Predicting individual treatment effects with the `partykit` package

Heidi Seibold,
Achim Zeileis & Torsten Hothorn

useR! 2016

June 29, 2016

Overall Treatment Effect

```r
basemodel <- model(response ~ treatment, data)
```

Personalised medicine?

```
basemodel <- model(response ~ treatment, data)

model_m <- model(response ~ treatment, data,
                 subset = gender == "male")

model_f <- model(response ~ treatment, data,
                 subset = gender == "female")
```
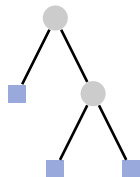
Personalised medicine?

```
basemodel <- model(response ~ treatment, data)


model_m <- model(response ~ treatment, data,
                 weights = as.numeric(gender == "male"))

model_f <- model(response ~ treatment, data,
                 weights = as.numeric(gender == "female"))
```

Model-based trees

Model-based trees find **subgroups of patients with similar treatment effect and/or expected response**.

Groups are formed based on patient characteristics.

```
model_sg1 <- model(response ~ treatment, data,
                   weights = as.numeric(subgroup == 1))
```
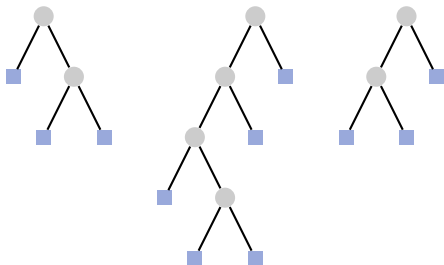
Personalised models

Weight patients higher who have a similar treatment
effect and/or expected response.

```
pmodel_i <- model(response ~ treatment, data,
                  weights = w_i)
```

Weights

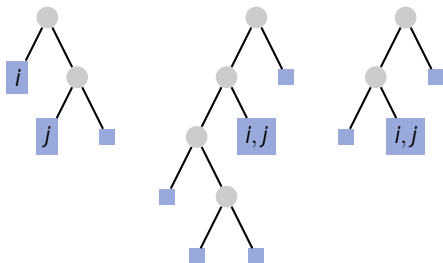Weights are computed using a model-based random forest.

Compute ensemble of trees, where each tree is based on:

- a subsample of the training data (per tree)
- a subsample of the eligible patient characteristics (per node)

Weights

Weights are computed using a model-based random forest.

How often are patients *i* and *j* assigned to the same terminal node / subgroup?
$\rightarrow$ 2 times $\Rightarrow w_{ij} = 2$

Personalised models

```
pmodel_i <- model(response ~ treatment, data,
                  weights = w_i)
```

$\rightarrow$ patient j enters $w_{ij} = 2$ times in pmodel$_i$.

Personalised models can be computed for
- in-sample patients
- new patients

PRO-ACT database

- Amyotrophic lateral sclerosis (ALS) patients
- Data of several clinical trials
  https://nctu.partners.org/ProACT/
- 3306 patients
- 18 patient characteristics

Base model:

```r
library("survival")
basemodel <- survreg(Surv(survival.time, cens) ~ Riluzole,
                     data = ALSsurvdata, dist = "weibull")
```

Computing the forest in R

```r
### computes model and returns score function
my.wb <- function(data, weights) {
  ## model
  mod <- survreg(Surv(survival.time, cens) ~ Riluzole,
                 data = data, weights = weights, subset = weights > 0,
                 dist = "weibull", init = c(6.7, 0))

  ## score function
  ef <- as.matrix(sandwich::estfun(mod))
  ret <- matrix(0, nrow = nrow(data), ncol = ncol(ef))
  ret[weights > 0,] <- ef
  ret
}


### forest
ALSforest <- cforest(survival.time + cens + Riluzole ~ age + gender + etc,
                     data = ALSsurvdata, ytrafo = my.wb,
                     ntree = 100, perturb = list(replace = FALSE))
```
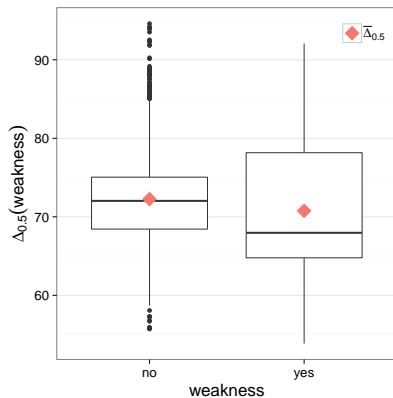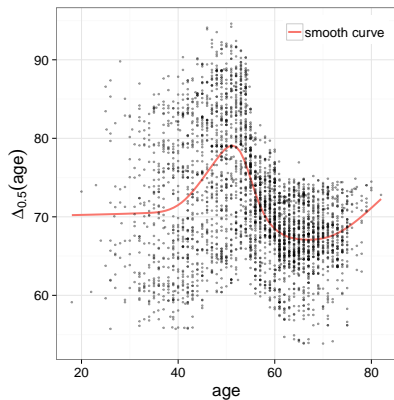
Computing personalised models in R

```r
### weights
weights <- predict(ALSforest, type = "weights", OOB = TRUE)
w_i <- weights[ , i]


### personalised model for patient i
pmodel_i <- survreg(Surv(survival.time, cens) ~ Riluzole,
                    data = ALSsurvdata, dist = "weibull",
                    weights = w_i)
```

Dependence plots

Is all this just overfitting?

Difference in log-likelihood between personalised models and base model:

$$71.47$$

but maximum only

$$0.96$$

on parametric bootstrap samples.

$$\text{log-likelihood} = \begin{cases} \sum\limits_{i=1}^{n} \ell((\text{response}, \text{treatment})_i, \text{pmodel}_i) \\ \sum\limits_{i=1}^{n} \ell((\text{response}, \text{treatment})_i, \text{basemodel}) \end{cases}$$

Papers on ArXiv:

H. Seibold, A. Zeileis, and T. Hothorn. Model-based Recursive Partitioning for Subgroup Analyses.

H. Seibold, A. Zeileis, and T. Hothorn. Individual Treatment Effect Prediction for ALS Patients.

Code:

partykit:
`https://cran.r-project.org/web/packages/partykit`

personalised models:
`https://github.com/HeidiSeibold/personalised_medicine`

Contact me:

`heidi.seibold@uzh.ch`