Package 'plasma'

July 20, 2022

Title Partial LeAst Squares for Multiomic Analysis

Version 0.5.12

Date 2022-07-13

Author Kevin R. Coombes, Kyoko Yamaguchi

Maintainer Kevin R. Coombes <krc@silicovore.com>

Description Contains tools for supervised analyes of incomplete, overlapping multi-omics datasets.

Depends R (>= 3.5.0)

Imports methods, stats, graphics, survival, plsRcox, pls, Polychrome (>= 1.5.0), viridisLite

Suggests knitr, rmarkdown, oompaBase

License Apache License (== 2.0)

LazyLoad yes

URL http://oompa.r-forge.r-project.org/

VignetteBuilder knitr

NeedsCompilation no

R topics documented:

Contribution-class *Class* "Contribution"

Description

The Contribution object class contains the weight matrix between variables and the PLS components. The values in the weight matrix are a numeric representation of how much a variable from the omics datasets contributed to defining the final PLS components.

Usage

```
getCompositeWeights(object, N, M)
getAllWeights(object, N)
getTop(object, N = 1)
pickSignificant(object, alpha)
## S4 method for signature 'Contribution'
summary(object, ...)
## S4 method for signature 'Contribution'
image(x, col = viridis(64), mai = c(1.82, 1.52, 0.32, 0.32), ...)
## S4 method for signature 'Contribution'
heat(object, main = "Contributions", col = viridis(64),
                                        mai = c(1.52, 0.32, 0.82, 1.82), ...)
```

Arguments

object	In the first four functions, an object of the plasma class. In the methods described here, an object of the Contributions class.
N	in the function getCompositeWeights, the name of the dataset being modeled. in the function getTop, the number of significant components you want to print.
M	name of the dataset being modeled pairwise with dataset N in the getCompositeWeights function.
alpha	level of significance used in the pickSignificant function.
...	other graphical parameters.
x	an object of the Contributions class.
main	A character vector of length one; the main plot title.
col	A vector of color descriptors.
mai	A vector of four nonnegative numbers.

Value

The plasma function returns a newly constructed object of the plasma class.

Objects from the Class

Objects are defined using the getAllWeights, getCompositeWeights, getTop, or pickSignificant functions. In the simplest scenario, one would enter an object of class plasma and any specific parameters associated with the function (see arguments section for more info).

Slots

`contrib`: a matrix of the original variables in dataset `N` as rows and the PLS components `M` as columns.

`datasets`: a character vector that stores the names of the datasets that were specified for the function.

Methods

`summary`: outputs summary statistics for the contributions of dataset `N` to components from all datasets in the case of `getAllWeights` or dataset `M` in the case of `getCompositeWeights`.

`image`: outputs a heatmap of the transposed `contrib` matrix.

`heat`: outputs a clustered heatmap of the `contrib` matrix.

Author(s)

Kevin R. Coombes <krc@silicovore.com>, Kyoko Yamaguchi <kyoko.yamaguchi@osumc.edu>

Examples

```
fls <- try(loadESCAdata())
if (inherits(fls, "try-error")) {
  stop("Unable to load data from remote server.")
}
# restrict data set size
MO <- prepareMultiOmics(assemble[c("ClinicalBin", "ClinicalCont", "RPPA")], Outcome)

splitVec <- rbinom(nrow(Outcome), 1, 0.6)
trainD <- MO[, splitVec == 1]
testD <- MO[, splitVec == 0]

firstPass <- fitCoxModels(trainD, "Days", "vital_status", "dead")
pl <- plasma(object = trainD, multi = firstPass)

getCompositeWeights(object = pl, N = "ClinicalBin", M = "RPPA")

cbin <- getAllWeights(object = pl, N = "ClinicalBin")
summary(cbin)
image(cbin)
heat(cbin, cexCol = 0.5)

cbin01 <- pickSignificant(object = cbin, alpha = 0.01)
image(cbin01)
heat(cbin01, cexCol = 0.5)

getTop(object = cbin01, N = 3)
```

MultiOmics-class *Class* "MultiOmics"

Description

The `prepareMultiOmics` function returns a new object of `MultiOmics` class for use in `fitCoxModel`.

Usage

```
prepareMultiOmics(datalist, outcome)
## S4 method for signature 'MultiOmics'
summary(object, ...)
## S4 method for signature 'MultiOmics,missing'
plot(x, y, ...)
```

Arguments

datalist	a list of dataframes formatted to have variables as rows (dimension D) and samples as columns (dimension N).
outcome	a dataframe of clinical outcomes formatted to have sample names as row indexes and variable names as column indexes
object	An object of the MultiOmics class.
x	An object of the MultiOmics class.
y	Nothing; ignored.
...	Extra graphical or other parameters.

Value

The prepareMultiOmics function returns a new object of the MultiOmics class.

Objects from the Class

Objects should be defined using the prepareMultiOmics constructor. In the simplest case, you enter two objects: a list of dataframes and a dataframe of clinical outcomes.

Slots

data: A list of dataframes with variables as rows or varying length and samples as columns of uniform length N, where N is the maximum value of non-missing samples in any given dataset. Note that NAs have been added to "pad" to make the column length uniform across data types.

outcome: A dataframe of clinical outcomes with variables as columns and samples as rows.

Methods

plot: Produces a visual representation of the dimensionalities of each dataframe in datalist. D corresponds to the number of variables in each omics dataframe, and N corresponds to samples (or members) whose variable is not entirely missing. Gray areas correspond to missing samples.

summary: Produces summary tables corresponding to datasets and outcomes.

Author(s)

Kevin R. Coombes <krc@silicovore.com>, Kyoko Yamaguchi <kyoko.yamaguchi@osumc.edu>

Examples

```
fls <- try(loadESCAdata())
if (inherits(fls, "try-error")) {
  stop("Unable to load data from remote server.")
}
MO <- prepareMultiOmics(datalist = assemble, outcome = Outcome)

plot(MO)
summary(MO)
```

MultiplePLSCoxModels-class

Class "MultiplePLSCoxModels"

Description

The MultiplePLSCoxModels object class ... The validMultipleCoxModels function checks if each data set contains the same set of samples. The fitCoxModels function fits many plsRcox-models and returns an S4 object of class MultiplePLSCoxModels. The getSizes function returns a matrix with the list of dataframes of the MultiOmics object as rownames and columns with NT, cNT, and p-values.

Usage

```
fitCoxModels(multi, timevar, eventvar, eventvalue, verbose)
## S4 method for signature 'MultiplePLSCoxModels'
summary(object, ...)
## S4 method for signature 'MultiplePLSCoxModels,missing'
plot(x, y, col = c("blue", "red"),
      lwd = 2, xlab = "", ylab = "Fraction Surviving",
       mark.time = TRUE, legloc = "topright", ...)
## S4 method for signature 'MultiplePLSCoxModels'
predict(object, newdata, type = c("components", "risk",
                 "split"), ...)
```

Arguments

multi	an object of class MultiOmics for fitting the model.
timevar	a column in the MultiOmics object in the outcome dataframe containing the time-to-event.
eventvar	a column in the MultiOmics object in the outcome dataframe containing the event.
eventvalue	a character string specifying the value of the event in eventvar.
verbose	logical; should the function report progress.
object	an object of class MultiplePLSCoxModels for outputting the summary.
x	an object of class MultiplePLSCoxModels for plotting the Kaplan-Meier curves.
y	An ignored argrument for the plot method.
col	A vector of color specifications. Default is c("blue", "red").

lwd	A vector specifying the line width. Default is "2".
xlab	A character string to label the x-axis. Default is "".
ylab	A character string to label the y-axis. Default is "Fraction Surviving".
mark.time	A logical value; should tickmarks indicate censored data? Default is TRUE.
legloc	A character string indicating where to put the legend. Default is "topright".
...	Other graphical parameters.
newdata	A MultiOmics object with the same structure as the training data.
type	An enumerated character value.

Value

The fitCoxModels function retuns a newly constructed object of the MultiplePLSCoxModels class. The plot method invisibly returns the object on which it was invoked. The summary method returns no value. The predict method returns a list of prediction results, each of which comes from the predict method for the SingleModel-class.

Slots

models: A list of SingleModel objects, one for each assay.

timevar: A character matching the name of the column containing the time-to-event.

eventvar: A character matching the name of the column containing the event.

eventvalue: A character specifying the event in eventvar.

Methods

plot: Plots Kaplan-Meier curves for each omics dataset split into Low Risk and High Risk groups.

summary: Returns a description of the MultiplePLSCoxModels object and the names of the omics datasets used to build the model.

predict: returns a list of numeric vectors of predicted risk per data type.

Author(s)

Kevin R. Coombes <krc@silicovore.com>, Kyoko Yamaguchi <kyoko.yamaguchi@osumc.edu>

See Also

fitSingleModel

Examples

```
fls <- try(loadESCAdata())
if (inherits(fls, "try-error")) {
  stop("Unable to load data from remote server.")
}
# restrict data set size
MO <- prepareMultiOmics(assemble[c("ClinicalBin", "ClinicalCont", "RPPA")], Outcome)

splitVec <- rbinom(nrow(Outcome), 1, 0.6)
trainD <- MO[, splitVec == 1]
testD <- MO[, splitVec == 0]
```

```
firstPass <- fitCoxModels(trainD, "Days", "vital_status", "dead")

summary(firstPass)
plot(firstPass)
getSizes(firstPass)
pre1 <- predict(firstPass, testD)
```

---

plasma-class *Class* "plasma"

Description

The plasma object class is returned after running the plasma function. The plasma function uses the PLSRCox components from one dataset as the predictor variables and the PLSRCox components of another dataset as the response variables to fit a partial least squares regression (plsr) model. Then, we take the mean of the predictions to create a final matrix of samples versus components.

The matrix of components described earlier is then used to fit a Cox Proportional Hazards (coxph) model with AIC stepwise variable selection to return a final object of class plasma which includes a coxph model with a reduced number of predictors.

Usage

```
plasma(object, multi)
## S4 method for signature 'plasma,missing'
plot(x, y, ...)
## S4 method for signature 'plasma'
predict(object, newdata = NULL, type = c("components", "risk",
                 "split"), ...)
```

Arguments

multi	an object of the MultiplePLSCoxModels class.
object	an object of the plasma class.
x	an object of class plasma for plotting the Kaplan-Meier curves.
y	An ignored argrument for the plot method.
newdata	A MultiOmics object with the same structure as the training data.
type	An enumerated character value.
...	Additional graphical parameters.

Value

The plasma function returns a newly constructed object of the plasma class. The plot method invisibly returns the object on which it was invoked. The predict method returns an object of the plasmaPredictions class.

Objects from the Class

Objects should be defined using the plasma function.

Slots

`traindata`: An object of class `MultiOmics` used for training the model.

`compModels`: A list containing objects in the form of `plsr`.

`fullModel`: A coxph object with variables (components) selected via AIC stepwise selection.

Methods

`plot`: Plots a Kaplan-Meier curve of the final `coxph` model that has been categorized into "low risk" and "high risk" based whether it is higher or lower, respectively, than the median value of risk.

`predict`: creates an object of class `plasmaPredictions`.

Author(s)

Kevin R. Coombes <krc@silicovore.com>, Kyoko Yamaguchi <kyoko.yamaguchi@osumc.edu>

See Also

plasmaPredictions, plsr

Examples

```
fls <- try(loadESCAdata())
if (inherits(fls, "try-error")) {
  stop("Unable to load data from remote server.")
}
# restrict data set size
MO <- prepareMultiOmics(assemble[c("ClinicalBin", "ClinicalCont", "RPPA")], Outcome)

splitVec <- rbinom(nrow(Outcome), 1, 0.6)
trainD <- MO[, splitVec == 1]
testD <- MO[, splitVec == 0]

firstPass <- fitCoxModels(trainD, "Days", "vital_status", "dead")
pl <- plasma(object = trainD, multi = firstPass)

plot(pl, legloc = "topright", main = "Training Data")
```

---

plasmaPredictions-class

Class "plasmaPredictions"

Description

The `plasmaPredictions` object class is returned when running the `predict` method on an object of class `plasma`.

Usage

```
## S4 method for signature 'plasmaPredictions,missing'
plot(x, y, col = c("blue", "red"),
                 lwd = 2, xlab = "", ylab = "Fraction Surviving",
                 mark.time = TRUE, legloc = "topright", ...)
```

Arguments

x	An object of the plasmaPredictions class for plotting the Kaplan-Meier curves.
y	An ignored argument for the plot method.
col	A vector of color specifications. Default is c("blue", "red").
lwd	A vactor specifying the line width. Default is "2".
xlab	A character string to label the x-axis. Default is "".
ylab	A character string to label the y-axis. Default is "Fraction Surviving".
mark.time	A logical value; should tickmarks indicate censored data? Default is TRUE.
legloc	A character string indicating where to put the legend. Default is "topright".
...	Other graphical parameters.

Value

The predict method on an object of the plasma class returns an object of the plasmaPredictions class. The plot method invisibly returns the value on which it was invoked.

Objects from the Class

Users shold not create objects of this class directly. They will be automatically created when you apply the predict method to a fully worked out plasma model.

Slots

meanPredictions: A matrix with samples as rows and factors as columns that is a result of taking the mean of the PLS component predictions from each dataset.

riskDF: Object of type data.frame containing the original outcome dataframe and additional columns for "Risk", and "Split", corresponding to the risk of the event calculated by the model, and patient assignment to low versus high-risk groups, respectively.

riskModel: Object of type coxph that uses predicted Risk (continuous) as the predictor variable and survival as the response variable. See documentation for link{coxph}.

splitModel: Object of type coxph that uses predicted Split (predicted Risk categorized into "high" and "low" risk by the median predicted Risk) as the predictor variable and survival as the response variable. See documentation for link{coxph}.

SF: Object of type survfit which is used by the plot method to plot Kaplan-Meier curves grouped by predicted Split. See documentation for link{survfit}.

Methods

plot: Produces Kaplan-Meier curves for the low risk and high risk groups.

Note

An object of plasmaPredictions class contains many models that are similar to an object of MultiplePLSCoxModels class.

Author(s)

Kevin R. Coombes <krc@silicovore.com>, Kyoko Yamaguchi <kyoko.yamaguchi@osumc.edu>

See Also

plasma

Examples

```
fls <- try(loadESCAdata())
if (inherits(fls, "try-error")) {
  stop("Unable to load data from remote server.")
}
# restrict data set size
MO <- prepareMultiOmics(assemble[c("ClinicalBin", "ClinicalCont", "RPPA")], Outcome)

splitVec <- rbinom(nrow(Outcome), 1, 0.6)
trainD <- MO[, splitVec == 1]
testD <- MO[, splitVec == 0]

firstPass <- fitCoxModels(trainD, "Days", "vital_status", "dead")
pl <- plasma(object = trainD, multi = firstPass)

testpred <- predict(pl, testD)
plot(testpred, main = "Testing", xlab = "Time (Days)")
```

SingleModel-class *Class* "SingleModel"

Description

The `fitSingleModel` function takes in an object of `MultiOmics` class and returns a new object of `SingleModel` class.

Usage

```
fitSingleModel(multi, N, timevar, eventvar, eventvalue)
## S4 method for signature 'SingleModel'
summary(object, ...)
## S4 method for signature 'SingleModel,missing'
plot(x, y, col = c("blue", "red"),
      lwd = 2, xlab = "", ylab = "Fraction Surviving",
       mark.time = TRUE, legloc = "topright", ...)
## S4 method for signature 'SingleModel'
predict(object, newdata, type = c("components", "risk",
                 "split"), ...)
```

Arguments

`multi`	an object of class `MultiOmics` for fitting the model.
`N`	A character string identifying the data set being modeled.
`timevar`	a column in the `MultiOmics` object in the `outcome` dataframe containing the time-to-event.
`eventvar`	a column in the `MultiOmics` object in the `outcome` dataframe containing the event.

eventvalue	a character string specifying the value of the event.
x	an object of class plsRcoxmodel for plotting the Kaplan-Meier curves.
y	An ignored argrument for the plot method.
col	A vector of color specifications.
lwd	A vactor specifying the line width.
xlab	A character string to label the x-axis.
ylab	A character string to label the y-axis.
mark.time	A logical value; should tickmarks indicate censored data?
legloc	A character string indicating where to put the legend.
object	an object of class SingleModel.
newdata	A MultiOmics object with the same structure as the training data.
type	An enumerated character valuee.
...	other parameters used in graphing or prediction.

Value

The fitSingleModel function returns a newly constructed object of the SingleModel class. The plot method invisibly returns the value on which it was invoked. The summary method returns an object summarizing the final model produced by PLS R cox regression. The predict method returns either a vector or matrix depending on the type of predictions requested.

Slots

plsmod: Object of class plsRcoxmodel containing the fitted model.

Xout: Object of type data.frame containing the original outcome dataframe and additional columns for "Risk", and "Split", corresponding to the risk of the event calculated by the model, and patient assignment to low versus high-risk groups, respectively.

dsname: A character vector of length one; the name of the data set being modeled from a MultiOmics object.

SF: Object of type survfit which is used by the plot method to plot Kaplan-Meier curves grouped by predicted Split. See documentation for link{survfit}.

riskModel: Object of type coxph that uses predicted Risk (continuous) as the predictor variable and survival as the response variable. See documentation for link{coxph}.

splitModel: Object of type coxph that uses predicted Split (predicted Risk categorized into "high" and "low" risk by the median predicted Risk) as the predictor variable and survival as the response variable. See documentation for link{coxph}.

Methods

plot: Plots Kaplan-Meier curves for each omics dataset split into Low Risk and High Risk groups.

summary: Returns a description of the MultiplePLSCoxModels object and the names of the omics datasets used to build the model.

predict: a numeric vector containing the predicted risk values.

Author(s)

Kevin R. Coombes <krc@silicovore.com>, Kyoko Yamaguchi <kyoko.yamaguchi@osumc.edu>

See Also

getSizes

Examples

```
fls <- try(loadESCAdata())
if (inherits(fls, "try-error")) {
  stop("Unable to load data from remote server.")
}
MO <- prepareMultiOmics(assemble, Outcome)
MO <- MO[c("ClinicalBin", "ClinicalCont", "RPPA"),]
set.seed(98765)
splitVec <- rbinom(nrow(Outcome), 1, 0.6)
trainD <- MO[, splitVec == 1]
testD <- MO[, splitVec == 0]

zerothPass <- fitSingleModel(trainD, N = "RPPA",
                             timevar = "Days", eventvar = "vital_status",
                             eventvalue = "dead")
summary(zerothPass)
plot(zerothPass)
pre0 <- predict(zerothPass, testD)
```

TCGA-ESCA *Esophageal carcinoma (ESCA) data from The Cancer Genome Atlas (TCGA)*

Description

The TCGA-ESCA dataset contains the objects assemble, Outcome, and m450info for building the MultiOmics object. Because its size exceeds the CRAN limits, the data is stored on a remote server and must be loaded using the function loadESCAdata.

Usage

```
loadESCAdata()
```

Format

The "TCGA-ESCA" dataset contains the following:

assemble A list of 7 different omics dataframes with varying numbers of features as rows (D) and varying number of patients as columns (N). Note that some of these omics dataframes had been manipulated to contain NAs, where these may be complete on the GDC Dat Portal from which these data originally came. This was done to illustrate the capability of the plasma package on working with missing data.

1. ClinicalBina dataframe (53x185) of clinical binary values.
2. ClinicalConta dataframe (6x185) of clinical continuous values.
3. MAFa dataframe (566x184) of minor allele frequencies (MAF) that have been converted to binary based on whether they had a MAF greater than 0.03 (1) or not (0).

4. Meth450a dataframe (1454x185) of continuous beta values from the Illumina Infinium HumanMethylation450 arrays. The features in this dataframe have been filtered on mean greater than 0.15 and a standard deviation greater than 0.3.
5. miRSeqa dataframe (926x166) of continuous counts values from microRNA (miRNA) sequencing. The features in this dataframe have been filtered on a standard deviation of 0.05.
6. mRNASeqa dataframe (2520x157) of continuous counts values from mRNA sequencing data. The features in this dataframe have been filtered on a mean greater than 4 and a standard deviation greater than 0.7.
7. RPPAa dataframe (192x126) of continuous protein expression values from reverse phase protein array (RPPA) assays.

Outcome a dataframe (185x5) containing the survival outcomes for the patients in assemble.

m450info a dataframe (1454x3) containing gene symbol, chromosome number, and genomic coordinate IDs corresponding to the features (or "probes") in Meth450.

Author(s)

Kevin R. Coombes <krc@silicovore.com>, Kyoko Yamaguchi <kyoko.yamaguchi@osumc.edu>

Source

https://portal.gdc.cancer.gov/projects/TCGA-ESCA

Examples

```
fls <- try(loadESCAdata())
if (inherits(fls, "try-error")) {
  stop("Unable to load data from remote server.")
}
```

Index