

Boosting Worksheet

Abhirup Sen

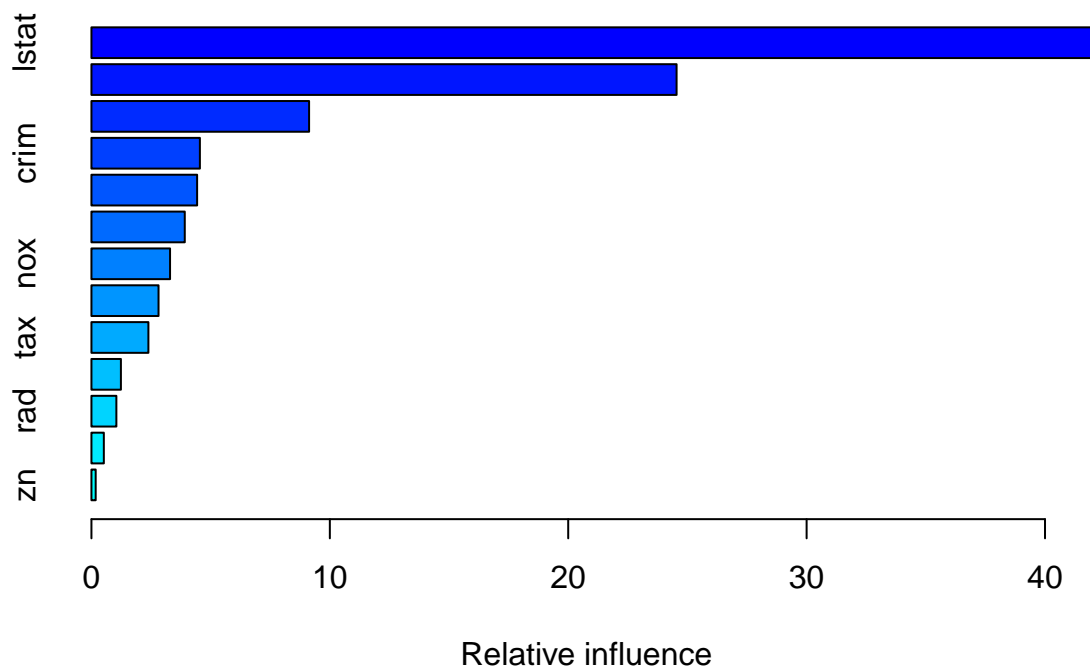
02/06/2021

```
#Boosting#
```

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

```
## Loaded gbm 2.1.8
```

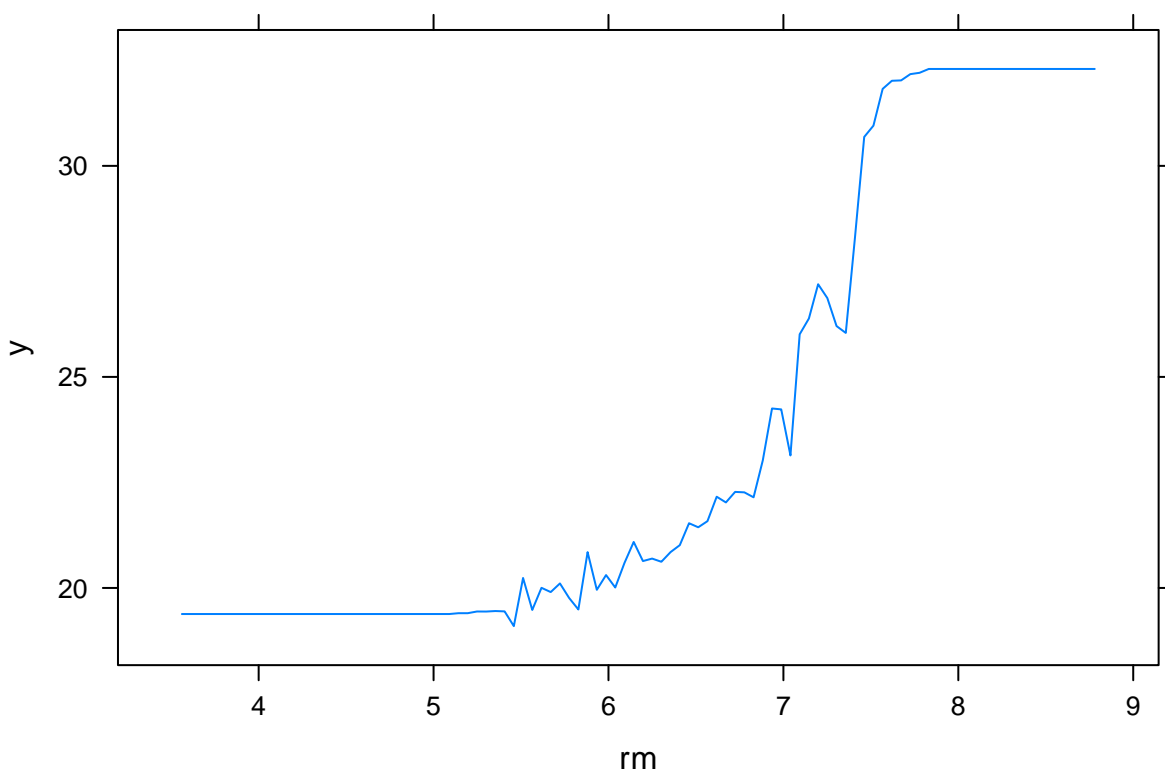
```
set.seed(10)
train = sample(1:nrow(Boston),nrow(Boston)/2)
boston.test = Boston[-train,"medv"]
boost.boston = gbm(medv~., data = Boston[train,], distribution = "gaussian",
                    n.trees = 5000, interaction.depth = 4)
summary(boost.boston)
```



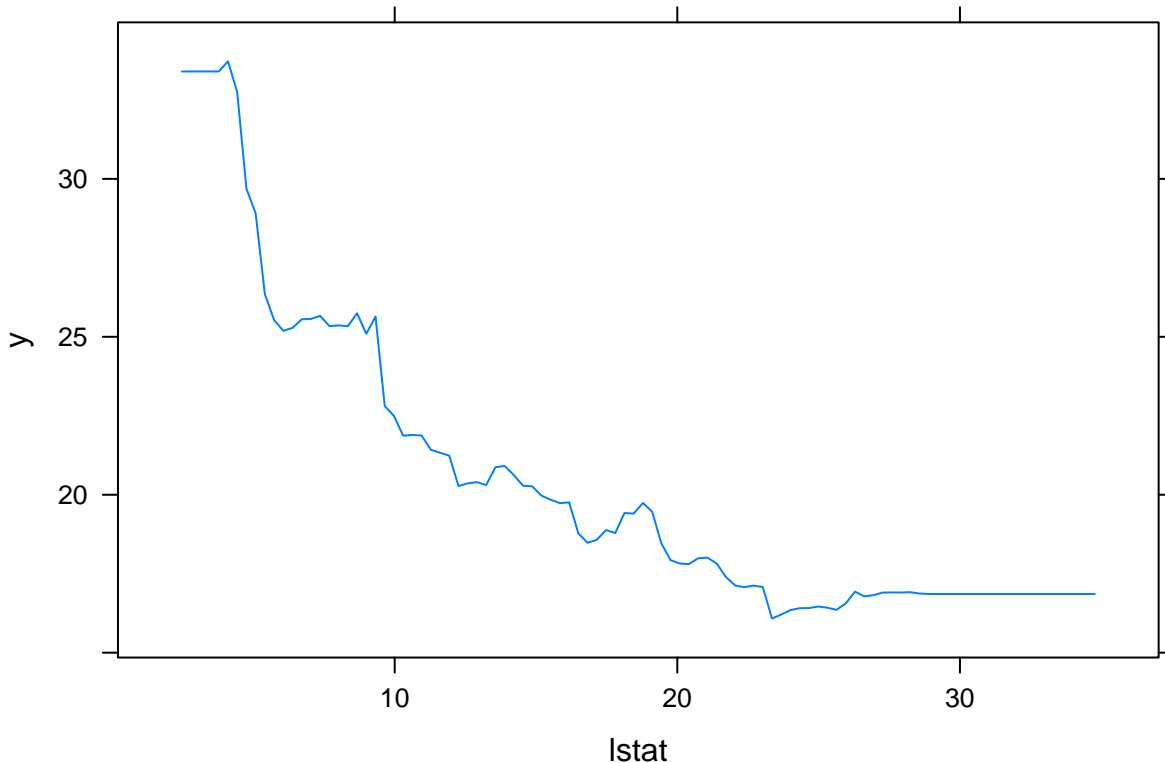
```
##          var    rel.inf
## lstat    lstat 41.9444150
## rm       rm   24.5459374
## dis      dis   9.1302324
## crim     crim  4.5495716
## age      age   4.4325354
## black    black 3.9151755
## nox      nox   3.2974569
## ptratio  ptratio 2.8146239
## tax      tax   2.3897066
## indus    indus 1.2364196
## rad      rad   1.0458959
## chas     chas  0.5198322
## zn       zn    0.1781975
```

rm and lstat are the most important variable

```
par(mfrow=c(1,2))
plot(boost.boston, i ="rm")
```



```
plot(boost.boston, i ="lstat")
```



Now we use the boosted models to predict medv on the test set

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

```
## [1] 10.41889
```

therefor the test MSE = 19.03 We introduce **shrinkage** parameters lambda

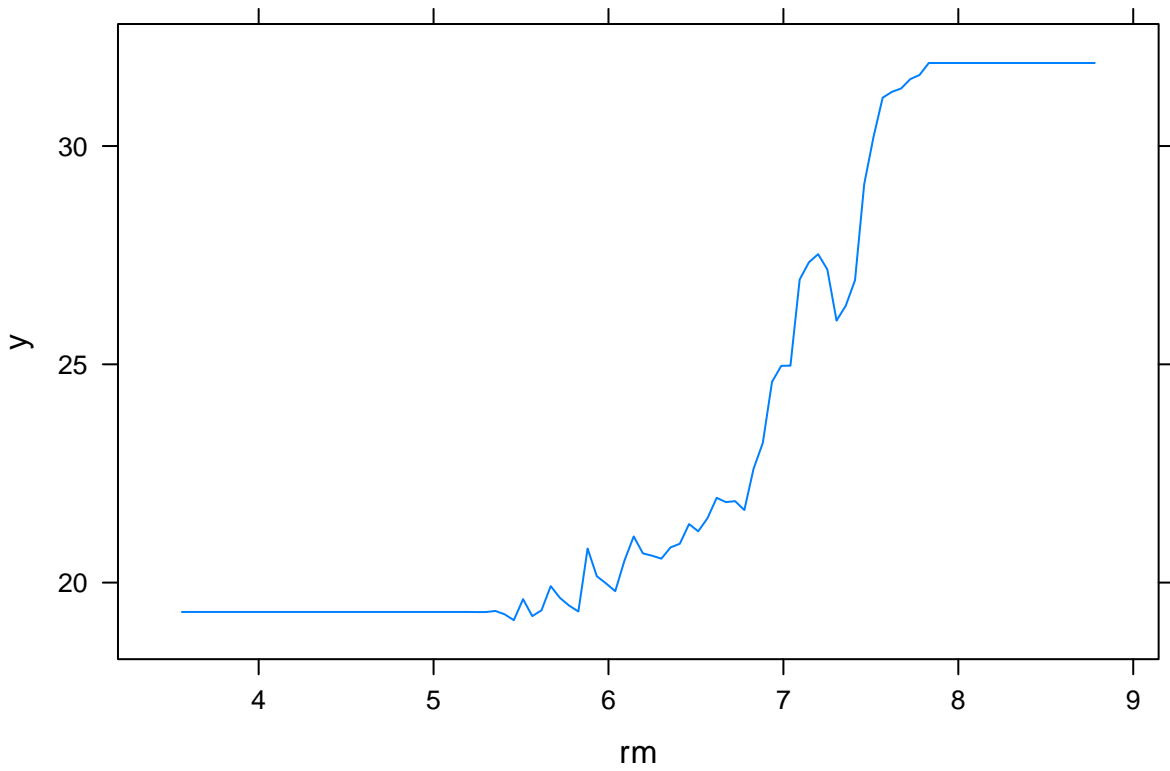
```
boost.boston = gbm(medv~., data = Boston[train,], distribution =
                    "gaussian",n.trees=5000, interaction.depth = 4,
                    shrinkage = 0.01,verbose = F)
yhat.boost = predict(boost.boston, newdata=Boston[-train,], n.trees=5000)
mean((yhat.boost-boston.test)^2)
```

```
## [1] 9.318381
```

```
boost.boston = gbm(medv~., data = Boston[train,], distribution =
                    "gaussian",n.trees=10000, interaction.depth = 4,
                    shrinkage = 0.02,verbose = F)
yhat.boost = predict(boost.boston, newdata=Boston[-train,], n.trees=5000)
mean((yhat.boost-boston.test)^2)
```

```
## [1] 9.825726
```

```
par(mfrow=c(1,2))  
plot(boost.boston, i ="rm")
```



```
plot(boost.boston, i ="lstat")
```

