ROC\_AUC

Animesh Kansal

2/4/2019

library(pROC)

## Type 'citation("pROC")' for a citation.

##   
## Attaching package: 'pROC'

## The following objects are masked from 'package:stats':  
##   
## cov, smooth, var

library(randomForest)

## randomForest 4.6-14

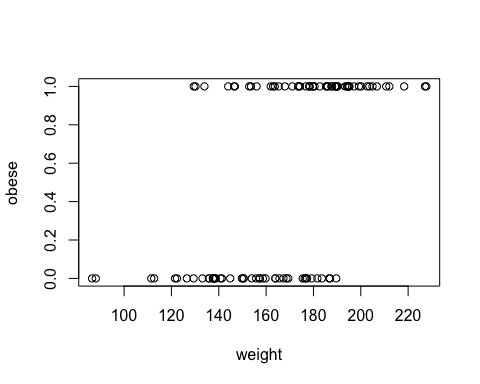
## Type rfNews() to see new features/changes/bug fixes.

set.seed(420)

num\_samples = 100  
weight = sort(rnorm(n = num\_samples,mean = 172,sd = 29))  
obese = ifelse(test = (runif(n=num\_samples) < (rank(weight)/100)), yes = 1,no = 0 )  
  
obese

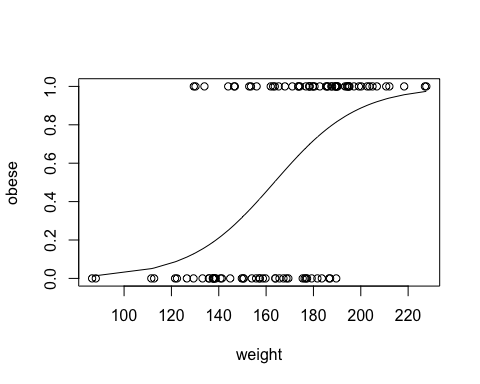
## [1] 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 0 1 1 0 0 1 0  
## [36] 0 0 0 1 1 1 0 0 1 0 0 1 0 0 1 1 1 1 0 0 1 0 0 1 1 1 0 1 1 1 0 1 0 1 1  
## [71] 1 0 0 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

plot(x= weight,y = obese)



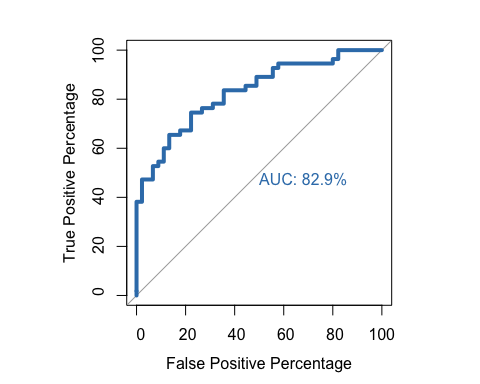
# Logistic regression

glm\_fit = glm(obese ~ weight, family = binomial)  
plot(x= weight,y = obese)  
lines(weight,glm\_fit$fitted.values)



# AUC ROC

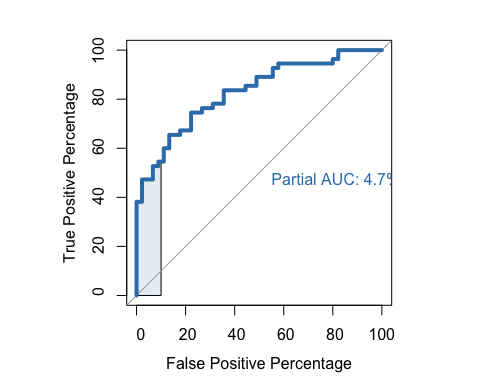
par(pty = "s")  
roc(  
obese,  
glm\_fit$fitted.values,  
plot = TRUE,  
legacy.axes = TRUE,  
percent = TRUE,  
xlab = "False Positive Percentage",  
ylab = "True Positive Percentage",  
col = "#377eb8",  
lwd = 4,  
print.auc = TRUE  
)



##   
## Call:  
## roc.default(response = obese, predictor = glm\_fit$fitted.values, percent = TRUE, plot = TRUE, legacy.axes = TRUE, xlab = "False Positive Percentage", ylab = "True Positive Percentage", col = "#377eb8", lwd = 4, print.auc = TRUE)  
##   
## Data: glm\_fit$fitted.values in 45 controls (obese 0) < 55 cases (obese 1).  
## Area under the curve: 82.91%

# Partial AUC ROC

par(pty = "s")  
roc(  
obese,  
glm\_fit$fitted.values,  
plot = TRUE,  
legacy.axes = TRUE,  
percent = TRUE,  
xlab = "False Positive Percentage",  
ylab = "True Positive Percentage",  
col = "#377eb8",  
lwd = 4,  
print.auc = TRUE,  
print.auc.x = 45,  
partial.auc=c(100,90),  
auc.polygon = TRUE,  
auc.polygon.col = '#377eb822'  
)



##   
## Call:  
## roc.default(response = obese, predictor = glm\_fit$fitted.values, percent = TRUE, plot = TRUE, legacy.axes = TRUE, xlab = "False Positive Percentage", ylab = "True Positive Percentage", col = "#377eb8", lwd = 4, print.auc = TRUE, print.auc.x = 45, partial.auc = c(100, 90), auc.polygon = TRUE, auc.polygon.col = "#377eb822")  
##   
## Data: glm\_fit$fitted.values in 45 controls (obese 0) < 55 cases (obese 1).  
## Partial area under the curve (specificity 100%-90%): 4.727%

roc\_info = roc(obese, glm\_fit$fitted.values, legacy.axes = TRUE)  
roc\_def = data.frame(  
 tpp = roc\_info$sensitivities \* 100,  
 fpp = (1 - roc\_info$specificities) \* 100,  
 thresholds = roc\_info$thresholds  
)  
print('head')

## [1] "head"

head(roc\_def)

## tpp fpp thresholds  
## 1 100 100.00000 -Inf  
## 2 100 97.77778 0.01349011  
## 3 100 95.55556 0.03245008  
## 4 100 93.33333 0.05250145  
## 5 100 91.11111 0.07017225  
## 6 100 88.88889 0.08798755

print('tail')

## [1] "tail"

tail(roc\_def)

## tpp fpp thresholds  
## 96 9.090909 0 0.9275222  
## 97 7.272727 0 0.9371857  
## 98 5.454545 0 0.9480358  
## 99 3.636364 0 0.9648800  
## 100 1.818182 0 0.9735257  
## 101 0.000000 0 Inf

# Choose best threshold

roc\_def[roc\_def$tpp>60 & roc\_def$tpp < 80 ,]

## tpp fpp thresholds  
## 42 78.18182 35.55556 0.5049310  
## 43 78.18182 33.33333 0.5067116  
## 44 78.18182 31.11111 0.5166680  
## 45 76.36364 31.11111 0.5287933  
## 46 76.36364 28.88889 0.5429351  
## 47 76.36364 26.66667 0.5589494  
## 48 74.54545 26.66667 0.5676342  
## 49 74.54545 24.44444 0.5776086  
## 50 74.54545 22.22222 0.5946054  
## 51 72.72727 22.22222 0.6227449  
## 52 70.90909 22.22222 0.6398136  
## 53 69.09091 22.22222 0.6441654  
## 54 67.27273 22.22222 0.6556705  
## 55 67.27273 20.00000 0.6683618  
## 56 67.27273 17.77778 0.6767661  
## 57 65.45455 17.77778 0.6802060  
## 58 65.45455 15.55556 0.6831936  
## 59 65.45455 13.33333 0.6917225  
## 60 63.63636 13.33333 0.6975300  
## 61 61.81818 13.33333 0.6982807

# Random forest classifier

rf\_model = randomForest(factor(obese) ~ weight)

par(pty = "s")  
roc(  
obese,  
glm\_fit$fitted.values,  
plot = TRUE,  
legacy.axes = TRUE,  
percent = TRUE,  
xlab = "False Positive Percentage",  
ylab = "True Positive Percentage",  
col = "#377eb8",  
lwd = 4,  
print.auc = TRUE  
)

##   
## Call:  
## roc.default(response = obese, predictor = glm\_fit$fitted.values, percent = TRUE, plot = TRUE, legacy.axes = TRUE, xlab = "False Positive Percentage", ylab = "True Positive Percentage", col = "#377eb8", lwd = 4, print.auc = TRUE)  
##   
## Data: glm\_fit$fitted.values in 45 controls (obese 0) < 55 cases (obese 1).  
## Area under the curve: 82.91%

plot.roc(obese,rf\_model$votes[,1],percent= TRUE,col = '#4daf4a',lwd = 4,print.auc=TRUE,add=TRUE, print.auc.y=40)  
  
legend("bottomright",legend=c("Random forest","Logistic Regression"),col =c("#4dac4a","#377eb8"),lwd= 4)

