1. Use the below given data set

Data Set

2. Perform the below given activities:

a. Create classification model using logistic regression model

b. verify model goodness of fit

c. Report the accuracy measures

d. Report the variable importance

e. Report the unimportant variables

f. Interpret the results

g. Visualize the results

Solution:

library(C50)

data(churn)

head(churnTrain)

head(churnTest)

#churnTrain = churnTrain[1:500,]

#churnTest = churnTest[1:500,]

# logistic regression model:

fit <- glm(churn~.,data = churnTrain,family = binomial(link='logit'))

summary(fit)

library(MASS)

step\_fit <- stepAIC(fit,method='backward')

summary(step\_fit)

confint(step\_fit)

#ANOVA on base model

anova(fit,test = 'Chisq')

#ANOVA from reduced model after applying the Step AIC

anova(step\_fit,test = 'Chisq')

#plot the fitted model

plot(fit$fitted.values)

pred\_link <- predict(fit,newdata = churnTest,type = 'link')

#check for multicollinearity

library(car)

vif(fit)

vif(step\_fit)

pred <- predict(fit,newdata = churnTest,type = 'response')

#check the AUC curve

library(pROC)

g <- roc(churn ~ pred, data = churnTest)

g

plot(g)

library(caret)

#with default prob cut 0.50

churnTest$pred\_churn <- ifelse(pred<0.7,'yes','no')

table(churnTest$pred\_churn,churnTest$churn)

#training split of churn classes

round(table(churnTrain$churn)/nrow(churnTrain),2)\*100

# test split of churn classes

round(table(churnTest$churn)/nrow(churnTest),2)\*100

#predicted split of churn classes

round(table(churnTest$pred\_churn)/nrow(churnTest),2)\*100

#create confusion matrix

confusionMatrix(churnTest$churn,churnTest$pred\_churn)

#how do we create a cross validation scheme

control <- trainControl(method = 'repeatedcv',

number = 10,

repeats = 3)

seed <-7

metric <- 'Accuracy'

set.seed(seed)

fit\_default <- train(churn~.,

data = churnTrain,

method = 'glm',

metric = metric,

trControl = control)

print(fit\_default)

library(caret)

varImp(step\_fit)

varImp(fit\_default)

library(devtools)

install\_github("riv","tomasgreif")

install\_github("woe","tomasgreif")

library(woe)

library(riv)

iv\_df <- iv.mult(churnTrain, y="churn", summary=TRUE, verbose=TRUE)

iv\_df

iv <- iv.mult(churnTrain, y="churn", summary=FALSE, verbose=TRUE)

# Plot information value summary

iv.plot.summary(iv\_df)

