Logistic Regression Demo

W. Grossmann

```
#-----------------------------------------------------
# Logistic Regression Demonstration
# ------------------------------------------------------------
# Libraries
#-------------------------------------------------------------
 library("e1071")         # naiveBayes

.library("klaR")          # NaiveBayes

 library("class")     # Functions for classifications
 library("MASS")      # Modern Applied Statistics with S
 library("pROC")      # for ROC

# ----------------------------------------------------------
# Data
#-----------------------------------------------------------

 demo1<- read.csv("LogisticData2.csv",header=T,sep=";")
head(demo1)

##   Duration ActInd UserType Quit
## 1     5.63   1.93  office     0
## 2     6.39   9.47  office     0
## 3     5.31   9.23  office     0
## 4     5.76  11.67  office     0
## 5     7.12   8.90  office     0
## 6     8.13   9.90  office     0

attach(demo1)
#--------------------------------------
#Logistic Regression
#--------------------------------------
mod.logistic<-glm(Quit~UserType+ActInd, data=demo1,
                  family = binomial())

summary(mod.logistic)

##
## Call:
## glm(formula = Quit ~ UserType + ActInd, family = binomial(),
##     data = demo1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.29456  -0.24724   0.03295   0.32015   2.46917
##
```

```
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.3847     1.8515   0.748   0.4545
## UserTypeprivate   3.0583     1.5978   1.914   0.0556 .
## ActInd           -0.5769     0.2735  -2.109   0.0349 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 33.271  on 23  degrees of freedom
## Residual deviance: 11.793  on 21  degrees of freedom
## AIC: 17.793
##
## Number of Fisher Scoring iterations: 6
```

```
drop1(mod.logistic)
```

```
## Single term deletions
##
## Model:
## Quit ~ UserType + ActInd
##          Df Deviance    AIC
## <none>        11.793 17.793
## UserType  1   16.135 20.135
## ActInd    1   17.864 21.864
```

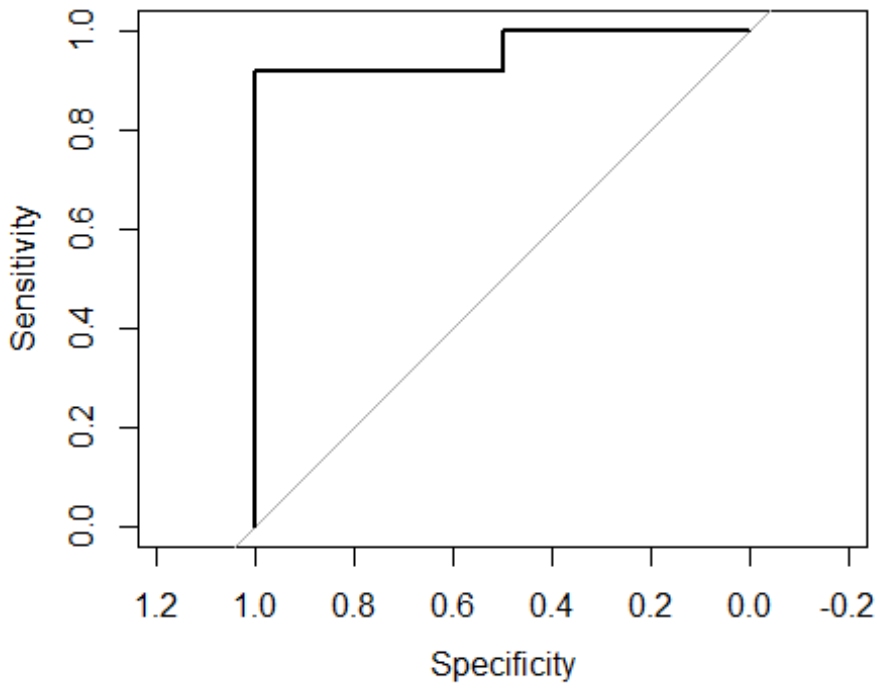
```
anova(mod.logistic)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Quit
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                        23     33.271
## UserType  1  15.4067        22     17.864
## ActInd    1   6.0715        21     11.793
```

```
prob<-fitted(mod.logistic)
head(prob)
```

```
##           1           2           3           4           5           6
## 0.567400498 0.016649223 0.019074477 0.004736114 0.022982599 0.013039188
```

```
plot.roc(Quit,prob )
```




```
auc(Quit,prob)
```

```
## Area under the curve: 0.9583
```

```
pred<-predict(mod.logistic, type = "response")
pred.class<-ifelse(pred >0.5, 1,0)
table(pred.class, Quit)
```

```
##           Quit
## pred.class  0  1
##          0 11  1
##          1  1 11
```