Problem Set 5 – Boosting

2016-04-26

Boosting

- Analyse the Boston data set using boosting using trees. The dependent variable is again *medv*. Split the data to training and testing set. Use the testing set to analyze the prediction quality of your model.
- Hint: function *gbm()* from the package *gbm* or *mboost* from the package *mboost*
- Estimate again a tree-based model with boosting, but now with shrinkage parameter $\nu = 0.3$
- Experiment with the options and report the best tree-based boosting model.

```
library(MASS)
library(gbm)
```

```
## Loading required package: survival
```

```
## Loading required package: lattice
```

```
## Loading required package: splines
```

```
## Loading required package: parallel
```

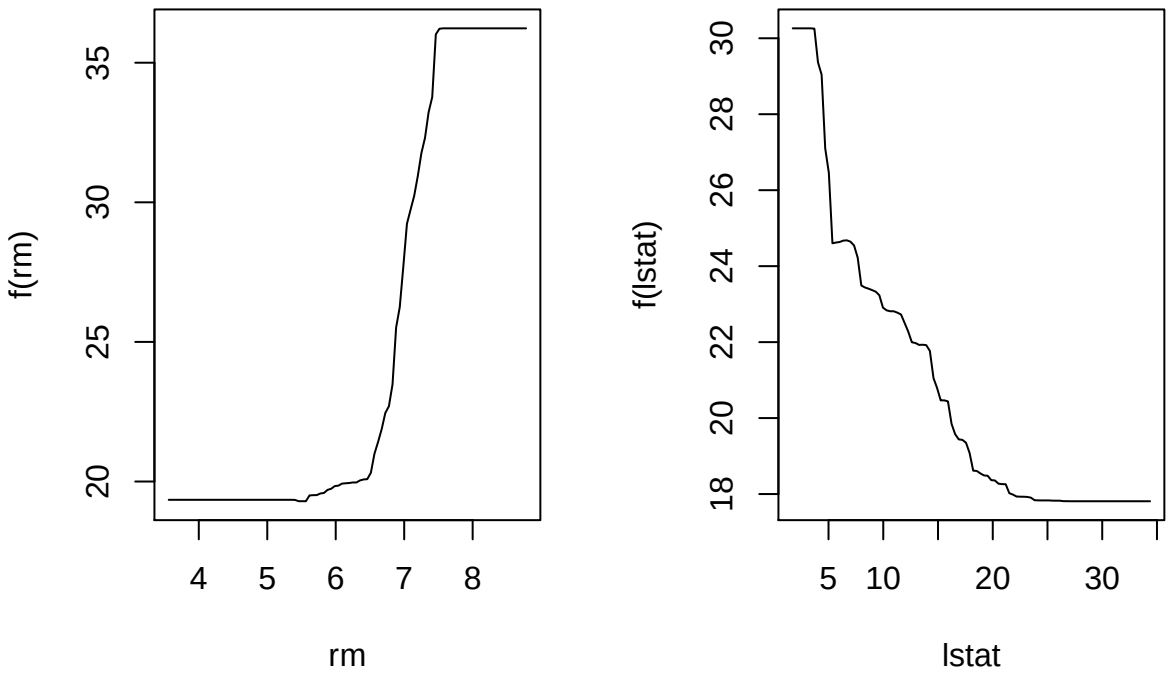
```
## Loaded gbm 2.1.1
```

```
library(mboost)
```

```
## Loading required package: stabs
```

```
## This is mboost 2.6-0. See 'package?mboost' and 'news(package  = "mboost")'
## for a complete list of changes.
```

```
set.seed(12345)
train = sample(1:nrow(Boston), floor(nrow(Boston)/2))
boost.boston = gbm(medv ~., data=Boston[train,], distribution="gaussian", n.trees=5000, interaction.dep
par(mfrow=c(1,2))
plot(boost.boston, i="rm")
plot(boost.boston, i="lstat")
```

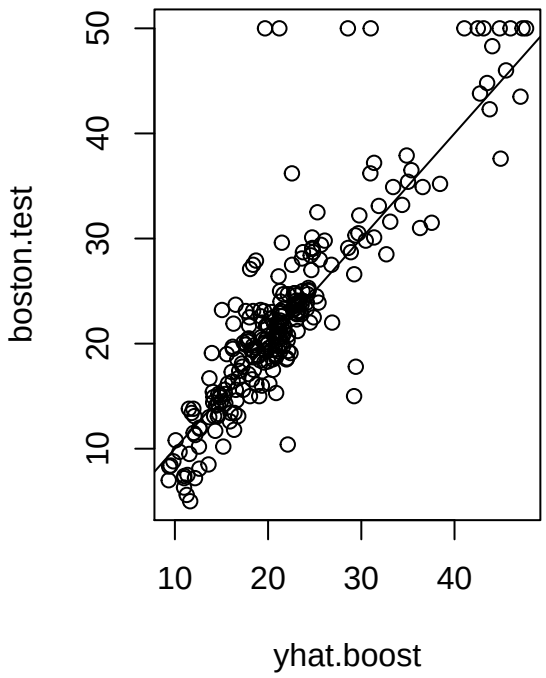



```
yhat.boost=predict(boost.boston, newdata=Boston[-train,], n.trees=5000)
boston.test=Boston[-train, "medv"]
mean((yhat.boost - boston.test)^2)
```

```
## [1] 21.49555
```

```
plot(yhat.boost, boston.test)
abline(0,1)
boost.boston = gbm(medv ~., data=Boston[train,], distribution="gaussian", n.trees=5000, interaction.dep
yhat.boost=predict(boost.boston, newdata=Boston[-train,], n.trees=5000)
mean((yhat.boost - boston.test)^2)
```

```
## [1] 19.39098
```


boston.test
yhat.boost
10
20
30
40
50
10
20
30
40