survXAI

Aleksandra Grudziąż

2018-06-27

Single prediction

In chosen time points we want to find out which variables support the "predicted" value of survival probability for new observation and which variables contradict it.

```
library(survxai)
```

Data and model

```
library(rms)
library(randomForestSRC)

data(pbc, package = "randomForestSRC")
pbc <- pbc[complete.cases(pbc),]

predict_times <- function(model, new_observation, times=1){
  prob <- rms::survest(model, new_observation, times=times)$surv
  return(prob)
}

cph_model <- cph(Surv(days/365, status)~., data=pbc[1:300,], surv=TRUE, x = TRUE, y=TRUE)
```

Explainer

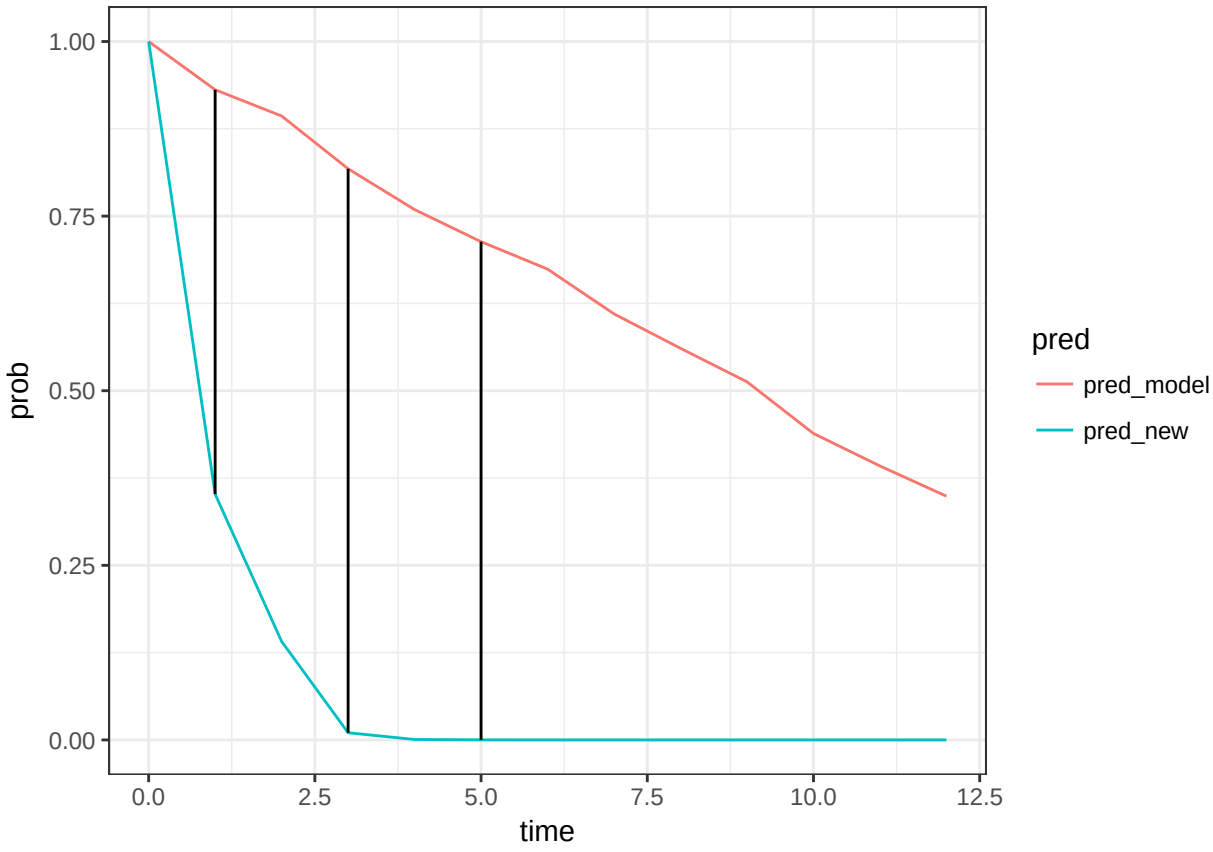
```
surve_cph <- surv_explain(model = cph_model,
                  data = pbc, time = pbc$days/365, status = pbc$status,
                  predict_function = predict_times)
```

Object with predictions

```
broken_list <- surv_breakdown(surve_cph, pbc[1,], times = c(1,3,5))
```

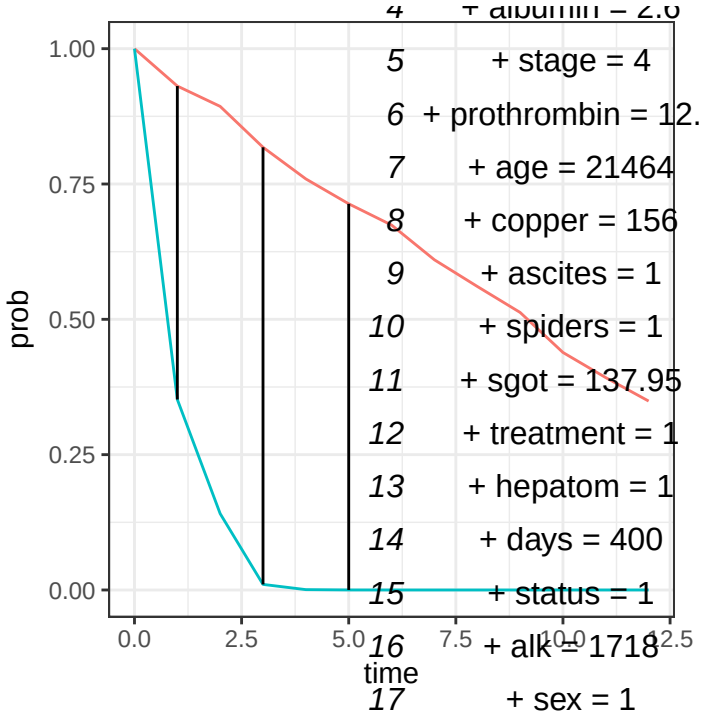
Plot survival curves

```
plot(broken_list)
```



Plot survival curves and contribution of each variables

```
plot(broken_list, table=T)
```



		+ albumin = 2.6		−1.087706e−01	−1.8
5		+ stage = 4		−8.940311e−02	−1.31
6	+ prothrombin = 12.2			−9.434601e−02	−8.53
7		+ age = 21464		−8.184380e−02	−4.83
8		+ copper = 156		−4.875775e−02	−2.15
9		+ ascites = 1		−2.866280e−02	−9.0(
10		+ spiders = 1		−2.493749e−02	−4.07
11		+ sgot = 137.95		−2.325737e−02	−2.76
12		+ treatment = 1		−2.235109e−02	−2.1(
13		+ hepatom = 1		−4.355044e−03	−3.0(
14		+ days = 400		0.000000e+00	0.0(
15		+ status = 1		0.000000e+00	0.0(
16		+ alk = 1718		9.549769e−05	3.34
17		+ sex = 1		1.512286e−02	7.01
18		+ trig = 172		1.644649e−02	1.11
19		+ chol = 261		1.871021e−02	1.39
20		+ platelet = 190		2.357602e−02	2.21

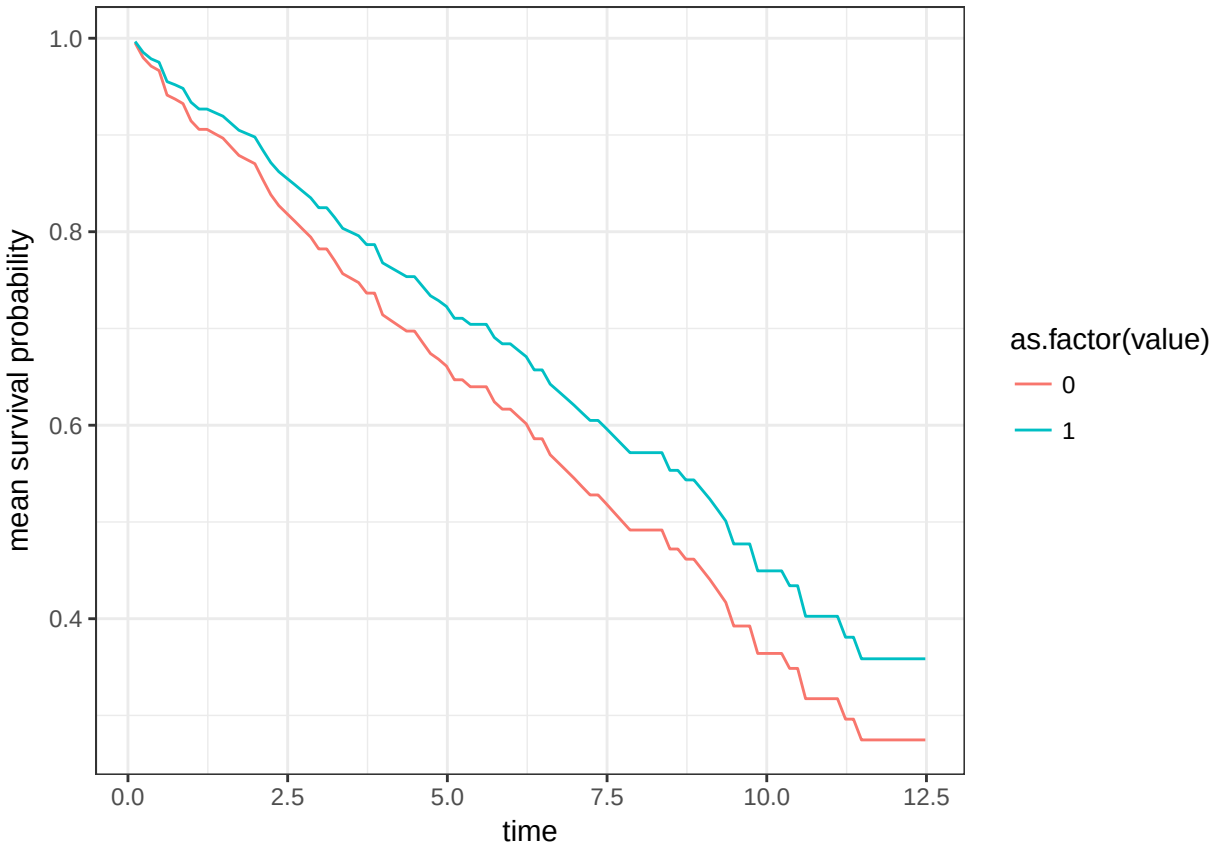
	time_1	time_3	time_5
del	0.930933893106637	0.817797963034403	0.713368750766265
ew	0.351945329731296	0.0102820890449102	4.31764456899885e−05

Variable response

We want to describe how each variable level impact the value of survival probability in time.

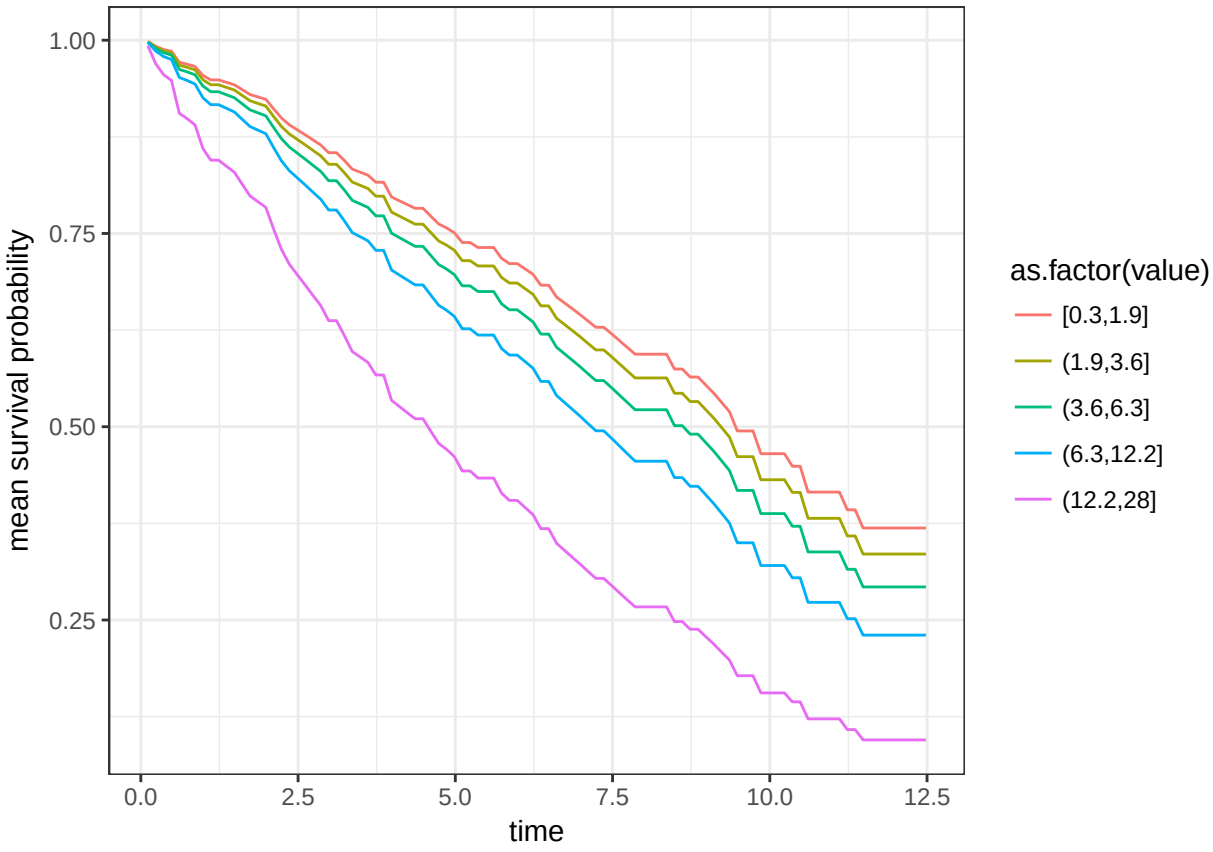
Variable response

```
svr_cph <- surv_variable_response(surve_cph, "sex")
plot(svr_cph)
```



For factor variable or numeric with few values we draw a survival curves for all values.

```
svr_cph_group <- surv_variable_response(surve_cph, "bili")
plot(svr_cph_group)
```



When we have a numerical variable with lots of values we plot not all values, because it might be difficult to read this plot, but we draw curves for quartiles. In example above we have variable `bili` has over 80 unique values, but we draw only 5 curves.