

Package ‘CSCNet’

August 16, 2022

Type Package

Title Fitting and Tuning Regularized Cause-Specific Cox Models with Elastic-Net Penalty

Version 0.1.0

Maintainer Shahin Roshani <s.roshani@nki.nl>

Description Flexible tools to fit, tune and obtain absolute risk predictions from regularized cause-specific cox models with elastic-net penalty.

License GPL (>= 3)

Encoding UTF-8

RoxygenNote 7.2.0

Depends R (>= 2.10), tidyverse (>= 1.3.1), magrittr (>= 2.0.3)

Imports dplyr (>= 1.0.9), purrr (>= 0.3.4), stringr (>= 1.4.0), tibble (>= 3.1.7), survival (>= 3.3-1), prodlim (>= 2019.11.13), riskRegression (>= 2022.03.22), glmnet (>= 4.1-4), caret (>= 6.0-92), recipes (>= 0.2.0), future (>= 1.26.1), furr (>= 0.3.0)

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Author Shahin Roshani [aut, cre, cph]
(<<https://orcid.org/0000-0001-7518-157X>>)

Repository CRAN

Date/Publication 2022-08-16 09:20:15 UTC

R topics documented:

penCSC	2
predict.penCSC	3
predictRisk.penCSC	5
print.penCSC	6
print.tune_penCSC	7
tune_penCSC	7

penCSC	<i>penCSC</i>
--------	---------------

Description

Function to fit penalized cause-specific-cox with elastic-net penalty.

Usage

```
penCSC(
  time,
  status,
  vars.list,
  data,
  alpha.list,
  lambda.list,
  standardize = TRUE,
  keep = NULL
)
```

Arguments

<code>time</code>	A character showing the name of the time variable in the data.
<code>status</code>	A character showing the name of the status/event variable in the data.
<code>vars.list</code>	A named list containing the variables to be included in each cause-specific model. Variables can be vectors of variable names or a one sided formula. Names of the list must be the events and exactly the same as values in the status variable. See ‘Examples’ for details.
<code>data</code>	A data frame containing the information of the variables.
<code>alpha.list</code>	A named list containing the single alpha values of each cause-specific model. Names of the list must be the events and exactly the same as values in the status variable. See ‘Examples’ for details.
<code>lambda.list</code>	A named list containing the single lambda values of each cause-specific model. Names of the list must be the events and exactly the same as values in the status variable. See ‘Examples’ for details.
<code>standardize</code>	Logical indicating whether the variables must be standardized or not. Default is TRUE.
<code>keep</code>	A character vector of the names of variables that should not be shrunk. Default is NULL.

Value

A named list containing all the information related to the used data and the fitted models for all causes. Use \$ to explore all the involved information.

Author(s)

Shahin Roshani

References

Friedman J, Hastie T, Tibshirani R (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software*, 33(1), 1-22. doi:10.18637/jss.v033.i01, <https://www.jstatsoft.org/v33/i01/>.

Therneau T (2022). A Package for Survival Analysis in R. R package version 3.3-1, <https://CRAN.R-project.org/package=survival>.

Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. *J Open Source Softw.* 2019 Nov 21;4(43):1686.

Bache S, Wickham H (2022). magrittr: A Forward-Pipe Operator for R. <https://magrittr.tidyverse.org>, <https://github.com/tidyverse/magrittr>.

Examples

```
library(riskRegression)

data(Melanoma)

v1 <- list('1'=c('age','sex','ulcer','thick'),
          '2'=~age+sex+epicel+thick+ici)

a1 <- list('1'=0, '2'=.5)

l1 <- list('1'=.01, '2'=.04)

penCSC(time='time', status='status', vars.list=v1,
        data=Melanoma, alpha.list=a1, lambda.list=l1)
```

predict.penCSC

predict.penCSC

Description

Flexible prediction method for the objects of class 'penCSC' including the absolute risk prediction.

Usage

```
## S3 method for class 'penCSC'
predict(object, newX, event = NULL, time, type = "lp", reference = "zero", ...)
```

Arguments

object	An object of class 'penCSC'.
newX	A data frame containing the information of variables related to new records. Information of variables not included in the model creation will be ignored.
event	A vector of event codes which we want predictions for. This must be the same as values in the status variable of the data that was used to create the models. If NULL, absolute risk will be calculated for all involved events. Default is NULL which returns values for all involved causes.
time	A vector of time horizons which we want absolute risk predictions at. Only applicable when type='absRisk'.
type	Type of the predictions. Valid values are: 'lp' or 'link' for linear predictors, 'risk' or 'response' for exp(lp) and finally 'absRisk' for semi-parametric estimates of absolute risk.
reference	Reference for centering predictions. Valid values are 'zero' and 'sample'. Default is 'zero'. For more information on referencing see details in ?predict.coxph.
...	Additional arguments. Not used by predict.penCSC.

Value

A tibble containing the predictions based on the input arguments.

Author(s)

Shahin Roshani

References

- Pfeiffer, R. M., & Gail, M. M. (2017). Absolute risk: Methods and applications in clinical management and public health.
- Aalen, O.O. (1978) Nonparametric Inference for a Family of Counting Processes. The Annals of Statistics, 6, 701-726. doi:10.1214/aos/1176344247.
- Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. J Open Source Softw. 2019 Nov 21;4(43):1686.
- Bache S, Wickham H (2022). magrittr: A Forward-Pipe Operator for R. <https://magrittr.tidyverse.org>, <https://github.com/tidyverse/magrittr>.
- Friedman J, Hastie T, Tibshirani R (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." Journal of Statistical Software, 33(1), 1-22. doi:10.18637/jss.v033.i01, <https://www.jstatsoft.org/v33/i01/>.

Examples

```
library(riskRegression)

data(Melanoma)
```

```

v1 <- list('1'=c('age','sex','ulcer','thick'),
           '2'=~age+sex+epicel+thick+ici)

a1 <- list('1'=0,'2'=.5)

l1 <- list('1'=.01,'2'=.04)

penfit <- penCSC(time='time',status='status',vars.list=v1,
                 data=Melanoma,alpha.list=a1,lambda.list=l1)

predict(penfit,Melanoma[1:5,],type='lp')

predict(penfit,Melanoma[1:5,],type='response')

predict(penfit,Melanoma[1:5,],type='absRisk',event=1:2,time=1825*(1:2))

```

predictRisk.penCSC *predictRisk.penCSC*

Description

predictRisk method for absolute risk prediction. This is mainly for compatibility of 'CSCNet' with functions of 'riskRegression' package.

Usage

```

## S3 method for class 'penCSC'
predictRisk(object, newdata, times, cause, ...)

```

Arguments

object	An object of class 'penCSC'.
newdata	A data frame containing the variable information of new records.
times	A vector of time horizons which we want the absolute risk predictions at.
cause	A single value indicating the event of interest which we want the absolute risk predictions for. This value should be one of the values in the status variable of the data.
...	Additional arguments. Not used by predictRisk.penCSC.

Value

A matrix with columns of absolute risk predictions of individuals for each requested time horizon.

Author(s)

Shahin Roshani

References

Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. *J Open Source Softw.* 2019 Nov 21;4(43):1686.

Bache S, Wickham H (2022). *magrittr: A Forward-Pipe Operator for R.* <https://magrittr.tidyverse.org>, <https://github.com/tidyverse/magrittr>.

See Also

<https://www.rdocumentation.org/packages/riskRegression/versions/1.3.7/topics/predictRisk>
 Details in: <https://rdrr.io/cran/riskRegression/man/Score.html>

Examples

```
library(riskRegression)

data(Melanoma)

v1 <- list('1'=c('age','sex','ulcer','thick'),
          '2'=~age+sex+epicel+thick+ici)

a1 <- list('1'=0,'2'=.5)

l1 <- list('1'=.01,'2'=.04)

penfit <- penCSC(time='time',status='status',vars.list=v1,
                data=Melanoma,alpha.list=a1,lambda.list=l1)

predictRisk(penfit,Melanoma[1:5,],times=1825*(1:2),cause=1)
```

`print.penCSC`

print.penCSC

Description

Internal method for printing the objects of class `penCSC`.

Usage

```
## S3 method for class 'penCSC'
print(x, ...)
```

Arguments

`x` An object of class `penCSC`.
`...` Other arguments. Not used by `print.penCSC`.

Value

A modified print of penCSC objects.

Author(s)

Shahin Roshani

`print.tune_penCSC` *print.tune_penCSC*

Description

Internal method for printing the objects of class `tune_penCSC`.

Usage

```
## S3 method for class 'tune_penCSC'  
print(x, ...)
```

Arguments

`x` An object of class `tune_penCSC`.
`...` Other arguments. Not used by `print.tune_penCSC`.

Value

A modified print of `tune_penCSC` objects.

Author(s)

Shahin Roshani

`tune_penCSC` *tune_penCSC*

Description

A flexible function for tuning the involved hyper-parameters of a penalized cause-specific-cox model with elastic net penalty using the linking idea.

Usage

```
tune_penCSC(
  time,
  status,
  vars.list,
  data,
  horizons,
  event,
  rhs = ~1,
  method = "cv",
  k = 10,
  times = 25,
  p = 0.7,
  metrics = "Brier",
  final.metric = NULL,
  alpha.grid = NULL,
  lambda.grid = NULL,
  nlambdas.list = NULL,
  grow.by = 0.01,
  standardize = TRUE,
  keep = NULL,
  preProc.fun = function(x) x,
  parallel = FALSE,
  preProc.pkgs = NULL,
  preProc.globals = NULL,
  core.num = future::availableCores()/2
)
```

Arguments

time	A character showing the name of the time variable in the data.
status	A character showing the name of the status/event variable in the data.
vars.list	A named list containing the variables to be included in each cause-specific model. Variables can be vectors of variable names or a one sided formula. Names of the list must be the events and exactly the same as values in the status variable. See 'Examples' for details.
data	A data frame containing the information of the variables.
horizons	A vector of time horizons which we want the absolute risk predictions to be evaluated at.
event	The value for event of interest which we want the absolute risk predictions to be evaluated for. This must be one of the values in the status variable of the data.
rhs	A right hand sided formula indicating the variables to be used in estimating the inverse probability of censoring weighting (IPCW) model. Default is ~ 1 .
method	Resampling method to be used for hyper-parameter tuning. Values can be: 'cv' for cross validation, 'repcv' for repeated cross validation, 'lgocv' for monte-carlo cross validation, 'loocv' for leave one out cross validation and 'boot' for bootstrap. Default is 'cv'.

<code>k</code>	Number of folds. Only applicable for <code>method='cv'</code> and <code>method='repcv'</code> . Default is 10.
<code>times</code>	Repeat number of the resampling process. Only applicable for <code>method='repcv'</code> , <code>method='lgocv'</code> and <code>method='boot'</code> . Default is 25.
<code>p</code>	The fraction of data to be used as the training set during resampling. Only applicable for <code>method='lgocv'</code> . Default is 0.7.
<code>metrics</code>	Evaluation metric (loss function) to be used. Values can be 'Brier' for IPCW brier score, 'AUC' for IPCW AUC or a vector of both. Default is 'Brier'.
<code>final.metric</code>	The evaluation metric to decide the best hyper-parameters set for the final fits on the whole data. When NULL which is the default value, it takes the value from <code>metrics</code> . If both 'Brier' and 'AUC' were specified in <code>metrics</code> and <code>final.metric</code> is NULL, 'Brier' will be used.
<code>alpha.grid</code>	A named list containing a sequence of alpha values to be evaluated for each cause-specific model. Names of the list must be the events and exactly the same as values in the status variable. Default is NULL which orders the function to set <code>seq(0, 1, .5)</code> for all cause-specific models. See 'Details' for more information.
<code>lambda.grid</code>	A named list containing a sequence of lambda values to be evaluated for each cause-specific model. Names of the list must be the events and exactly the same as values in the status variable. Default is NULL which orders the function to calculate exclusive lambda sequences for all causes. See 'Details' for more information.
<code>nlambda.list</code>	A names list of single integers indicating the length of lambda sequences which are calculated automatically by the function for each cause. Only applicable when <code>lambda.grid=NULL</code> . Default is NULL which sets all lengths to 5. See 'Details' for more information.
<code>grow.by</code>	Difference between the values in the growing sequence of lambda values to find the maximum value that makes the null model. Only applicable when <code>lambda.grid=NULL</code> . Default is 0.01. See 'Details' for more information.
<code>standardize</code>	Logical indicating whether the variables must be standardized or not during model fitting procedures. Default is TRUE.
<code>keep</code>	A character vector of the names of variables that should not be shrunk in all model fitting procedures. Default is NULL.
<code>preProc.fun</code>	A function that accepts a data frame and returns a modified data that has gone through the user's desired pre-processing steps. All modifications from this function will be done during the resampling procedures to avoid data leakage. Default is <code>function(x) x</code> .
<code>parallel</code>	Logical indicating whether the tuning process should be performed in parallel or not. Default is FALSE.
<code>preProc.pkgs</code>	A character vector containing the names of packages that was used in creating user's <code>preProc.fun</code> while using parallel computation. Only applicable if <code>parallel=T</code> and <code>preProc.fun</code> is a user specified function using functions from other packages. See 'Examples' for details.
<code>preProc.globals</code>	A character vector containing names of objects included in <code>preProc.fun</code> to be considered as global objects while using parallel computation. The most

frequent ones are the names of the user specified pre processing function or functions within this function. Only applicable if `parallel=T` and `preProc.fun` is a user specified function. See 'Examples' for details.

`core.nums` Number of CPU cores to be used for parallel computation. Only applicable if `parallel=T`. Default is `future::availableCores()/2`.

Details

`tune_penCSC` has the ability to automatically determine the candidate sequences of alpha & lambda values. Setting any of `alpha.grid` & `lambda.grid` to `NULL` will order the function to calculate them automatically. The process of determining the lambda values automatically is by:

1. Starting from `lambda=0`, the algorithm fits LASSO models until finding a lambda value that creates a `NULL` model where all variables were shrunk to be exactly zero.
2. The obtained lambda value will be used as the maximum value of a sequence starting from 0. The length of this sequence is controlled by values in `nlambda.list`.

This will be done for each cause-specific model to create exclusive sequences of lambdas for each of them.

Value

A list containing the detailed information of the hyper-parameter tuning and the validation process, best combination of hyper-parameters and the final fits based on the whole data using the best obtained hyper-parameters. Use `$` to explore all the involved information.

Author(s)

Shahin Roshani

References

- Friedman J, Hastie T, Tibshirani R (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software*, 33(1), 1-22. doi:10.18637/jss.v033.i01, <https://www.jstatsoft.org/v33/i01/>.
- Saadati, M, Beyersmann, J, Kopp-Schneider, A, Benner, A. Prediction accuracy and variable selection for penalized cause-specific hazards models. *Biometrical Journal*. 2018; 60: 288– 306. doi:10.1002/bimj.201600242.
- Gerds TA, Kattan MW (2021). *Medical Risk Prediction Models: With Ties to Machine Learning* (1st ed.). Chapman and Hall/CRC. doi:10.1201/9781138384484
- Pfeiffer, R. M., & Gail, M. M. (2017). *Absolute risk: Methods and applications in clinical management and public health*.
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5), 1–26. doi:10.18637/jss.v028.i05.
- Bengtsson H (2021). "A Unifying Framework for Parallel and Distributed Processing in R using Futures." *The R Journal*, 13(2), 208–227. doi:10.32614/RJ2021048.
- Vaughan D, Dancho M (2022). *furrr: Apply Mapping Functions in Parallel using Futures*. <https://github.com/DavisVaughan/furrr>, <https://furrr.futureverse.org/>.

Therneau T (2022). A Package for Survival Analysis in R. R package version 3.3-1, <https://CRAN.R-project.org/package=survival>.

Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. J Open Source Softw. 2019 Nov 21;4(43):1686.

Bache S, Wickham H (2022). magrittr: A Forward-Pipe Operator for R. <https://magrittr.tidyverse.org>, <https://github.com/tidyverse/magrittr>.

Examples

```
library(riskRegression)

data(Melanoma)

v1 <- list('1'=~age+sex+epicel+ici,
           '2'=c('age','ulcer','thick','invasion'))

a1 <- list('1'=0,'2'=c(.5,1))

#External standardization function with data frame as its input and output

library(recipes)

std.fun <- function(data){
  cont_vars <- data %>% select(where(~is.numeric(.))) %>% names
  cont_vars <- cont_vars[-which(cont_vars %in% c('time','status'))]

  #External functions from recipes package are being used

  recipe(~.,data=data) %>%
    step_center(all_of(cont_vars)) %>%
    step_scale(all_of(cont_vars)) %>%
    prep(training=data) %>% juice
}

set.seed(233)

test <- tune_penCSC(time='time',status='status',vars.list=v1,data=Melanoma,horizons=1825,
                   event=1,method='cv',k=5,metrics='AUC',alpha.grid=a1,standardize=FALSE,
                   preProc.fun=std.fun,parallel=TRUE,preProc.pkgs='recipes')

test
```


Index

penCSC, [2](#)
predict.penCSC, [3](#)
predictRisk.penCSC, [5](#)
print.penCSC, [6](#)
print.tune_penCSC, [7](#)

tune_penCSC, [7](#)