Package 'VRPM'

July 13, 2015

Type Package

Title Visualizing risk prediction models

Version 1.0

Date 2015-05-08

Author Vanya Van Belle

Maintainer Vanya Van Belle <vanya.vanbelle@esat.kuleuven.be>

Description This is a package to visualize risk prediction models. For each predictor, a color bar represents the contribution to the linear predictor. A conversion from the linear predictor to the estimated risk or survival is also given. (Cumulative) contribution charts enable to visualize how the estimated risk for one particular observation is obtained by the model. Several options allow to choose different color maps, and to select the zero level of the contributions. The package is able to deal with glm, coxph and mfp objects.

Imports Hmisc, survival, fields

Suggests mfp

License GPL-2

R topics documented:

ccchart 1
cchart 3
colplot 5

Index **8**

`ccchart` *Cumulative contribution chart.*

Description

Display a graph explaining how the risk prediction for a new observation is obtained from the risk prediction model. All contributions are added in a cumulative way to end up with the linear predictor that is transformed into a risk estimate.

Usage

```
ccchart(x, obs, filename, zerolevel = "zero", risklabel, riskcutoff = 0.1,
  type, sorted = FALSE, time)
```

Arguments

x	glm, coxph or mfp object.
obs	A data.frame containing the predictor values of the observation for which the chart should be made.
filename	The name of the resulting pdf file, if the result is to be saved.
zerolevel	The value of the contributions that should be put to zero. If "zero", the contributions are represented as they are. If "min", for each predictor or set of predictors contributing to an interaction, the minimal observed value of the contribution in the training data is substracted from the contribution to ensure that the contribution is always positive. If "median" or "mean", the median or mean value is substracted from the contributions, respectively (default="zero"). See below for more details.
risklabel	A character string representing the label for the represented risk.
riskcutoff	A value between 0 and 1, indicating the change in color for the represented score and risk bars (default: 0.1). A risk lower than the riskcutoff is visualized in green, a risk higher than the riskcutoff is visualized in red.
type	A string specifying the type of plot: "logistic" for logistic regression models, and "survival" for cox proportional hazard models. This should not be provided by the user when using glm, coxph or mfp objects.
sorted	logical. If TRUE the contributions are sorted in increasing order (default=FALSE).
time	The time at which the estimated survival should be calculated. As default, the estimated survival at median survival time is reported. If the median survival time can not be calculated, the estimated survival at the latest event time is reported. For objects that are not a member of the coxph class, this is redundant.

Details

The cumulative contribution chart constitutes from different bars, representing the contribution of the predictors to the score (translated linear predictor) in a cumulative way. Depending on the value of zerolevel, the visualized contributions are slightly different. If zerolevel="zero", the contribution for variable x^p is $\beta_p x^p$, with β_p the model coefficient corresponding to this predictor. If zerolevel is "min", "median" or "mean", a value equal to the minimum, median and mean of the contribution $\beta_p x^p$ in the training data, respectively, is substracted from the contribution. At the bottom of the chart, the score of this observation is reported next to the maximal score observed in the training data. The risk corresponding to both of these scores are represented as well. Risk predictions above the riskcutoff result in red bars, risk predictions below this cutoff result in green bars.

Note

This graph can not (yet) be used for multinomial logistic regression and cox proportional hazard regression including strata.

For coxph models, it is necessary to include model=TRUE in the model fit.

Author(s)

Vanya Van Belle

References

Van Belle V., Van Calster B., *Visualizing Risk Prediction Models*, PLoS One, 2015. In Press.

See Also

colplot, cchart

Examples

```
#### logistic regression
mydata <- read.csv("http://www.ats.ucla.edu/stat/data/binary.csv")
mydata$rank <- factor(mydata$rank)
fit <- glm(admit ~ gre + gpa + rank, data = mydata, family = "binomial")
patient1=data.frame(gre=386,gpa=3.58,rank=3)
# The option zerolevel only changes the contributions of each input variable.
# The estimated risk remains the same.
ccchart(fit,obs=patient1)
ccchart(fit,obs=patient1,zerolevel="min")
ccchart(fit,obs=patient1,zerolevel="median")
# the contributions can be sorted by setting sorted to TRUE
ccchart(fit,obs=patient1,zerolevel="min",sorted=TRUE)

#### cox proportional hazard regression
library(mfp)
data(GBSG)
fit<-coxph(Surv(rfst, cens) ~ age+tumsize+posnodal+prm+esm+menostat+tumgrad, data = GBSG,
model=TRUE)
patient1=data.frame(age=52,tumsize=45,posnodal=23,prm=1025,esm=562,menostat=2,tumgrad=3)
patient2=data.frame(age=52,tumsize=25,posnodal=3,prm=100,esm=562,menostat=1,tumgrad=3)
# Depending on the risk cutoff, the score and risk are represented by means of green or
# red bars.
ccchart(fit,patient2)
ccchart(fit,patient2,riskcutoff=0.7)
# Represent the estimated survival at 500 days.
ccchart(fit,patient1,time=500)
# Change the risk label
ccchart(fit,patient1,,"median",time=500,risklabel="mylabel",sorted=TRUE)


#### cox proportional hazard regression in combination with mfp
library(mfp)
data(GBSG)
fit <- mfp(Surv(rfst, cens) ~ age+fp(tumsize)+fp(posnodal)+fp(prm)+fp(esm)+menostat+
tumgrad, family = cox, data = GBSG)
ccchart(fit,patient1,time=500)
```

---

cchart *Contribution chart.*

Description

Display a graph explaining how the risk prediction of a new observation is obtained from the model. All contributions to the linear predictor are visualized and summed to obtain the linear predictor. The latter is then transformed into a risk estimate.

Usage

```
cchart(x, obs, filename, zerolevel = "zero", risklabel, sorted = FALSE,
  time)
```

Arguments

x	glm, coxph or mfp object.
obs	A data.frame containing the predictor values of the observation for which the chart should be made.
filename	The name of the resulting pdf file, if the result is to be saved.
zerolevel	The value of the contributions that should be put to zero. If "zero", the contributions are represented as they are. If "min", for each predictor or set of predictors contributing to an interaction, the minimal observed value of the contribution in the training data is substracted from the contribution to ensure that the contribution is always positive. If "median" or "mean", the median or mean value is substracted from the contributions, respectively (default="zero"). See below for more details.
risklabel	A character string representing the label for the represented risk.
sorted	logical. If TRUE the contributions are sorted in increasing order (default=FALSE).
time	The time at which the estimated survival should be calculated. As default, the estimated survival at median survival time is reported. If the median survival time can not be calculated, the estimated survival at the latest event time is reported. For objects that are not a member of the coxph class, this is redundant.

Details

The contribution chart is a bar plot representing the contribution of each predictor or set of predictors to the score (translated linear predictor) by means of bars. Depending on the value of zerolevel, the visualized contributions are slightly different. If zerolevel="zero", the contribution for variable x^p is $\beta_p x^p$, with β_p the model coefficient corresponding to this predictor. If zerolevel is "min", "median" or "mean", a value equal to the minimum, median and mean of the contribution $\beta_p x^p$ in the training set, respectively, is substracted from the contribution. The sum of all the contributions is the score (translated linear predictor) which is transformed into the risk estimate. The range of all contributions in the training set are visualized by means of black horizontal lines.

Note

This graph can not (yet) be used for multinomial logistic regression and cox proportional hazard regression including strata.

For coxph models, it is necessary to include model=TRUE in the model fit.

Author(s)

Vanya Van Belle

References

Van Belle V., Van Calster B., *Visualizing Risk Prediction Models*, PLoS One, 2015. In Press.

See Also

colplot, ccchart

Examples

```
#### logistic regression
mydata <- read.csv("http://www.ats.ucla.edu/stat/data/binary.csv")
mydata$rank <- factor(mydata$rank)
fit <- glm(admit ~ gre + gpa + rank, data = mydata, family = "binomial")
patient1=data.frame(gre=386,gpa=3.58,rank=3)
# The option zerolevel only changes the contributions of each input variable.
# The estimated risk remains the same.
cchart(fit,obs=patient1)
cchart(fit,obs=patient1,zerolevel="min")
cchart(fit,obs=patient1,zerolevel="median")
# the contributions can be sorted by setting sorted to TRUE
cchart(fit,obs=patient1,zerolevel="min",sorted=TRUE)

#### cox proportional hazard regression
library(mfp)
data(GBSG)
fit<-coxph(Surv(rfst, cens) ~ age+tumsize+posnodal+prm+esm+menostat+tumgrad, data = GBSG,
model=TRUE)
patient1=data.frame(age=52,tumsize=45,posnodal=23,prm=1025,esm=562,menostat=2,tumgrad=3)
cchart(fit,patient1)
cchart(fit,patient1,time=500)
# Indicate the risk estimate at 500 days
cchart(fit,patient1,,"min",,,500)
# Change the risk label
cchart(fit,patient1,time=500,risklabel="mylabel")
# cox proportional hazard regression with interactions between the predictors
fit<-coxph(Surv(rfst, cens) ~ age*tumsize+posnodal*prm+esm+menostat*tumgrad,
data = GBSG, model=TRUE)
cchart(fit,patient1,,"median",,,500)

#### cox proportional hazard regression in combination with mfp
library(mfp)
data(GBSG)
fit <- mfp(Surv(rfst, cens) ~ age+fp(tumsize)+fp(posnodal)+fp(prm)+fp(esm)+menostat+
tumgrad, family = cox, data = GBSG)
cchart(fit,patient1,time=500)
```

`colplot` *Visualize a risk prediction model by means of colored bars.*

Description

Display a graph in which the contributions of each predictor or set of predictors is represented in a colored bar. The color indicates the value of the contribution.

Usage

```
colplot(x, filename, coloroptions = 2, zerolevel = "zero", risklabel,
  adverse, obs, q5, q95, time)
```

Arguments

x	glm, coxph or mfp object.
filename	The name of the resulting pdf file, if the result is to be saved.
coloroptions	If 1, the rainbow color map is used. If 2, a sequential color map is used. If 3, a diverging color map is used. If 4, a black-and-white color map is used. (default=2)
zerolevel	The value of the contributions that should be put to zero. If "zero", the contributions are represented as they are. If "min", for each predictor or set of predictors contributing to an interaction, the minimal observed value of the contribution in the training data is substracted from the contribution to ensure that the contribution is always positive. If "median" or "mean", the median or mean value is substracted from the contributions, respectively (default="zero"). See below for more details.
risklabel	A character string representing the label for the represented risk.
adverse	A logical indicating whether the score and risk range in the adverse direction (default=FALSE, i.e. high score corresponds to a high risk).
obs	A data.frame containing the predictor values of the observation that should be added to the plot.
q5	A data.frame containing the predictor values of the 5th percentiles of the predictors that should be added to the plot. This only impacts the plot for continous variables.
q95	A data.frame containing the predictor values of the 95th percentiles of the predictors that should be added to the plot. This only impacts the plot for continous variables.
time	The time at which the estimated survival should be calculated. As default, the estimated survival at median survival time is reported. If the median survival time can not be calculated, the estimated survival at the latest event time is reported. For objects that are not a member of the coxph class, this is redundant.

Details

The colorplot is a chart that visualizes the contribution of a predictor or a set of predictors by means of colored bars. Depending on the value of zerolevel, the visualized contributions are slightly different. If zerolevel="zero", the contribution for variable x^p is $\beta_p x^p$, with β_p the model coefficient corresponding to this predictor. If zerolevel is "min", "median" or "mean", a value equal to the minimum, median and mean of the contribution $\beta_p x^p$ in the training set, respectively, is substracted from the contribution. coloroptions enables to choose between diffent color maps. It is recommended to use a sequential color map when zerolevel="min" and a diverging color map when zerolevel is "median" or "mean". For the latter case, a white color will correspond to zero points. In the color bar converting the score to a risk, white will correspond to the median or mean observed risk in the training data, respectively.

Note

This graph can not (yet) be used for multinomial logistic regression and cox proportional hazard regression including strata.

For coxph models, it is necessary to include model=TRUE in the model fit.

Author(s)

Vanya Van Belle

References

Van Belle V., Van Calster B., *Visualizing Risk Prediction Models*, PLoS One, 2015. In Press.

See Also

cchart, ccchart

Examples

```
#### logistic regression
# default options
mydata <- read.csv("http://www.ats.ucla.edu/stat/data/binary.csv")
mydata$rank <- factor(mydata$rank)
fit <- glm(admit ~ gre + gpa + rank, data = mydata, family = "binomial")
colplot(fit)
# the minimum of each contribution is put to zero
colplot(fit,zerolevel="min")
# the median of each contribution is put to zero, a diverging colormap is used
colplot(fit,zerolevel="median",coloroptions=3)

#### cox proportional hazard regression
library(mfp)
data(GBSG)
fit<-coxph(Surv(rfst, cens) ~ age+tumsize+posnodal+prm+esm+menostat+tumgrad, data = GBSG,
model=TRUE)
colplot(fit)
# indicate the predicted risk (1-survival) at 500 days for a new patient
patient=data.frame(age=52,tumsize=45,posnodal=23,prm=1025,esm=562,menostat=2,tumgrad=3)
colplot(fit,obs=patient,time=500)
# cox proportional hazard regression with interactions between the predictors
fit<-coxph(Surv(rfst, cens) ~ age*tumsize+posnodal*prm+esm+menostat*tumgrad,
data = GBSG, model=TRUE)
colplot(fit,zerolevel="median",coloroptions=3)

#### cox proportional hazard regression in combination with mfp
library(mfp)
data(GBSG)
fit <- mfp(Surv(rfst, cens) ~ age+fp(tumsize)+fp(posnodal)+fp(prm)+fp(esm)+menostat+
tumgrad, family = cox, data = GBSG)
colplot(fit,time=500)
# diverging colormap, median of contributions is put to zero, use "mylabel" as risk label,
#  survival time=350
colplot(fit,,3,"median","mylabel",,,,,350)
# add two reference observations
q5=data.frame(age=46,tumsize=20,posnodal=1,prm=7,esm=8,menostat=1,tumgrad=1)
q95=data.frame(age=61,tumsize=35,posnodal=7,prm=132,esm=114,menostat=2,tumgrad=3)
colplot(fit,,3,"median","mylabel",,,q5,q95,350)
```

Index

ccchart, 1, *4*, *7*
cchart, *3*, 3, *7*
colorplot.coxph *(*colplot*)*, 5
colorplot.default *(*colplot*)*, 5
colorplot.glm *(*colplot*)*, 5
colplot, *3*, *4*, 5