Boosting Compare

nclivio

Saturday, October 25, 2014

Boosting Models

It was originally designed for classification problems, but it can profitably be extended to regression as well. The motivation for boosting was a procedure that combines the outputs of many "weak" classifiers to produce a powerful "committee." From this perspective boosting bears a resemblance to bagging and other committee-based approaches

Based on http://topepo.github.io/caret/Boosting.html I will describe each model and predictions for the ausair data.

```
library(fpp)
```

```
## Loading required package: forecast
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
##
## Loading required package: timeDate
## This is forecast 5.5
##
## Loading required package: fma
## Loading required package: tseries
## Loading required package: expsmooth
## Loading required package: lmtest
```

```
library(caret)
```

```
## Loading required package: lattice
## Loading required package: ggplot2
```

```
library(ggplot2)
data(ausair)

ts1<-ausair
years<-c(1970:2009)
data<-data.frame(Años=years,Pasajeros=ts1)
inTrain<-createDataPartition(y=data$Años, p=0.75, list=FALSE)
training<-data[inTrain,]
testing<-data[-inTrain,]
```

Models and Predictions

```
#Boosted Generalized Additive Model
modbsgam <- train(Pasajeros ~ ., method = "gamboost", data = training) #RMSE=2.74
```

```
## Loading required package: mboost
## Loading required package: parallel
## Loading required package: stabs
## This is mboost 2.4-0. See 'package?mboost' and the NEWS file
## for a complete list of changes.
##
##
## Attaching package: 'mboost'
##
## The following object is masked from 'package:ggplot2':
##
##     %+%
```

```
predbsgam <- predict(modbsgam, testing) #RMSE=2.71
#rmse(testrf$Pasajeros,predbsgam)
#[1] 2.207791

#Stochastic Gradient Boosting
#modbsgbm <- train(Pasajeros ~ ., method = "gbm", data = training)**    #No se pudo

 #Cubist
modbscub <- train(Pasajeros ~ ., method = "cubist", data = training)   #RMSE=1.6
```

```
## Loading required package: Cubist
```

```
predbscub <- predict(modbscub, testing)
#rmse(testrf$Pasajeros,predbscub)
#[1] 2.337806

#C5.0 y Cost-Sensitive C5.0
#modbsc5c <- train(Pasajeros ~ ., method = "C5.0Cost", data = training) #No se pudo
#modbsc5 <- train(Pasajeros ~ ., method = "C5.0", data = training, model=) #Nspudo

#Boosted Tree
modbstt <- train(Pasajeros ~ ., method = "bstTree", data = training)   #RMSE= 4.3
```

```
## Loading required package: bst
## Loading required package: plyr
##
## Attaching package: 'plyr'
##
## The following object is masked from 'package:fma':
##
##     ozone
```

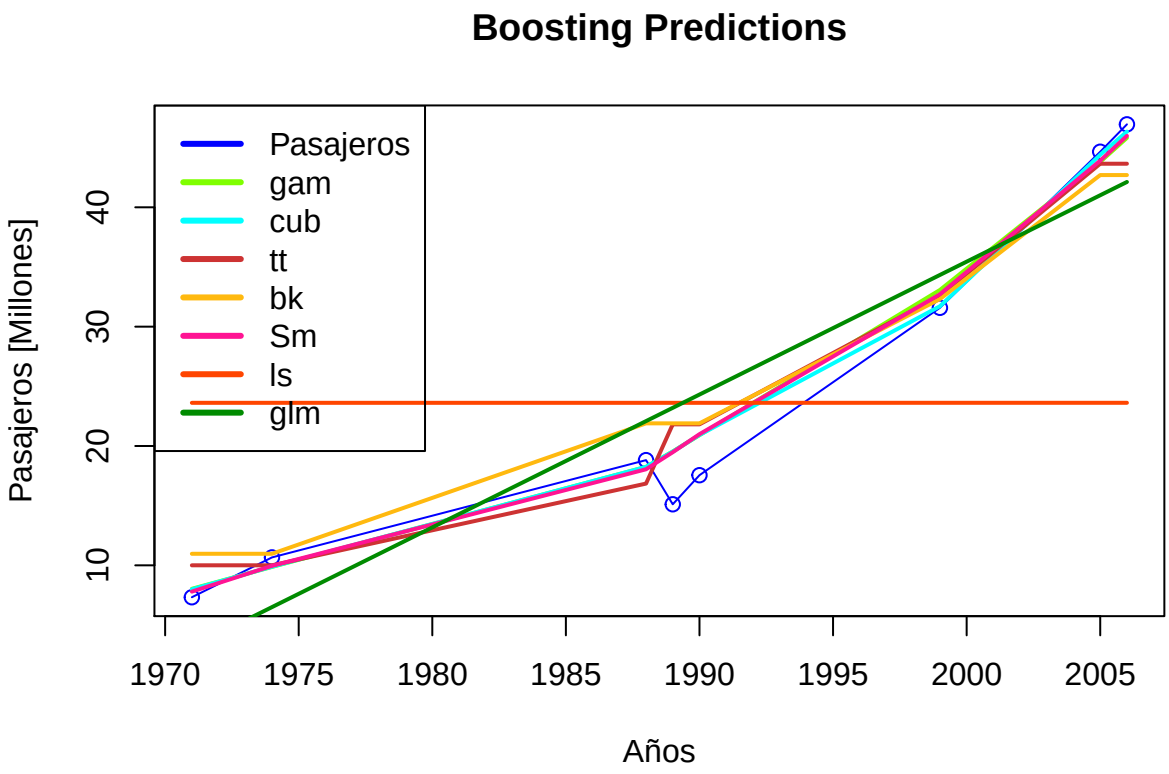
```r
predbstt <- predict(modbstt, testing)
#rmse(testrf$Pasajeros,predbstt)
#[1] 2.361692

#Boosted Tree
modbsbk <- train(Pasajeros ~ ., method = "blackboost", data = training) #RMSE= 4.55
```

```
## Loading required package: party
## Loading required package: grid
## Loading required package: sandwich
## Loading required package: strucchange
## Loading required package: modeltools
## Loading required package: stats4
##
## Attaching package: 'modeltools'
##
## The following object is masked from 'package:plyr':
##
##     empty
```

```r
predbsbk <- predict(modbsbk, testing)
#rmse(testrf$Pasajeros,predbsbk)
#[1] 3.021642

#Boosted Smoothing Spline
modbsSm <- train(Pasajeros ~ ., method = "bstSm", data = training)   #RMSE=1.27
predbsSm <- predict(modbsSm, testing)
#rmse(testrf$Pasajeros,predbsSm)
#[1] 2.245

#Boosted Logistic Regression
#modbslb <- train(Pasajeros ~ ., method = "LogitBoost", data = training)   #nsp

#Boosted Linear Model
modbsls <- train(Pasajeros ~ ., method = "bstLs", data = training) #RMSE= 13.9
predbsls <- predict(modbsls, testing)
#rmse(testrf$Pasajeros,predbsls)
#12.8992

#Boosted Generalized Linear Model
modbsglm <- train(Pasajeros ~ ., method = "glmboost", data = training) #RMSE= 3.97
predbsglm <- predict(modbsglm, testing)
#rmse(testrf$Pasajeros,predbsglm)
#[1] 4.372003

#Boosted Classification Trees
#modbsada <- train(Pasajeros ~ ., method = "ada", data = training) # error
#predbsada <- predict(modbsada, testing)
```

Plotting some Boosting predictions

```
plot(testing,xlab="Años",ylab="Pasajeros [Millones]",col="blue",
     main="Boosting Predictions")
lines(testing,lwd=1,col="blue")

#Boosted Generalized Additive Model
lines(testing$Años,predbsgam,lwd=2,col="chartreuse")

#Cubist
lines(testing$Años,predbscub,lwd=2,col="cyan1")

#Boosted Tree
lines(testing$Años,predbstt,lwd=2,col="brown3")
lines(testing$Años,predbsbk,lwd=2,col="darkgoldenrod1")

#Boosted Smoothing Spline
lines(testing$Años,predbsSm,lwd=2,col="deeppink")

#Boosted Linear Model
lines(testing$Años,predbsls,lwd=2,col="orangered")

#Boosted Generalized Linear Model
lines(testing$Años,predbsglm,lwd=2,col="green4")

legend("topleft",legend=c("Pasajeros","gam","cub","tt","bk","Sm","ls","glm"),
       col=c("blue","chartreuse","cyan1","brown3","darkgoldenrod1","deeppink"
             ,"orangered","green4"),lty=1,lwd=3)
```



The *RMSE* of the models are:

Boosted Generalized Additive Model predbsgam rmse=2.207791

Cubist predbscub rmse=2.337806

Boosted Tree predbstt rmse=2.361692

Boosted Tree predbsbk rmse 3.021642

Boosted Smoothing Spline predbsSm rmse=2.245

Boosted Linear Model predbsls rmse=12.8992

Boosted Generalized Linear Model predbsglm rmse=4.372003