

iboost: Inference for Model-based Boosting

David Rügamer

Add-on package for the R package `mboost` to calculate p-values and confidence intervals for model parameters. Package currently supports models fitted with Gaussian family including linear, group and spline base-learners.

```
library("iboost")
library("mboost")
set.seed(0)

n <- 200
x1 <- rnorm(n)
x2 <- rnorm(n) + 0.25 * x1
x3 <- rnorm(n)
eta <- 3 * sin(x1) + x2^2
y <- scale(eta + rnorm(n), scale = FALSE)

spline1 <- bbs(x1, knots = 20, df = 4)
knots.x2 <- quantile(x2, c(0.25, 0.5, 0.75))
spline2 <- bbs(x2, knots = knots.x2, df = 4)
spline3 <- bbs(x3, knots = 20, df = 4)

data <- data.frame(y=y, x1=x1, x2=x2, x3=x3)

mod1 <- mboost(y ~ spline1 + spline2 + spline3,
              control=boost_control(mstop = 73), offset = 0,
              data = data)

# calculate p-values and intervals for model with
# fixed stopping iteration:
# this is done with only B = 100 samples for
# demonstrative purposes and should be increased
# for actual research questions
res <- iboost(mod1, method = "impsamp", B = 100)

# do the same with crossvalidation

fixFolds <- cv(weights = model.weights(mod1),
              type = "kfold", B = 10)
cvr <- cvrisk(mod1, folds = fixFolds, papply = lapply)
modf <- mod1[mstop(cvr)]

# define corresponding refit function
modFun <- function(y){

  mod <- mboost_fit(response = y,
```

```
      blg = blList,  
      offset = 0,  
      control = boost_control(mstop = 73))  
cvr <- cvrisk(mod, folds = fixFolds, papply = lapply)  
return(mod[mstop(cvr)])  
}  
  
# this will take a while  
(res <- iboost(modf, refit.mboost = modFun, method = "impsamp", B = 1000))
```

References

Rügamer, D., Greven, S. Selective inference after likelihood- or test-based model selection in linear models, *Statistics & Probability Letters* 140, 7-12 (2018). <https://doi.org/10.1016/j.spl.2018.04.010>.

Rügamer, D., Greven, S. Inference for L_2 -Boosting. *Stat Comput* 30, 279–289 (2020). <https://doi.org/10.1007/s11222-019-09882-0>.

Rügamer, D., Baumann, P., Greven, S. Selective inference for additive and linear mixed models, *Computational Statistics & Data Analysis* 167, 107350 (2022). <https://doi.org/10.1016/j.csda.2021.107350>.