ntrees = 2, nfeatures = 5, minleaf = 10)
mod
```

---

comp.rf *Compositional Random Forests*

Description

Compositional Random Forests.

Usage

```
comp.rf(xnew = x, y, x, type = "alr", ntrees, nfeatures, minleaf, ncores = 1)
```

Arguments

xnew	A matrix with the new predictor variables whose compositional response values are to be predicted.
y	The response compositional data. Zero values are not allowed.
x	A matrix with the predictor variables data.
type	If the responses are alreay transformed with the additive log-ratio transformation type 0, otherwise, if they are compositional data, leave it equal to "alr", so that the data will be transformed.
ntrees	The number of trees to construct in the random forest.
nfeatures	The number of randomly selected predictor variables considered for a split in each regression tree node, which must be less than the number of input precictors.

`minleaf` Minimum number of observations in the leaf node. If a node has less than or equal to minleaf observations, there will be no splitting in that node and this node will be considered as a leaf node. The number evidently must be less than or equal to the sample size.

`ncores` The number of cores to use. If greater than 1, parallel computing will take place. It is advisable to use it if you have many observations and or many variables, otherwise it will slow down the process. The default is 1, meaning that code is executed serially.

Details

The compositional are first log-transformed using the additive log-ratio transformation and then the multivariate random forest algorithm of Rahman, Otridge and Pal (2017) is applied.

Value

A matrix with the estimated compositional response values.

Author(s)

Michail Tsagris and Christos Adam.

R implementation and documentation: Michail Tsagris <mtsagris@uoc.gr> and Christos Adam <pada4m4@gmail.com>.

References

Rahman R., Otridge J. and Pal R. (2017). IntegratedMRF: random forest-based framework for integrating prediction from different data types. Bioinformatics, 33(9): 1407–1410.

Segal M. and Xiao Y. (2011). Multivariate random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1): 80–87.

See Also

cv.comprf

Examples

```
y <- as.matrix(iris[, 1:4])
y <- y/ rowSums(y)
x <- matrix( rnorm(150 * 10), ncol = 10 )
mod <- comp.rf(x[1:10, ], y, x, ntrees = 2, nfeatures = 5, minleaf = 10)
mod
```

cv.alfacomprf *Cross-Validation of the Compositional Random Forests using the alpha-transformation*

Description

Cross-Validation of the Compositional Random Forests using the alpha-transformation.

Usage

```
cv.alfacomprf(y, x, a = seq(-1, 1, by = 0.1), ntrees = c(100, 500, 1000),
nfeatures, minleaf, folds = NULL, nfolds = 10, seed = NULL, ncores = 1)
```

Arguments

y	The response compositional data. Zero values are not allowed.
x	A matrix with the predictor variables data.
a	A vector of α values.
ntrees	A vector with the possible number of trees to consider each time.
nfeatures	A vector with the number of randomly selected predictor variables considered for a split in each regression tree node.
minleaf	A vector with the minimum number of observations in the leaf node.
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 in the cross validation.
seed	You can specify your own seed number here or leave it NULL.
ncores	The number of cores to use. If greater than 1, parallel computing will take place. It is advisable to use it if you have many observations and or many variables, otherwise it will slow down the process.

Details

K-fold cross-validation for the multivariate random forest with compositional responses is performed.

Value

A list including:

kl	A matrix with the configurations of hyper-parameters tested and the estimated Kullback-Leibler divergence, for each configuration.
js	A matrix with the configurations of hyper-parameters tested and the estimated Jensen-Shannon divergence, for each configuration.

Author(s)

Michail Tsagris.

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

References

Rahman R., Otridge J. and Pal R. (2017). IntegratedMRF: random forest-based framework for integrating prediction from different data types. Bioinformatics, 33(9): 1407–1410.

Segal M. and Xiao Y. (2011). Multivariate random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1): 80–87.

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

comp.rf

Examples

```
y <- as.matrix(iris[, 1:4])
y <- y/ rowSums(y)
x <- matrix( rnorm(150 * 10), ncol = 10 )
mod <- cv.comprf(y, x, ntrees = 2, nfeatures = 5, minleaf = 10, nfolds = 2)
```

---

cv.comprf — *Cross-Validation of the Compositional Random Forests*

Description

Cross-Validation of the Compositional Random Forests.

Usage

```
cv.comprf(y, x, ntrees = c(50, 100, 500, 1000), nfeatures, minleaf,
folds = NULL, nfolds = 10, seed = NULL, ncores = 1)
```

Arguments

y	The response compositional data. Zero values are not allowed.
x	A matrix with the predictor variables data.
ntrees	A vector with the possible number of trees to consider each time.
nfeatures	A vector with the number of randomly selected predictor variables considered for a split in each regression tree node.
minleaf	A vector with the minimum number of observations in the leaf node.

`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 in the cross validation.
`seed`	You can specify your own seed number here or leave it NULL.
`ncores`	The number of cores to use. If greater than 1, parallel computing will take place. It is advisable to use it if you have many observations and or many variables, otherwise it will slow down the process.

Details

K-fold cross-validation for the multivariate random forest with compositional responses is performed.

Value

A list including:

`kl`	A matrix with the configurations of hyper-parameters tested and the estimated Kullback-Leibler divergence, for each configuration.
`js`	A matrix with the configurations of hyper-parameters tested and the estimated Jensen-Shannon divergence, for each configuration.

Author(s)

Michail Tsagris and Christos Adam.

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

References

Rahman R., Otridge J. and Pal R. (2017). IntegratedMRF: random forest-based framework for integrating prediction from different data types. Bioinformatics, 33(9): 1407–1410.

Segal M. and Xiao Y. (2011). Multivariate random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1): 80–87.

See Also

comp.rf

Examples

```
y <- as.matrix(iris[, 1:4])
y <- y/ rowSums(y)
x <- matrix( rnorm(150 * 10), ncol = 10 )
mod <- cv.comprf(y, x, ntrees = 2, nfeatures = 5, minleaf = 10, nfolds = 2)
```

mrf *Multivariate Random Forests*

Description

Multivariate Random Forests.

Usage

```
mrf(xnew, y, x, ntrees, nfeatures, minleaf, ncores = 1)
```

Arguments

xnew	A matrix with the new predictor variables whose multivariate response values are to be predicted.
y	The response multivariate data.
x	A matrix with the predictor variables data.
ntrees	The number of trees to construct in the random forest.
nfeatures	The number of randomly selected predictor variables considered for a split in each regression tree node, which must be less than the number of input precictors.
minleaf	Minimum number of observations in the leaf node. If a node has less than or equal to minleaf observations, there will be no splitting in that node and this node will be considered as a leaf node. The number evidently must be less than or equal to the sample size.
ncores	The number of cores to use. If greater than 1, parallel computing will take place. It is advisable to use it if you have many observations and or many variables, otherwise it will slow down the process. The default is 1, meaning that code is executed serially.

Details

Multivariate random forest algorithm of Rahman, Otridge and Pal (2017) is applied.

Value

A matrix with the estimated multivariate response values.

Author(s)

Christos Adam.

R implementation and documentation: Christos Adam <pada4m4@gmail.com>.

References

Rahman R., Otridge J. and Pal R. (2017). IntegratedMRF: random forest-based framework for integrating prediction from different data types. Bioinformatics, 33(9): 1407–1410.

Segal M. and Xiao Y. (2011). Multivariate random forests. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1): 80–87.

See Also

`comp.rf`

Examples

```
y <- as.matrix(iris[, 1:4])
x <- matrix( rnorm(150 * 10), ncol = 10 )
mod <- mrf(x[1:10, ], y, x, ntrees = 2, nfeatures = 5, minleaf = 10)
mod
```

Index