

Boosting

W. Grossmann

```
#-----  
#Classification CRM data, boosting  
#-----  
#Libraries  
#-----  
  
library(adabag)          # Boosting  
library(pROC)  
  
## Warning: package 'pROC' was built under R version 3.2.5  
## Type 'citation("pROC")' for a citation.  
  
##  
## Attaching package: 'pROC'  
  
## Die folgenden Objekte sind maskiert von 'package:stats':  
##  
##      cov, smooth, var  
  
#-----  
#Data  
#-----  
  
data<-read.csv("CRM_class1.csv",header=T,sep=";")  
summary(data)  
  
##           ID           Age           sex           student           office  
## Min.      : 1.0      Min.      :1.000    female:678    no :866    no :435  
## 1st Qu.: 328.5    1st Qu.:1.000    male  :633    yes:445    yes:876  
## Median : 656.0    Median :1.000  
## Mean     : 656.0    Mean     :1.583  
## 3rd Qu.: 983.5    3rd Qu.:2.000  
## Max.     :1311.0    Max.     :3.000  
##  
##           Sales           SalesIndex           Duration_CR           UseS9  
## Min.      : 0.00      Min.      : 0.010    Min.      : 12.03    no :563  
## 1st Qu.: 17.48      1st Qu.: 1.605    1st Qu.: 31.80    yes:748  
## Median : 58.03      Median : 4.510    Median : 68.50  
## Mean     : 197.57    Mean     : 12.089    Mean     : 78.38  
## 3rd Qu.: 178.01      3rd Qu.: 12.630    3rd Qu.:119.36  
## Max.     :5869.48    Max.     :348.350    Max.     :188.13  
##  
##           Train  
## Min.      :0.0000
```

```

## 1st Qu.:0.0000
## Median :1.0000
## Mean   :0.6903
## 3rd Qu.:1.0000
## Max.   :1.0000
##
#-----
#classification of all data
#-----
data.boost<-subset(data, select = Age:UseS9)
data.boost<-as.data.frame(data.boost)
attach(data.boost)
dim(data.boost)

## [1] 1311    8

variable.names(data.boost)

## [1] "Age"      "sex"      "student"  "office"   "Sales"
## [6] "SalesIndex" "Duration_CR" "UseS9"

mod.boost<-boosting(UseS9~., data=data.boost, mfinal = 10)
summary(mod.boost)

##           Length Class  Mode
## formula      3  formula call
## trees        10 -none- list
## weights      10 -none- numeric
## votes       2622 -none- numeric
## prob         2622 -none- numeric
## class        1311 -none- character
## importance    7  -none- numeric
## terms         3  terms  call
## call          4  -none- call

pred<- predict(mod.boost, data.boost)
ls(pred)

## [1] "class"      "confusion" "error"      "formula"    "prob"       "votes"

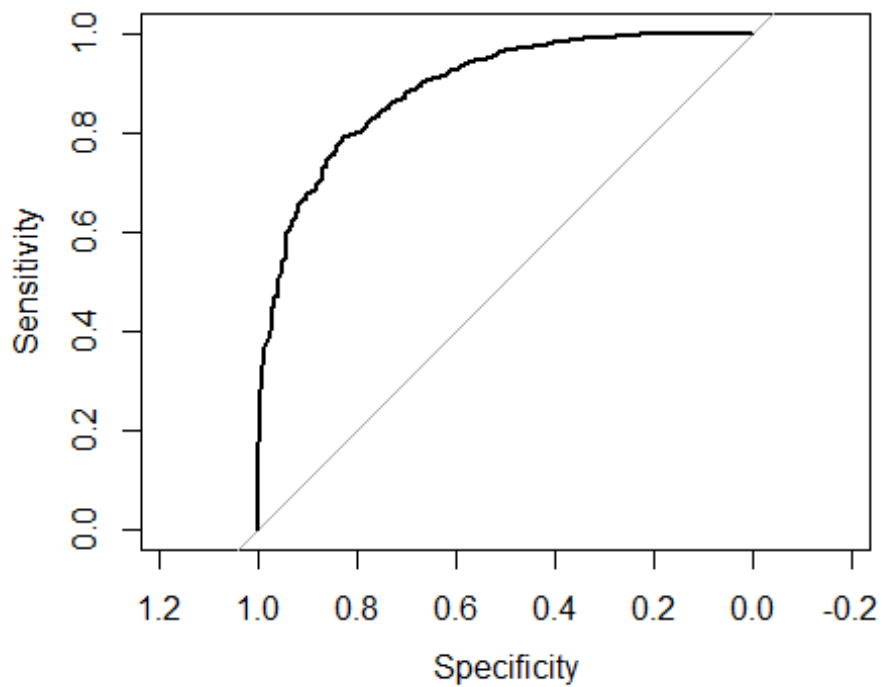
pred$confusion

##           Observed Class
## Predicted Class no yes
##           no 396 96
##           yes 167 652

#accuracy

```

```
#-----  
#ROC curve  
#-----  
prob<-pred$prob  
head(prob)  
  
##           [,1]      [,2]  
## [1,] 1.0000000 0.0000000  
## [2,] 0.4512196 0.5487804  
## [3,] 0.5422712 0.4577288  
## [4,] 0.7255404 0.2744596  
## [5,] 0.4985995 0.5014005  
## [6,] 0.3948482 0.6051518  
  
use9<-as.numeric(UseS9)-1  
  
plot.roc(use9,prob[,2])
```



```
auc(use9,prob[,2])  
## Area under the curve: 0.8907
```