

Package ‘multiridge’

March 30, 2021

Type Package

Title Fast Cross-Validation for Multi-Penalty Ridge Regression

Version 1.4

Date 2021-03-29

Author Mark A. van de Wiel

Maintainer Mark A. van de Wiel <mark.vdwiel@amsterdamumc.nl>

Depends R (>= 3.4.0), survival, pROC, risksetROC, mgcv, snowfall

Description The package implements multi-penalty linear, logistic and cox ridge regression, including estimation of the penalty parameters by efficient (repeated) cross-validation and marginal likelihood maximization. It allows for multiple high-dimensional data types that require penalization, as well as unpenalized variables. Moreover, it allows a paired penalty for paired data types, and preferential data types can be specified.

License GPL (>=3)

R topics documented:

multiridge-package	2
augment	3
betasout	3
createXblocks	4
createXXblocks	5
CVfolds	6
CVscore	7
doubleCV	8
fastCV2	9
IWLSCoxridge	10
IWLSridge	12
mgcv_lambda	13
mlikCV	14
optLambdas	15
optLambdasWrap	17
optLambdas_mgcv	18
optLambdas_mgcvWrap	20
predictIWLS	21
Scoring	22
setupParallel	23
SigmaFromBlocks	24

Index**25**

multiridge-package	<i>Fast cross-validation for multi-penalty ridge regression</i>
--------------------	---

Description

The package implements multi-penalty linear, logistic and cox ridge regression, including estimation of the penalty parameters by efficient (repeated) cross-validation or marginal likelihood maximization. It allows for multiple high-dimensional data types that require penalization, as well as unpenalized variables. Moreover, it allows a paired penalty for paired data types, and preferential data types can be specified.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

betasout: Coefficient estimates from (converged) IWLS fit

createXXblocks: Creates list of (unscaled) sample covariance matrices

CVscore: Cross-validated score for given penalty parameters

doubleCV: Double cross-validation for estimating performance

fastCV2: Fast cross-validation per data block; no dependency

IWLSCoxridge: Iterative weighted least squares algorithm for Cox ridge regression

IWLSridge: Iterative weighted least squares algorithm for linear and logistic ridge regression

mlikCV: Cross-validation for estimating performance of marginal likelihood estimation

optLambdasWrap: Find optimal ridge penalties by cross-validation

optLambdas_mgcVWrap: Find optimal ridge penalties in terms of marginal likelihood

predictIWLS: Predictions from ridge fits

setupParallel: Setting up parallel computing

SigmaFromBlocks: Create penalized sample cross-product matrix

Author(s)

Mark A. van de Wiel (mark.vdwiel@amsterdamumc.nl)

References

Mark A. van de Wiel, Mirrelijn van Nee, Armin Rauschenberger (2021). Fast cross-validation for high-dimensional ridge regression. J Comp Graph Stat

See Also

A full demo and data are available from:

<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzeG506nBwgW64e4>

augment	<i>Augment data with zeros.</i>
---------	---------------------------------

Description

This function augments data with zeros to allow pairing of data on the same variables, but from DIFFERENT samples

Usage

```
augment(Xdata1, Xdata2)
```

Arguments

Xdata1	Data frame or data matrix of dimension $n_1 \times p$.
Xdata2	Data frame or data matrix of dimension $n_2 \times p$

Details

Xdata1 and Xdata2 should have the same number of columns. These columns represent variables. Augments both data matrices with zeros, such that the matrices can be paired using [createXblocks](#) on the output of this function.

Value

List	
Xaug1	Augmented data matrix 1
Xaug2	Augmented data matrix 2

betasout	<i>Coefficient estimates from (converged) IWLS fit</i>
----------	--

Description

Extracts estimated regression coefficients from the final Iterative Weighted Least Squares fit, as obtained from linear, logistic, or Cox ridge regression.

Usage

```
betasout(IWLSfit, Xblocks, penalties, pairing = NULL)
```

Arguments

IWLSfit	List object, see details
Xblocks	List of data frames or matrices, representing $b=1, \dots, B$ data blocks of dimensions $n \times p_b$.
penalties	Numerical vector.
pairing	Numerical vector of length 3 or NULL.

Details

IWLSfit should be the output of either [IWLSridge](#) or [IWLScoxridge](#). Xblocks may be created by [createXblocks](#).

Value

List. Number of components equals number of components of Xblocks plus one, as the output is augmented with an intercept estimate (first component, NULL if absent). Each component is a numerical vector representing regression parameter estimates. Lengths of vectors match column dimensions of Xblocks (nr of variables for given data type)

See Also

[createXblocks](#). A full demo and data are available from:
<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzveG506nBwgW64e4>

createXblocks	<i>Create list of paired data blocks</i>
---------------	--

Description

Create list of paired data blocks

Usage

```
createXblocks(datablocks, which2pair = NULL)
```

Arguments

datablocks	List of data frames or matrices representing $b=1, \dots, B$ data blocks of dimensions $n \times p_b$.
which2pair	Integer vector of size 2 (or NULL)

Details

Only use this function when you wish to pair two data blocks. If `which2pair = NULL` the output matches the input. If not, the function adds a paired data block, pairing the two data blocks corresponding to the elements of `which2pair`.

Value

List. Same as `datablocks` when `which2pair = NULL`, or augmented with one paired data block.

See Also

[createXXblocks](#). A full demo and data are available from:
<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzveG506nBwgW64e4>

createXXblocks	<i>Creates list of (unscaled) sample covariance matrices</i>
----------------	--

Description

Creates list of (unscaled) sample covariance matrices $X_b \%*\% t(X_b)$ for data blocks $b = 1, \dots, B$.

Usage

```
createXXblocks(datablocks, datablocksnew = NULL, which2pair = NULL)
```

Arguments

datablocks	List of data frames or matrices
datablocksnew	List of data frames or matrices
which2pair	Integer vector of size 2 (or NULL)

Details

The efficiency of `multiridge` for high-dimensional data relies largely on this function: all iterative calculation are performed on the out put of this function, which contains B blocks of $n \times n$ matrices. If `which2pair != NULL`, the function adds a paired covariance block, pairing the two data blocks corresponding to the elements of `which2pair`. If predictions for new samples are desired, one also needs to specify `datablocksnew`, which should have the exact same format as `datablocks` with matching column dimension (number of variables).

Value

List. Same number of component as `datablocks` when `which2pair = NULL`, or augmented with one paired data block. Dimension is $n \times n$ for all components.

See Also

[createXblocks](#), which is required when parameter estimates are desired (not needed for prediction). A full demo and data are available from:

<https://drive.google.com/open?id=1NUfeOtN8-KZ8A2HZzveG506nBwgW64e4>

Examples

```
#Example
#Simulate
Xb11 <- matrix(rnorm(1000),nrow=10)
Xb12 <- matrix(rnorm(2000),nrow=10)

#check whether dimensions are correct
ncol(Xb11)==nrow(Xb12)

#create cross-product
XXb1 <- createXXblocks(list(Xb11,Xb12))

#suppose penalties for two data types equal 5,10, respectively
Sigma <- SigmaFromBlocks(XXb1,c(5,10))
```

```
#check dimensions (should be n x n)
dim(Sigma)
```

CVfolds	<i>Creates (repeated) cross-validation folds</i>
---------	--

Description

Creates (repeated) cross-validation folds for samples

Usage

```
CVfolds(Y, model = NULL, balance = TRUE, kfold = 10, fixedfolds = TRUE, nrepeat = 1)
```

Arguments

Y	Response vector: numeric, binary, factor or survival.
model	Character. Any of c("linear", "logistic", "cox"). Is inferred from Y when NULL.
balance	Boolean. Should the splits be balanced in terms of response labels?
kfold	Integer. Desired fold.
fixedfolds	Boolean. Should fixed splits be used for reproducibility?
nrepeat	Numeric. Number of repeats.

Details

Creates (repeated), possibly balanced, splits of the samples. Computing time will often largely depend on on $kfold \times nrepeat$, the number of training-test splits evaluated.

Value

List object with $kfold \times nrepeat$ elements containing the sample indices of the left-out samples per split.

See Also

A full demo and data are available from:

<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzeG506nBwgW64e4>

CVscore	<i>Cross-validated score</i>
---------	------------------------------

Description

Cross-validated score for given penalty parameters.

Usage

```
CVscore(penalties, XXblocks, Y, X1 = NULL, pairing = NULL, folds, intercept =
ifelse(class(Y) == "Surv", FALSE, TRUE), frac1 = NULL, score = "loglik", model =
NULL, eps = 1e-07, maxItr = 100, trace = FALSE, printCV = TRUE, save = FALSE,
parallel = F)
```

Arguments

penalties	Numeric vector.
XXblocks	List of nxn matrices. Usually output of createXXblocks .
Y	Response vector: numeric, binary, factor or survival.
X1	Matrix. Dimension $n \times p_0$, $p_0 < n$, representing unpenalized covariates
pairing	Numerical vector of length 3 or NULL when pairs are absent. Represents the indices (in XXblocks) of the two data blocks involved in pairing, plus the index of the paired block.
folds	List of integer vector. Usually output of CVfolds .
intercept	Boolean. Should an intercept be included?
frac1	Scalar. Prior fraction of cases. Only relevant for model="logistic".
score	Character. See Details.
model	Character. Any of c("linear", "logistic", "cox"). Is inferred from Y when NULL.
eps	Scalar. Numerical bound for IWLS convergence.
maxItr	Integer. Maximum number of iterations used in IWLS.
trace	Boolean. Should the output of the IWLS algorithm be traced?
printCV	Boolean. Should the CV-score be printed on screen?
save	Boolean. If TRUE appends the penalties and resulting CVscore to global variable allscores
parallel	Boolean. Should computation be done in parallel? If TRUE, requires to run setupParallel first.

Details

See [Scoring](#) for details on score.

Value

Numeric, cross-validated prediction score for given penalties

See Also

[doubleCV](#) for double cross-validation, used for performance evaluation

doubleCV

*Double cross-validation for estimating performance of multiridge***Description**

Double cross-validation for estimating performance of `multiridge`. Outer fold is for testing, inner fold for penalty parameter tuning

Usage

```
doubleCV(penaltiesinit, XXblocks, Y, X1 = NULL, pairing = NULL, outfold = 5,
         infold = 10, nrepeatout = 1, nrepeatin = 1, balance = TRUE, fixedfolds =
         TRUE, intercept = ifelse(class(Y) == "Surv", FALSE, TRUE), frac1 = NULL,
         score = "loglik", model = NULL, eps = 1e-07, maxItr = 10, trace = FALSE,
         printCV = TRUE, reltol = 1e-04, optmethod1 = "SANN", optmethod2 =
         ifelse(length(penaltiesinit) == 1, "Brent", "Nelder-Mead"), maxItropt1 = 10,
         maxItropt2 = 25, save = FALSE, parallel = FALSE, pref = NULL, fixedpen = NULL)
```

Arguments

<code>penaltiesinit</code>	Numeric vector. Initial values for penalty parameters. May be obtained from fastCV2 .
<code>XXblocks</code>	List of $n \times n$ matrices. Usually output of createXXblocks .
<code>Y</code>	Response vector: numeric, binary, factor or survival.
<code>X1</code>	Matrix. Dimension $n \times p_0$, $p_0 < n$, representing unpenalized covariates
<code>pairing</code>	Numerical vector of length 3 or NULL when pairs are absent. Represents the indices (in <code>XXblocks</code>) of the two data blocks involved in pairing, plus the index of the paired block.
<code>outfold</code>	Integer. Outer fold for test samples.
<code>infold</code>	Integer. Inner fold for tuning penalty parameters.
<code>nrepeatout</code>	Integer. Number of repeated splits for outer fold.
<code>nrepeatin</code>	Integer. Number of repeated splits for inner fold.
<code>balance</code>	Boolean. Should the splits be balanced in terms of response labels?
<code>fixedfolds</code>	Boolean. Should fixed splits be used for reproducibility?
<code>intercept</code>	Boolean. Should an intercept be included?
<code>frac1</code>	Scalar. Prior fraction of cases. Only relevant for <code>model="logistic"</code> .
<code>score</code>	Character. See Details.
<code>model</code>	Character. Any of <code>c("linear", "logistic", "cox")</code> . Is inferred from <code>Y</code> when NULL.
<code>eps</code>	Scalar. Numerical bound for IWLS convergence.
<code>maxItr</code>	Integer. Maximum number of iterations used in IWLS.
<code>trace</code>	Boolean. Should the output of the IWLS algorithm be traced?
<code>printCV</code>	Boolean. Should the CV-score be printed on screen?
<code>reltol</code>	Scalar. Relative tolerance for optimization methods.

optmethod1	Character. First, global search method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but simulated annealing by "Sann" is recommended to search a wide landscape. Other unconstrained methods offered by optim may also be used, but have not been tested.
optmethod2	Character. Second, local search method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but "Nelder-Mead" is generally recommended. Other unconstrained methods offered by optim may also be used, but have not been tested.
maxItropt1	Integer. Maximum number of iterations for optmethod1.
maxItropt2	Integer. Maximum number of iterations for optmethod2.
save	Boolean. If TRUE appends the penalties and resulting CVscore to global variable allscores
parallel	Boolean. Should computation be done in parallel? If TRUE, requires to run setupParallel first.
pref	Integer vector or NULL. Contains indices of data types in XXblocks that are preferential.
fixedpen	Integer vector or NULL. Contains indices of data types of which penalty is fixed to the corresponding value in penaltiesinit.

Details

WARNING: this function may be very time-consuming. The number of evaluations may equal $nrepeatout \times outerfold \times nrepeatin \times innerfold \times maxItr \times (maxItropt1 + maxItropt2)$. Computing time may be estimated by multiplying computing time of [optLambdasWrap](#) by $nrepeatout \times outerfold$. See [Scoring](#) for details on score.

Value

List with the following components:

sampleindex	Numerical vector: sample indices
true	True responses
linpred	Cross-validated linear predictors

See Also

[optLambdas](#), [optLambdasWrap](#) which optimize the penalties. [Scoring](#) which may applied to output of this function to obtain overall cross-validated performance score. A full demo and data are available from:

<https://drive.google.com/open?id=1NUfeOtN8-KZ8A2HZzveG506nBwgW64e4>

fastCV2

Fast cross-validation per data block

Description

Fast cross-validation for high-dimensional data. Finds optimal penalties separately per data block. Useful for initialization.

Usage

```
fastCV2(XXblocks, Y, X1 = NULL, kfold = 10, intercept = ifelse(class(Y) ==
"Surv", FALSE, TRUE), parallel = FALSE, fixedfolds = TRUE, model = NULL,
eps = 1e-10, lambdamax = 10^6)
```

Arguments

XXblocks	List of data frames or matrices, representing $b=1, \dots, B$ data blocks of dimensions $n \times p_b$.
Y	Response vector: numeric, binary, factor or survival.
X1	Matrix. Dimension $n \times p_0, p_0 < n$, representing unpenalized covariates.
kfold	Integer. Desired fold.
intercept	Boolean. Should an intercept be included?
parallel	Boolean. Should computation be done in parallel? If TRUE, requires to run setupParallel first.
fixedfolds	Boolean. Should fixed splits be used for reproducibility?
model	Character. Any of <code>c("linear", "logistic", "cox")</code> . Is inferred from Y when NULL.
eps	Scalar. Numerical bound for IWLS convergence.
lambdamax	Numeric. Upperbound for lambda.

Details

This function is basically a wrapper for applying [optLambdas](#) per data block separately using Brent optimization.

Value

Numerical vector containing penalties optimized separately per data block. Useful for initialization.

See Also

[optLambdas](#), [optLambdasWrap](#) which optimize the penalties jointly. A full demo and data are available from:
<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzveG506nBwgW64e4>

IWLS Coxridge

Iterative weighted least squares algorithm for Cox ridge regression.

Description

Iterative weighted least squares algorithm for Cox ridge regression. Updates the weights and linear predictors until convergence.

Usage

```
IWLS Coxridge(XXT, Y, X1 = NULL, intercept = FALSE, eps = 1e-07, maxItr = 25,
trace = FALSE, E0 = NULL)
```

Arguments

XXT	Matrix. Dimensions $n \times n$. Sample cross-product from penalized variables, usually computed by SigmaFromBlocks .
Y	Response vector: class survival.
X1	Matrix. Dimension $n \times p_0$, $p_0 < n$, representing unpenalized covariates.
intercept	Boolean. Should an intercept be included?
eps	Scalar. Numerical bound for IWLS convergence.
maxItr	Integer. Maximum number of iterations used in IWLS.
trace	Boolean. Should the output of the IWLS algorithm be traced?
E0	Numerical vector or NULL. Optional initial values for linear predictor. Same length as Y. Usually NULL, which initializes linear predictor with 0.

Details

Usually, Cox ridge regression does not use an intercept, as this is part of the baseline hazard. The latter is estimated using the Breslow estimator. To keep the function computationally efficient it returns the linear predictors (which suffice for predictions), instead of parameter estimates. These may be obtained by applying the [betasout](#) function to the output of this function.

Value

List, containing:

etas	Numerical vector: Final linear predictors
Ypred	Predicted survival
convergence	Boolean: has IWLS converged?
nIt	Number of iterations
Hres	Auxiliary list object. Passed on to other functions
linearized	Linearized predictions
unpen	Boolean: are there any unpenalized covariates involved? Passed on to other functions
intercept	Boolean: Is an intercept included?
eta0	Numerical vector: Initial linear predictors

References

Mark A. van de Wiel, Mirrelijn van Nee, Armin Rauschenberger (2021). Fast cross-validation for high-dimensional ridge regression. J Comp Graph Stat

See Also

[IWLSridge](#) for linear and logistic ridge. [betasout](#) for obtaining parameter estimates. [predictIWLS](#) for predictions on new samples. A full demo and data are available from:
<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzveG506nBwgW64e4>

IWLSridge	<i>Iterative weighted least squares algorithm for linear and logistic ridge regression.</i>
-----------	---

Description

Iterative weighted least squares algorithm for linear and logistic ridge regression. Updates the weights and linear predictors until convergence.

Usage

```
IWLSridge(XXT, Y, X1 = NULL, intercept = TRUE, frac1 = NULL, eps = 1e-07,
maxItr = 25, trace = FALSE, model = NULL, E0 = NULL)
```

Arguments

XXT	Matrix. Dimensions nxn. Sample cross-product from penalized variables, usually computed by SigmaFromBlocks .
Y	Response vector: numeric, binary, or two-class factor
X1	Matrix. Dimension n x p_0, p_0 < n, representing unpenalized covariates.
intercept	Boolean. Should an intercept be included?
frac1	Scalar. Prior fraction of cases. Only relevant for model="logistic".
eps	Scalar. Numerical bound for IWLS convergence.
maxItr	Integer. Maximum number of iterations used in IWLS.
trace	Boolean. Should the output of the IWLS algorithm be traced?
model	Character. Any of c("linear", "logistic"). Is inferred from Y when NULL. Note that the cox model for survival is covered by the function IWLScoxridge .
E0	Numerical vector or NULL. Optional initial values for linear predictor. Same length as Y. Usually NULL, which initializes linear predictor with 0.

Details

An (unpenalized) intercept is included by default. To keep the function computationally efficient it returns the linear predictors (which suffice for predictions), instead of parameter estimates. These may be obtained by applying the [betasout](#) function to the output of this function.

Value

List, containing:

etas	Numerical vector: Final linear predictors
Ypred	Predicted survival
convergence	Boolean: has IWLS converged?
nIt	Number of iterations
Hres	Auxiliary list object. Passed on to other functions
linearized	Linearized predictions
unpen	Boolean: are there any unpenalized covariates involved? Passed on to other functions
intercept	Boolean: Is an intercept included?
eta0	Numerical vector: Initial linear predictors

References

Mark A. van de Wiel, Mirrelijn van Nee, Armin Rauschenberger (2021). Fast cross-validation for high-dimensional ridge regression. J Comp Graph Stat

See Also

[IWLSCoxridge](#) for Cox ridge. [betasout](#) for obtaining parameter estimates. [predictIWLS](#) for predictions on new samples. A full demo and data are available from:

<https://drive.google.com/open?id=1NUfeOtN8-KZ8A2HZzveG506nBwgW64e4>

mgcv_lambda	<i>Maximum marginal likelihood score</i>
-------------	--

Description

Computed maximum marginal likelihood score for given penalty parameters using mgcv.

Usage

```
mgcv_lambda(penalties, XXblocks, Y, model=NULL, printscore=TRUE, pairing=NULL, sigmasq = 1,
  opt.sigma=ifelse(model=="linear", T, F))
```

Arguments

penalties	Numeric vector.
XXblocks	List of nxn matrices. Usually output of createXXblocks .
Y	Response vector: numeric, binary, factor or survival.
model	Character. Any of c("linear", "logistic", "cox"). Is inferred from Y when NULL.
printscore	Boolean. Should the score be printed?
pairing	Numerical vector of length 3 or NULL when pairs are absent. Represents the indices (in XXblocks) of the two data blocks involved in pairing, plus the index of the paired block.
sigmasq	Default error variance.
opt.sigma	Boolean. Should the error variance be optimized as well? Only relevant for model="linear".

Details

See [gam](#) for details on how the marginal likelihood is computed.

Value

Numeric, marginal likelihood score for given penalties

References

Wood, S. N. (2011), Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, J. Roy. Statist. Soc., B 73(1), 3-36.

See Also

[CVscore](#) for cross-validation alternative. A full demo and data are available from:
<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzveG506nBwgW64e4>

mlikCV	<i>Outer-loop cross-validation for estimating performance of marginal likelihood based multiridge</i>
--------	---

Description

Outer-loop cross-validation for estimating performance of marginal likelihood based multiridge. Outer fold is for testing; penalty parameter tuning is performed by marginal likelihood estimation

Usage

```
mlikCV(penaltiesinit, XXblocks, Y, pairing = NULL, outfold = 5, nrepeatout = 1,
balance = TRUE, fixedfolds = TRUE, model = NULL, intercept = ifelse(class(Y) ==
"Surv", FALSE, TRUE), reltol = 1e-04, trace = FALSE, optmethod1 = "SANN",
optmethod2 = ifelse(length(penaltiesinit) == 1, "Brent", "Nelder-Mead"),
maxItropt1 = 10, maxItropt2 = 25, parallel = FALSE, pref = NULL,
fixedpen = NULL, sigmasq = 1, opt.sigma = ifelse(model == "linear", T, F))
```

Arguments

penaltiesinit	Numeric vector. Initial values for penalty parameters. May be obtained from fastCV2 .
XXblocks	List of nxn matrices. Usually output of createXXblocks .
Y	Response vector: numeric, binary, factor or survival.
pairing	Numerical vector of length 3 or NULL when pairs are absent. Represents the indices (in XXblocks) of the two data blocks involved in pairing, plus the index of the paired block.
outfold	Integer. Outer fold for test samples.
nrepeatout	Integer. Number of repeated splits for outer fold.
balance	Boolean. Should the splits be balanced in terms of response labels?
fixedfolds	Boolean. Should fixed splits be used for reproducibility?
intercept	Boolean. Should an intercept be included?
model	Character. Any of c("linear", "logistic", "cox"). Is inferred from Y when NULL.
trace	Boolean. Should the output of the IWLS algorithm be traced?
reltol	Scalar. Relative tolerance for optimization methods.
optmethod1	Character. First, global search method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but simulated annealing by "Sann" is recommended to search a wide landscape. Other unconstrained methods offered by optim may also be used, but have not been tested.
optmethod2	Character. Second, local search method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but "Nelder-Mead" is generally recommended. Other unconstrained methods offered by optim may also be used, but have not been tested.

maxItrOpt1	Integer. Maximum number of iterations for optmethod1.
maxItrOpt2	Integer. Maximum number of iterations for optmethod2.
parallel	Boolean. Should computation be done in parallel? If TRUE, requires to run setupParallel first.
pref	Integer vector or NULL. Contains indices of data types in XXblocks that are preferential.
fixedpen	Integer vector or NULL. Contains indices of data types of which penalty is fixed to the corresponding value in penaltiesinit.
sigmaSq	Default error variance.
opt.sigma	Boolean. Should the error variance be optimized as well? Only relevant for model="linear".

Details

WARNING: this function may be very time-consuming. The number of evaluations may equal $nrepeatout \times outerfold \times (maxItrOpt1 + maxItrOpt2)$. Computing time may be estimated by multiplying computing time of [optLambdas_mgcvWrap](#) by $nrepeatout \times outerfold$.

Value

List with the following components:

sampleindex	Numerical vector: sample indices
true	True responses
linpred	Cross-validated linear predictors

See Also

[optLambdas_mgcv](#), [optLambdas_mgcvWrap](#) which optimize the penalties. [Scoring](#) which may applied to output of this function to obtain overall cross-validated performance score. [doubleCV](#) for double cross-validation counterpart. A full demo and data are available from: <https://drive.google.com/open?id=1NUfeOtN8-KZ8A2HZzveG506nBwgW64e4>

optLambdas	<i>Find optimal ridge penalties.</i>
------------	--------------------------------------

Description

Optimizes a cross-validated score w.r.t. ridge penalties for multiple data blocks.

Usage

```
optLambdas(penaltiesinit = NULL, XXblocks, Y, X1 = NULL, pairing = NULL, folds,
  intercept = ifelse(class(Y) == "Surv", FALSE, TRUE), frac1 = NULL, score = "loglik",
  model = NULL, epsIWLS = 0.001, maxItrIWLS = 25, traceCV = TRUE, reltol = 1e-04,
  optmethod = ifelse(length(penaltiesinit) == 1, "Brent", "Nelder-Mead"), maxItrOpt = 500,
  save = FALSE, parallel = FALSE, fixedpen = NULL, fixedseed = TRUE)
```

Arguments

penaltiesinit	Numeric vector. Initial values for penaltyparameters. May be obtained from fastCV2 .
XXblocks	List of nxn matrices. Usually output of createXXblocks .
Y	Response vector: numeric, binary, factor or survival.
X1	Matrix. Dimension $n \times p_0$, $p_0 < n$, representing unpenalized covariates
pairing	Numerical vector of length 3 or NULL when pairs are absent. Represents the indices (in XXblocks) of the two data blocks involved in pairing, plus the index of the paired block.
folds	List, containing the splits of the samples. Usually obtained by CVfolds
intercept	Boolean. Should an intercept be included?
frac1	Scalar. Prior fraction of cases. Only relevant for model="logistic".
score	Character. See Details.
model	Character. Any of c("linear", "logistic", "cox"). Is inferred from Y when NULL.
epsIWLS	Scalar. Numerical bound for IWLS convergence.
maxItrIWLS	Integer. Maximum number of iterations used in IWLS.
traceCV	Boolean. Should the output of the IWLS algorithm be traced?
reltol	Scalar. Relative tolerance for optimization methods.
optmethod	Character. Optimization method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but "Nelder-Mead" is generally recommended. Other unconstrained methods offered by optim may also be used, but have not been tested.
maxItropt	Integer. Maximum number of iterations for optmethod.
save	Boolean. If TRUE appends the penalties and resulting CVscore to global variable allscores
parallel	Boolean. Should computation be done in parallel? If TRUE, requires to run setupParallel first.
fixedpen	Integer vector or NULL. Contains indices of data types of which penalty is fixed to the corresponding value in penaltiesinit.
fixedseed	Boolean. Should the initialization be fixed? For reproducibility.

Details

See [Scoring](#) for details on score. We highly recommend to use smooth scoring functions, in particular "loglik". For ranking-based criteria like auc and cindex we advise to use repeated CV (see [CVfolds](#)) to avoid ending up in any of the many local optima.

Value

List, with components:

optres	Output of the optimizer
optpen	Vector with determined optimal penalties
allsc	Matrix with CV scores for all penalty parameter configurations used by the optimizer

See Also

[optLambdasWrap](#) for i) (recommended) optimization in two steps: first global, then local; and ii) sequential optimization when some data types are preferred over others. [fastCV2](#) for initialization of penalties. A full demo and data are available from:

<https://drive.google.com/open?id=1NUfeOtN8-KZ8A2HZzeG506nBwgW64e4>

optLambdasWrap

Find optimal ridge penalties with sequential optimization.

Description

Sequentially optimizes a cross-validated score w.r.t. ridge penalties for multiple data blocks. Also implements preferential ridge, which allows to first optimize for the preferential data types.

Usage

```
optLambdasWrap(penaltiesinit = NULL, XXblocks, Y, X1 = NULL, pairing = NULL,
  folds, intercept = ifelse(class(Y) == "Surv", FALSE, TRUE), frac1 = NULL,
  score = "loglik", model = NULL, epsIWLS = 0.001, maxItrIWLS = 25,
  traceCV = TRUE, reltol = 1e-04, optmethod1 = "SANN", optmethod2 =
  ifelse(length(penaltiesinit) == 1, "Brent", "Nelder-Mead"), maxItropt1 = 10,
  maxItropt2 = 25, save = FALSE, parallel = FALSE, pref = NULL, fixedpen = NULL)
```

Arguments

penaltiesinit	Numeric vector. Initial values for penaltyparameters. May be obtained from fastCV2 .
XXblocks	List of nxn matrices. Usually output of createXXblocks .
Y	Response vector: numeric, binary, factor or survival.
X1	Matrix. Dimension $n \times p_0, p_0 < n$, representing unpenalized covariates
pairing	Numerical vector of length 3 or NULL when pairs are absent. Represents the indices (in XXblocks) of the two data blocks involved in pairing, plus the index of the paired block.
folds	List, containing the splits of the samples. Usually obtained by CVfolds
intercept	Boolean. Should an intercept be included?
frac1	Scalar. Prior fraction of cases. Only relevant for model="logistic".
score	Character. See Details.
model	Character. Any of c("linear", "logistic", "cox"). Is inferred from Y when NULL.
epsIWLS	Scalar. Numerical bound for IWLS convergence.
maxItrIWLS	Integer. Maximum number of iterations used in IWLS.
traceCV	Boolean. Should the output of the IWLS algorithm be traced?
reltol	Scalar. Relative tolerance for optimization methods.
optmethod1	Character. First, global search method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but simulated annealing by "Sann" is recommended to search a wide landscape. Other unconstrained methods offered by optim may also be used, but have not been tested.

optmethod2	Character. Second, local search method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but "Nelder-Mead" is generally recommended. Other unconstrained methods offered by <code>optim</code> may also be used, but have not been tested.
maxItropt1	Integer. Maximum number of iterations for optmethod1.
maxItropt2	Integer. Maximum number of iterations for optmethod2.
save	Boolean. If TRUE appends the penalties and resulting CVscore to global variable allscores
parallel	Boolean. Should computation be done in parallel? If TRUE, requires to run <code>setupParallel</code> first.
pref	Integer vector or NULL. Contains indices of data types in XXblocks that are preferential.
fixedpen	Integer vector or NULL. Contains indices of data types of which penalty is fixed to the corresponding value in penaltiesinit.

Details

As opposed to `optLambdas` this function first searches globally, then locally. Hence, more time-consuming, but better guarded against multiple local optima.

See `Scoring` for details on score. We highly recommend to use smooth scoring functions, in particular "loglik". For ranking-based criteria like "auc" and "cindex" we advise to use repeated CV (see `CVfolds`) to avoid ending up in any of the many local optima.

Value

List, with components:

res	Outputs of all optimizers used
lambdas	List of penalties found by the optimizers
optpen	Numerical vector with final, optimal penalties

See Also

`optLambdas` for one-pass optimization. `fastCV2` for initialization of penalties. A full demo and data are available from:

<https://drive.google.com/open?id=1NUfeOtN8-KZ8A2HZzveG506nBwgW64e4>

optLambdas_mgcv

Find optimal ridge penalties with maximum marginal likelihood

Description

Optimizes a marginal likelihood score w.r.t. ridge penalties for multiple data blocks.

Usage

```
optLambdas_mgcv(penaltiesinit=NULL, XXblocks,Y, pairing=NULL, model=NULL, reltol=1e-4,
  optmethod=ifelse(length(penaltiesinit)==1,"Brent", "Nelder-Mead"),maxItropt=500,
  tracescore=TRUE, fixedpen=NULL, fixedseed =TRUE, sigmasq = 1,
  opt.sigma=ifelse(model=="linear",T, F))
```

Arguments

penaltiesinit	Numeric vector. Initial values for penalty parameters. May be obtained from fastCV2 .
XXblocks	List of nxn matrices. Usually output of createXXblocks .
Y	Response vector: numeric, binary, factor or survival.
pairing	Numerical vector of length 3 or NULL when pairs are absent. Represents the indices (in XXblocks) of the two data blocks involved in pairing, plus the index of the paired block.
model	Character. Any of c("linear", "logistic", "cox"). Is inferred from Y when NULL.
reltol	Scalar. Relative tolerance for optimization methods.
optmethod	Character. Optimization method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but "Nelder-Mead" is generally recommended. Other unconstrained methods offered by optim may also be used, but have not been tested.
maxItropt	Integer. Maximum number of iterations for optmethod.
tracescore	Boolean. Should the output of the scores be traced?
fixedpen	Integer vector or NULL. Contains indices of data types of which penalty is fixed to the corresponding value in penaltiesinit.
fixedseed	Boolean. Should the initialization be fixed? For reproducibility.
sigmasq	Default error variance.
opt.sigma	Boolean. Should the error variance be optimized as well? Only relevant for model="linear".

Details

See [gam](#) for details on how the marginal likelihood is computed.

Value

List, with components:

optres	Output of the optimizer
optpen	Vector with determined optimal penalties
allsc	Matrix with marginal likelihood scores for all penalty parameter configurations used by the optimizer

See Also

[optLambdas_mgcvWrap](#) for i) (recommended) optimization in two steps: first global, then local; and ii) sequential optimization when some data types are preferred over others. A full demo and data are available from:

<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzveG506nBwgW64e4>

optLambdas_mgcvWrap *Find optimal ridge penalties with sequential optimization.*

Description

Sequentially optimizes a marginal likelihood score w.r.t. ridge penalties for multiple data blocks.

Usage

```
optLambdas_mgcvWrap(penaltiesinit=NULL, XXblocks,Y, pairing=NULL, model=NULL, reltol=1e-4,
  optmethod1= "SANN", optmethod2 =ifelse(length(penaltiesinit)==1,"Brent", "Nelder-Mead"),
  maxItropt1=10,maxItropt2=25, tracescore=TRUE,fixedseed =TRUE, pref=NULL, fixedpen=NULL,
  sigmasq = 1, opt.sigma=ifelse(model=="linear",T, F))
```

Arguments

penaltiesinit	Numeric vector. Initial values for penaltyparameters. May be obtained from fastCV2 .
XXblocks	List of nxn matrices. Usually output of createXXblocks .
Y	Response vector: numeric, binary, factor or survival.
pairing	Numerical vector of length 3 or NULL when pairs are absent. Represents the indices (in XXblocks) of the two data blocks involved in pairing, plus the index of the paired block.
model	Character. Any of c("linear", "logistic", "cox"). Is inferred from Y when NULL.
reltol	Scalar. Relative tolerance for optimization methods.
optmethod1	Character. First, global search method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but simulated annealing by "Sann" is recommended to search a wide landscape. Other unconstrained methods offered by optim may also be used, but have not been tested.
optmethod2	Character. Second, local search method. Any of the methods c("Brent", "Nelder-Mead", "Sann") may be used, but "Nelder-Mead" is generally recommended. Other unconstrained methods offered by optim may also be used, but have not been tested.
maxItropt1	Integer. Maximum number of iterations for optmethod1.
maxItropt2	Integer. Maximum number of iterations for optmethod2.
tracescore	Boolean. Should the output of the scores be traced?
fixedseed	Boolean. Should the initialization be fixed? For reproducibility.
pref	Integer vector or NULL. Contains indices of data types in XXblocks that are preferential.
fixedpen	Integer vector or NULL. Contains indices of data types of which penalty is fixed to the corresponding value in penaltiesinit.
sigmasq	Default error variance.
opt.sigma	Boolean. Should the error variance be optimized as well? Only relevant for model="linear".

Details

As opposed to [optLambdas_mgcv](#) this function first searches globally, then locally. Hence, more time-consuming, but better guarded against multiple local optima. See [gam](#) for details on how the marginal likelihood is computed.

Value

List, with components:

res	Outputs of all optimizers used
lambdas	List of penalties found by the optimizers
optpen	Numerical vector with final, optimal penalties

References

Wood, S. N. (2011), Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models, J. Roy. Statist. Soc., B 73(1), 3-36.

See Also

[optLambdas_mgcv](#) for one-pass optimization. A full demo and data are available from:
<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzveG506nBwgW64e4>

predictIWLS	<i>Predictions from ridge fits</i>
-------------	------------------------------------

Description

Produces predictions from ridge fits for new data.

Usage

```
predictIWLS(IWLSfit, X1new = NULL, Sigmanew)
```

Arguments

IWLSfit	List, containing fits from either IWLSridge (linear, logistic ridge) or IWLScoxridge
X1new	Matrix. Dimension nnew x p_0, representing unpenalized covariates for new data.
Sigmanew	Matrix. Dimensions nnew x n. Sample cross-product from penalized variables, usually computed by first applying createXXblocks and then SigmaFromBlocks .

Details

Predictions rely purely on the linear predictors, and do not require producing the parameter vector.

Value

Numerical vector of linear predictor for the test samples.

See Also

[IWLsRidge](#) ([IWLSCoxridge](#)) for fitting linear and logistic ridge (Cox ridge). [betasout](#) for obtaining parameter estimates. [Scoring](#) to evaluate the predictions. A full demo and data are available from: <https://drive.google.com/open?id=1NUfeOtN8-KZ8A2HZzeG506nBwgW64e4>

Examples

```
#Example below shows how to create the input argument Sigmanew (for simulated data)
#Simulate
Xbl1 <- matrix(rnorm(1000),nrow=10)
Xbl2 <- matrix(rnorm(2000),nrow=10)
Xbl1new <- matrix(rnorm(200),nrow=2)
Xbl2new <- matrix(rnorm(400),nrow=2)

#check whether dimensions are correct
nrow(Xbl1)==nrow(Xbl1new)
nrow(Xbl2)==nrow(Xbl2new)
ncol(Xbl1)==nrow(Xbl2)
ncol(Xbl1new)==ncol(Xbl2new)

#create cross-product
XXbl <- createXXblocks(list(Xbl1,Xbl2),list(Xbl1new,Xbl2new))

#suppose penalties for two data types equal 5,10, respectively
Sigmanew <- SigmaFromBlocks(XXbl,c(5,10))

#check dimensions (should be nnew x n)
dim(Sigmanew)
```

Scoring

Evaluate predictions

Description

Evaluates predictions by a score suitable for the corresponding response

Usage

```
Scoring(lp, Y, model = NULL, score = ifelse(model == "linear", "mse", "loglik"),
  print = TRUE)
```

Arguments

lp	Numerical vector. Linear predictor.
Y	Response vector: numeric, binary, factor or survival.
score	Character. See Details.
model	Character. Any of c("linear", "logistic", "cox"). Is inferred from Y when NULL.
print	Boolean. Should the score be printed on screen.

Details

Several scores are allowed, depending on the type of output. For `model = "linear"`, score equals any of `c("loglik", "mse", "abserror", "cor", "kendall", "spearman")`, denoting CV-ed log-likelihood, mean-squared error, mean absolute error, Pearson (Kendall, Spearman) correlation with response. For `model = "logistic"`, score equals any of `c("loglik", "auc", "brier")`, denoting CV-ed log-likelihood, area-under-the-ROC-curve, and brier score a.k.a. MSE. For `model = "cox"`, score equals any of `c("loglik", "cindex")`, denoting CV-ed log-likelihood, and c-index.

Value

Numerical value.

See Also

[CVscore](#) for obtaining the cross-validated score (for given penalties), and [doubleCV](#) to obtain doubly cross-validated linear predictors to which `Scoring` can be applied to estimated predictive performance by double cross-validation. A full demo and data are available from:

<https://drive.google.com/open?id=1NUfe0tN8-KZ8A2HZzveG506nBwgW64e4>

setupParallel

Setting up parallel computing

Description

This function sets up parallel computing by the package `snowfall`.

Usage

```
setupParallel(ncpus = 1, sourcefile = NULL, sourcelibraries =
c("multiridge", "survival", "pROC", "risksetROC"))
```

Arguments

<code>ncpus</code>	Integer. Number of cpus to use.
<code>sourcefile</code>	Character. Additional source files to be loaded in parallel. Only required when parallel computing is also desired for functions not available in <code>multiridge</code> .
<code>sourcelibraries</code>	Character vector. Libraries to be loaded in parallel. Defaults to the libraries <code>multiridge</code> depends on.

Details

Parallel computing is available for several functions that rely on cross-validation. If double CV is used, parallel computing is applied to the outer loop, to optimize efficiency.

Value

No output.

See Also

Snowfall package for further documentation on parallel computing. A full demo and data are available from:

<https://drive.google.com/open?id=1NUfeOtN8-KZ8A2HZzeG506nBwgW64e4>

SigmaFromBlocks

Create penalized sample cross-product matrix

Description

Creates penalized sample cross-product matrix, dimension $n \times n$.

Usage

```
SigmaFromBlocks(XXblocks, penalties, pairing = NULL)
```

Arguments

XXblocks	List of $n \times n$ matrices. Usually output of <code>createXXblocks</code> .
penalties	Numeric vector, representing penalty parameters.
pairing	Numerical vector of length 3 or NULL when pairs are absent. Represents the indices (in XXblocks) of the two data blocks involved in pairing, plus the index of the paired block.

See Also

A full demo and data are available from:

<https://drive.google.com/open?id=1NUfeOtN8-KZ8A2HZzeG506nBwgW64e4>

Examples

```
#Example
#Simulate
Xbl1 <- matrix(rnorm(1000),nrow=10)
Xbl2 <- matrix(rnorm(2000),nrow=10)

#check whether dimensions are correct
ncol(Xbl1)==nrow(Xbl2)

#create cross-product
XXbl <- createXXblocks(list(Xbl1,Xbl2))

#suppose penalties for two data types equal 5,10, respectively
Sigma <- SigmaFromBlocks(XXbl,c(5,10))

#check dimensions (should be n x n)
dim(Sigma)
```


Index

- * **package**
 - [multiridge-package](#), [2](#)
- [augment](#), [3](#)
- [betasout](#), [2](#), [3](#), [11–13](#), [22](#)
- [createXblocks](#), [4](#), [4](#), [5](#)
- [createXXblocks](#), [2–4](#), [5](#), [7](#), [8](#), [13](#), [14](#), [16](#), [17](#),
[19–21](#), [24](#)
- [CVfolds](#), [6](#), [7](#), [16–18](#)
- [CVscore](#), [2](#), [7](#), [14](#), [23](#)
- [doubleCV](#), [2](#), [7](#), [8](#), [15](#), [23](#)
- [fastCV2](#), [2](#), [8](#), [9](#), [14](#), [16–20](#)
- [gam](#), [13](#), [19](#), [21](#)
- [IWLScoxridge](#), [2](#), [4](#), [10](#), [12](#), [13](#), [21](#), [22](#)
- [IWLSridge](#), [2](#), [4](#), [11](#), [12](#), [21](#), [22](#)
- [mgcv_lambda](#), [13](#)
- [mlikCV](#), [2](#), [14](#)
- [multiridge \(multiridge-package\)](#), [2](#)
- [multiridge-package](#), [2](#)
- [optim](#), [9](#), [14](#), [16–20](#)
- [optLambdas](#), [9](#), [10](#), [15](#), [18](#)
- [optLambdas_mgcv](#), [15](#), [18](#), [21](#)
- [optLambdas_mgcvWrap](#), [2](#), [15](#), [19](#), [20](#)
- [optLambdasWrap](#), [2](#), [9](#), [10](#), [17](#), [17](#)
- [predictIWLS](#), [2](#), [11](#), [13](#), [21](#)
- [Scoring](#), [7](#), [9](#), [15](#), [16](#), [18](#), [22](#), [22](#)
- [setupParallel](#), [2](#), [7](#), [9](#), [10](#), [15](#), [16](#), [18](#), [23](#)
- [SigmaFromBlocks](#), [2](#), [11](#), [12](#), [21](#), [24](#)