Chapter 08 Lab 4 - Boosting

p330

Use half the data for training and half for testing/validation.

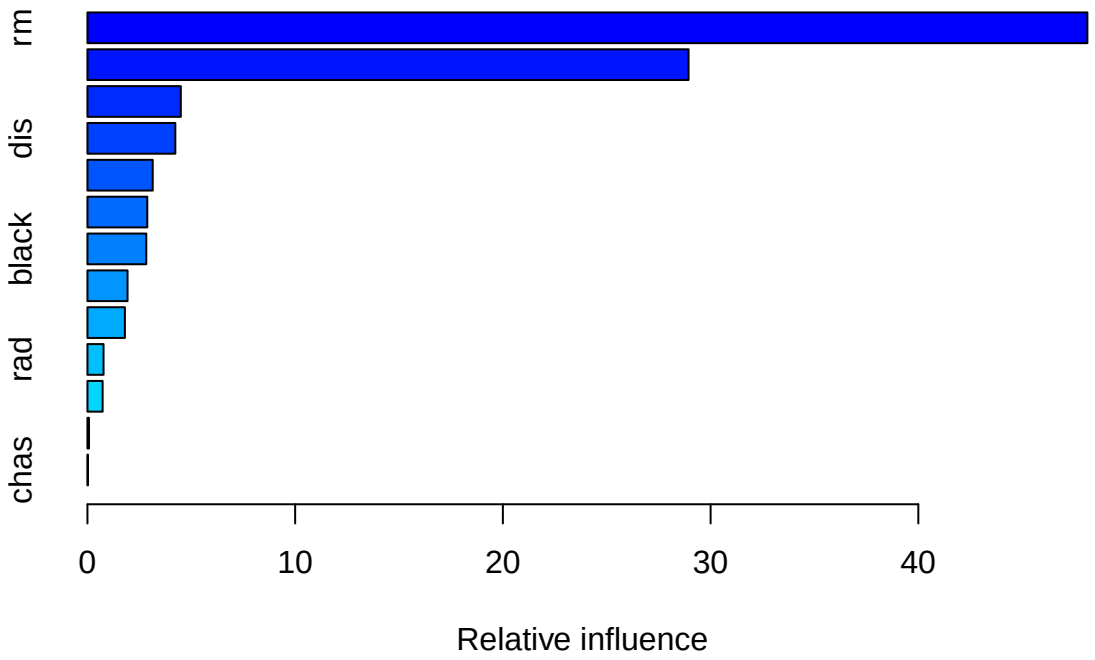
```
library(MASS)
library(gbm) # Generalized Boosted Regression Modeling
```

```
## Loaded gbm 2.1.8
```

```
set.seed(1)
train = sample(1:nrow(Boston), nrow(Boston)/2)

#Probably shouldn't do this because set train is different here? seed=1 so ok?
boston.test.y=Boston[-train, "medv"] # Need to import f other Lab section
```

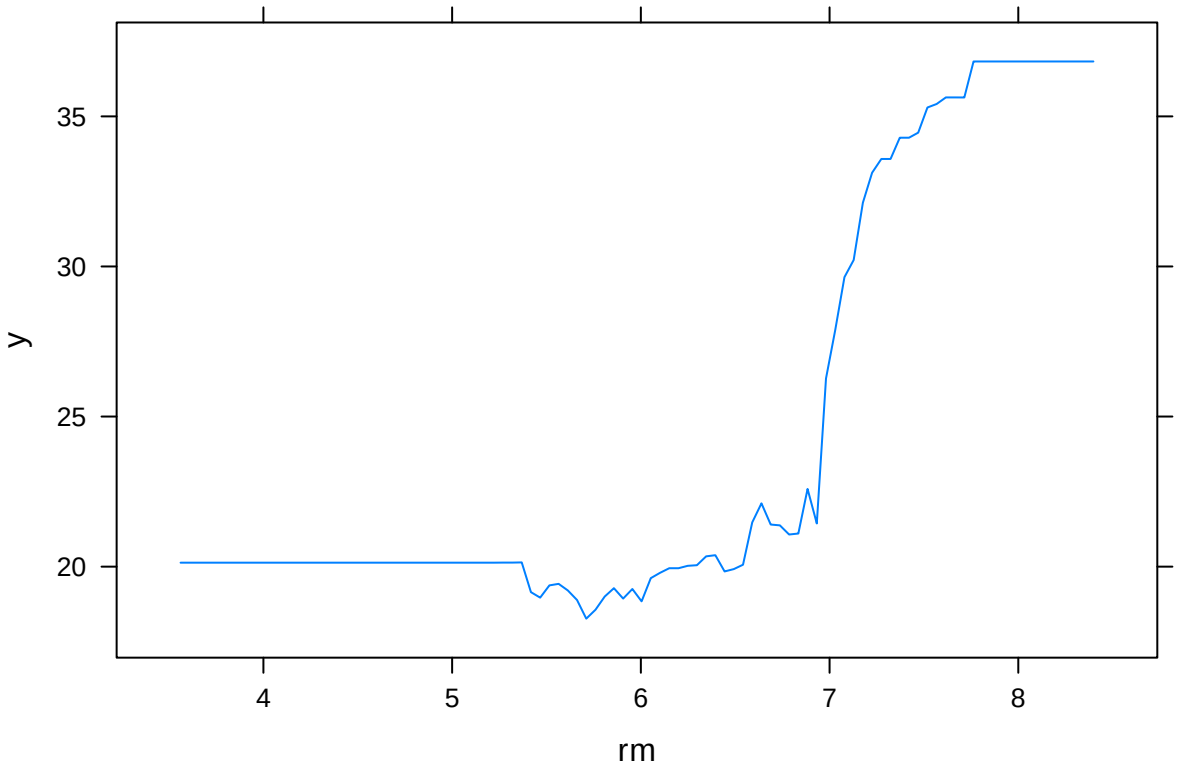
```
boost.boston=gbm(medv ~ ., data=Boston[train,], distribution="gaussian",
                 n.trees=5000, interaction.depth=4)
summary(boost.boston)
```



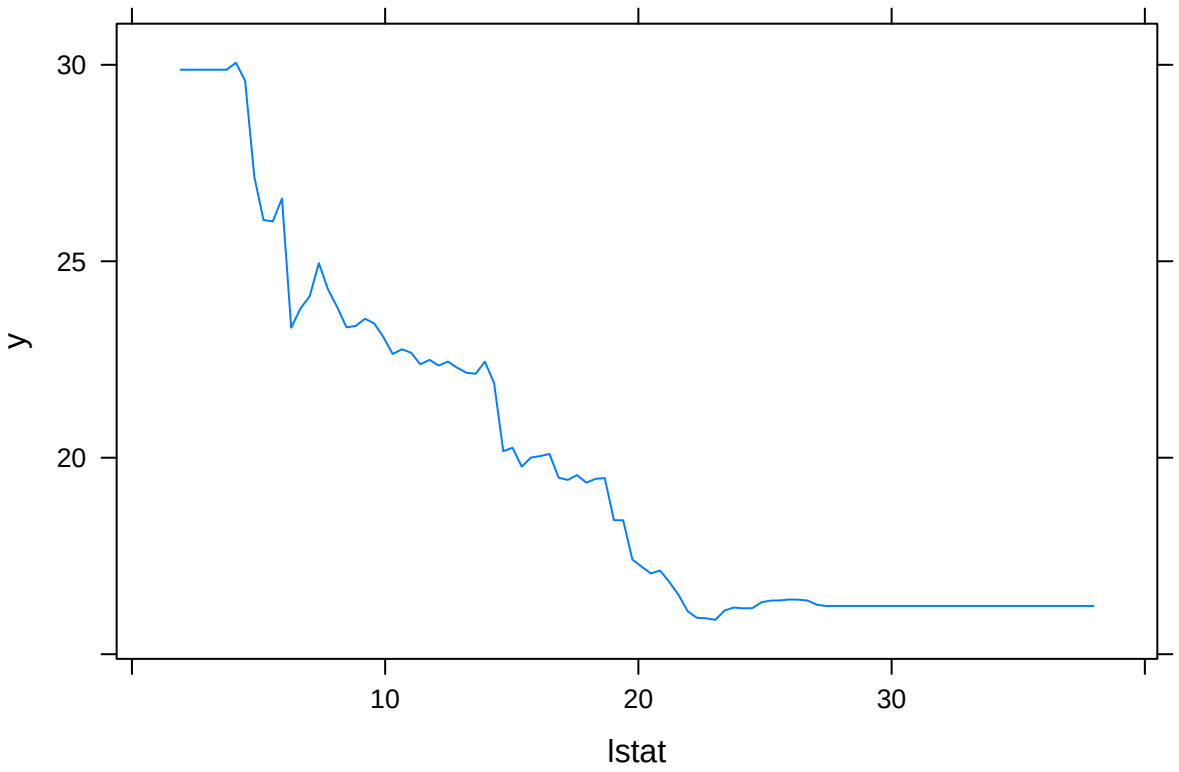

```
##             var     rel.inf
## rm           rm 48.13967682
## lstat     lstat 28.93851185
## crim       crim  4.49413146
## dis         dis  4.23182696
## age         age  3.14221169
## nox         nox  2.88094283
## black     black  2.83238772
## ptratio ptratio  1.93050932
```

```
## tax         tax  1.80427054
## rad         rad  0.77569461
## indus     indus  0.73110525
## zn           zn  0.07442923
## chas       chas  0.02430170
```

```
par(mfrow=c(1,2))
plot(boost.boston, i="rm") # average number of rooms per dwelling
```




```
plot(boost.boston, i="lstat") # lower status of the population (percent)
```



Probably shouldn't do this because set train is different here? seed=1 so ok?

```
#boston.test=Boston[-train, "medv"] # Need to import f other Lab section
```

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

```
## [1] 19.37033
```

In this case, using $\lambda = 0.2$ leads to a slightly lower test MSE than $\lambda = 0.00$

```
boost.boston=gbm(medv ~ .,data=Boston[train,], distribution="gaussian",
                 n.trees=5000, interaction.depth=4,
                 shrinkage=0.2, verbose=F)

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

```
## [1] 18.68911
```