RandomForest example

Abhirup Sen

01/06/2021

```
library(randomForest)
```

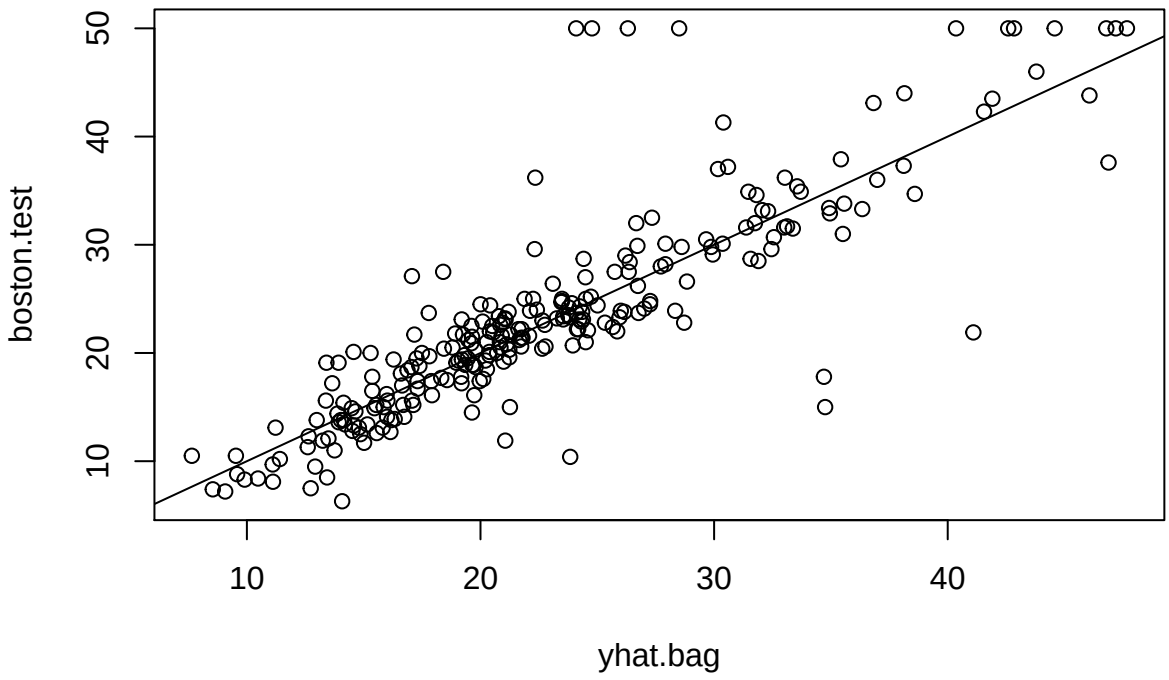
```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
library(MASS)
attach(Boston)
set.seed(1)
train = sample(1:nrow(Boston),nrow(Boston)/2)
bag.boston = randomForest(medv~.,data = Boston, subset=train, mtry=13, importance = TRUE)
bag.boston
```

```
##
## Call:
##  randomForest(formula = medv ~ ., data = Boston, mtry = 13, importance = TRUE,      subset = train)
##                Type of random forest: regression
##                      Number of trees: 500
## No. of variables tried at each split: 13
##
##           Mean of squared residuals: 11.33119
##                     % Var explained: 85.26
```

```
yhat.bag=predict(bag.boston, newdata=Boston[-train,])
boston.test = Boston[-train,"medv"]
plot(yhat.bag, boston.test)
abline(0,1)
```




```
mean((yhat.bag-boston.test)^2)
```

```
## [1] 23.4579
```

```
bag.boston = randomForest(medv~., data=Boston, subset=train,mtry=13,ntree=25)
yhat.bag = predict(bag.boston,newdata=Boston[-train,])
mean((yhat.bag-boston.test)^2)
```

```
## [1] 22.99145
```

```
set.seed(1)
rf.boston = randomForest(medv~., data=Boston,subset = train, mtry = 6, importance = TRUE)
yhat.rf = predict(rf.boston, newdata = Boston[-train,])
mean((yhat.rf - boston.test)^2)
```

```
## [1] 19.62021
```

```
importance(rf.boston)
```

```
##           %IncMSE IncNodePurity
## crim    16.697017    1076.08786
## zn       3.625784      88.35342
## indus    4.968621     609.53356
```

```
## chas     1.061432      52.21793
## nox     13.518179     709.87339
## rm      32.343305    7857.65451
## age     13.272498     612.21424
## dis      9.032477     714.94674
## rad      2.878434      95.80598
## tax      9.118801     364.92479
## ptratio  8.467062     823.93341
## black    7.579482     275.62272
## lstat   27.129817    6027.63740
```

```
varImpPlot(rf.boston)
```