LogisticRegression.R

Amisha Sancheti

2020-04-10

```
emp_churn <- read.csv("C:/Users/Amisha Sancheti/Desktop/MITA
sem2/Multivariate Analysis/Project/attritiondata.csv")

##########Exploring the Dataset###########
str(emp_churn)
```

```
## 'data.frame':    1470 obs. of  30 variables:
##  $ Age                     : int  41 49 37 33 27 32 59 30 38 36 ...
##  $ Attrition               : Factor w/ 2 levels "No","Yes": 2 1 2 1 1 1 1
1 1 1 ...
##  $ BusinessTravel          : Factor w/ 3 levels "Non-
Travel","Travel_Frequently",..: 3 2 3 2 3 2 3 3 2 3 ...
##  $ DailyRate               : int  1102 279 1373 1392 591 1005 1324 1358
216 1299 ...
##  $ Department              : Factor w/ 3 levels "Human Resources",..: 3 2
2 2 2 2 2 2 2 2 ...
##  $ DistanceFromHome        : int  1 8 2 3 2 2 3 24 23 27 ...
##  $ Education               : int  2 1 2 4 1 2 3 1 3 3 ...
##  $ EducationField          : Factor w/ 6 levels "Human Resources",..: 2 2
5 2 4 2 4 2 2 4 ...
##  $ EmployeeNumber          : int  1 2 4 5 7 8 10 11 12 13 ...
##  $ EnvironmentSatisfaction : int  2 3 4 4 1 4 3 4 4 3 ...
##  $ Gender                  : Factor w/ 2 levels "Female","Male": 1 2 2 1 2
2 1 2 2 2 ...
##  $ HourlyRate              : int  94 61 92 56 40 79 81 67 44 94 ...
##  $ JobInvolvement          : int  3 2 2 3 3 3 4 3 2 3 ...
##  $ JobRole                 : Factor w/ 9 levels "Healthcare
Representative",..: 8 7 3 7 3 3 3 3 5 1 ...
##  $ JobSatisfaction         : int  4 2 3 3 2 4 1 3 3 3 ...
##  $ MaritalStatus           : Factor w/ 3 levels "Divorced","Married",..: 3
2 3 2 2 3 2 1 3 2 ...
##  $ MonthlyIncome           : int  5993 5130 2090 2909 3468 3068 2670 2693
9526 5237 ...
##  $ MonthlyRate             : int  19479 24907 2396 23159 16632 11864 9964
13335 8787 16577 ...
##  $ NumCompaniesWorked      : int  8 1 6 1 9 0 4 1 0 6 ...
##  $ OverTime                : Factor w/ 2 levels "No","Yes": 2 1 2 2 1 1 2
1 1 1 ...
##  $ PerformanceRating       : int  3 4 3 3 3 3 4 4 4 3 ...
##  $ RelationshipSatisfaction: int  1 4 2 3 4 3 1 2 2 2 ...
##  $ StockOptionLevel        : int  0 1 0 0 1 0 3 1 0 2 ...
##  $ TotalWorkingYears       : int  8 10 7 8 6 8 12 1 10 17 ...
```

```
##  $ TrainingTimesLastYear   : int  0 3 3 3 3 2 3 2 2 3 ...
##  $ WorkLifeBalance         : int  1 3 3 3 3 2 2 3 3 2 ...
##  $ YearsAtCompany          : int  6 10 0 8 2 7 1 1 9 7 ...
##  $ YearsInCurrentRole      : int  4 7 0 7 2 7 0 0 7 7 ...
##  $ YearsSinceLastPromotion : int  0 1 0 3 2 3 0 0 1 7 ...
##  $ YearsWithCurrManager    : int  5 7 0 0 2 6 0 0 8 7 ...
```

```
library(data.table)
setDT(emp_churn)
class(emp_churn)
```

```
## [1] "data.table" "data.frame"
```

```
library(caTools)

#we will partition the data into test and train with 80:20 ratio.
partn = floor(0.80*nrow(emp_churn))

set.seed(123)
train_ind = sample(seq_len(nrow(emp_churn)),size = partn)

empchurn_train = emp_churn[train_ind,]
empchurn_test= emp_churn[-train_ind,]

dim(emp_churn)
```

```
## [1] 1470   30
```

```
dim(empchurn_train)
```

```
## [1] 1176   30
```

```
empchurn_train[is.na(empchurn_train)] <- 0

## applying logistic regression:
logreg_model <- glm(Attrition ~.,family=binomial(link='logit'),na.action =
na.exclude,data=empchurn_train)
summary(logreg_model)
```

```
##
## Call:
## glm(formula = Attrition ~ ., family = binomial(link = "logit"),
##     data = empchurn_train, na.action = na.exclude)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7622  -0.5101  -0.2662  -0.1009   3.3234
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -1.065e+01  6.679e+02  -0.016 0.987272
```

```
## Age                                -5.014e-02  1.609e-02  -3.115 0.001838 **
## BusinessTravelTravel_Frequently     1.827e+00  4.607e-01   3.966 7.31e-05
***
## BusinessTravelTravel_Rarely         1.067e+00  4.252e-01   2.510 0.012082 *
## DailyRate                          -1.458e-04  2.450e-04  -0.595 0.551700
## DepartmentResearch & Development    1.374e+01  6.679e+02   0.021 0.983585
## DepartmentSales                     1.382e+01  6.679e+02   0.021 0.983493
## DistanceFromHome                    3.343e-02  1.201e-02   2.783 0.005388 **
## Education                           1.411e-02  9.801e-02   0.144 0.885528
## EducationFieldLife Sciences        -1.163e+00  9.551e-01  -1.218 0.223264
## EducationFieldMarketing            -5.014e-01  1.007e+00  -0.498 0.618383
## EducationFieldMedical              -1.179e+00  9.525e-01  -1.237 0.215978
## EducationFieldOther                -9.692e-01  1.003e+00  -0.966 0.334011
## EducationFieldTechnical Degree     -1.828e-01  9.771e-01  -0.187 0.851624
## EmployeeNumber                     -1.454e-04  1.691e-04  -0.860 0.389804
## EnvironmentSatisfaction            -3.869e-01  8.967e-02  -4.315 1.60e-05
***
## GenderMale                          1.387e-01  2.037e-01   0.681 0.495712
## HourlyRate                         -1.077e-03  4.885e-03  -0.220 0.825524
## JobInvolvement                     -5.121e-01  1.364e-01  -3.755 0.000173
***
## JobRoleHuman Resources              1.466e+01  6.679e+02   0.022 0.982482
## JobRoleLaboratory Technician        1.527e+00  5.121e-01   2.983 0.002856 **
## JobRoleManager                     -5.328e-03  1.037e+00  -0.005 0.995899
## JobRoleManufacturing Director       3.082e-01  5.599e-01   0.550 0.581989
## JobRoleResearch Director           -1.046e+00  1.032e+00  -1.014 0.310796
## JobRoleResearch Scientist           5.291e-01  5.210e-01   1.016 0.309851
## JobRoleSales Executive              7.345e-01  1.227e+00   0.599 0.549410
## JobRoleSales Representative         1.825e+00  1.288e+00   1.416 0.156772
## JobSatisfaction                    -3.968e-01  8.948e-02  -4.435 9.23e-06
***
## MaritalStatusMarried                4.126e-01  3.035e-01   1.360 0.173880
## MaritalStatusSingle                 1.237e+00  3.924e-01   3.152 0.001623 **
## MonthlyIncome                       4.883e-06  6.718e-05   0.073 0.942056
## MonthlyRate                         7.525e-06  1.388e-05   0.542 0.587620
## NumCompaniesWorked                  1.799e-01  4.299e-02   4.184 2.86e-05
***
## OverTimeYes                         1.803e+00  2.124e-01   8.489  < 2e-16
***
## PerformanceRating                  -2.464e-02  2.755e-01  -0.089 0.928732
## RelationshipSatisfaction           -2.218e-01  9.120e-02  -2.432 0.015028 *
## StockOptionLevel                   -1.241e-01  1.774e-01  -0.699 0.484250
## TotalWorkingYears                  -4.359e-02  3.133e-02  -1.391 0.164156
## TrainingTimesLastYear              -2.048e-01  8.086e-02  -2.532 0.011330 *
## WorkLifeBalance                    -3.157e-01  1.338e-01  -2.359 0.018322 *
## YearsAtCompany                      8.220e-02  4.180e-02   1.966 0.049255 *
## YearsInCurrentRole                 -1.560e-01  4.928e-02  -3.166 0.001544 **
## YearsSinceLastPromotion             1.571e-01  4.603e-02   3.413 0.000642
***
## YearsWithCurrManager               -9.910e-02  5.183e-02  -1.912 0.055879 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1036.88  on 1175  degrees of freedom
## Residual deviance:  708.17  on 1132  degrees of freedom
## AIC: 796.17
##
## Number of Fisher Scoring iterations: 15
```

```
round(exp(coef(logreg_model)), 2)
```

```
##                       (Intercept)                               Age
##                              0.00                              0.95
##  BusinessTravelTravel_Frequently      BusinessTravelTravel_Rarely
##                              6.22                              2.91
##                         DailyRate DepartmentResearch & Development
##                              1.00                         927852.79
##                   DepartmentSales                 DistanceFromHome
##                         1002682.22                              1.03
##                         Education      EducationFieldLife Sciences
##                              1.01                              0.31
##          EducationFieldMarketing            EducationFieldMedical
##                              0.61                              0.31
##              EducationFieldOther   EducationFieldTechnical Degree
##                              0.38                              0.83
##                    EmployeeNumber          EnvironmentSatisfaction
##                              1.00                              0.68
##                        GenderMale                        HourlyRate
##                              1.15                              1.00
##                    JobInvolvement           JobRoleHuman Resources
##                              0.60                        2336669.04
##     JobRoleLaboratory Technician                   JobRoleManager
##                              4.61                              0.99
##    JobRoleManufacturing Director         JobRoleResearch Director
##                              1.36                              0.35
##        JobRoleResearch Scientist           JobRoleSales Executive
##                              1.70                              2.08
##      JobRoleSales Representative                   JobSatisfaction
##                              6.20                              0.67
##             MaritalStatusMarried              MaritalStatusSingle
##                              1.51                              3.44
##                    MonthlyIncome                       MonthlyRate
##                              1.00                              1.00
##               NumCompaniesWorked                       OverTimeYes
##                              1.20                              6.07
##                PerformanceRating         RelationshipSatisfaction
##                              0.98                              0.80
##                 StockOptionLevel                 TotalWorkingYears
```

```
##                              0.88                              0.96
##            TrainingTimesLastYear                   WorkLifeBalance
##                              0.81                              0.73
##                   YearsAtCompany               YearsInCurrentRole
##                              1.09                              0.86
##          YearsSinceLastPromotion             YearsWithCurrManager
##                              1.17                              0.91

logreg_pred <- predict(logreg_model,empchurn_test,type="response" )
logreg_pred

##            1            2            3            4            5
6
## 6.881501e-01 4.491745e-02 8.665631e-01 5.494477e-01 9.464370e-03
9.151841e-01
##            7            8            9           10           11
12
## 9.831324e-02 8.249842e-01 1.437346e-01 1.862026e-02 3.789682e-01
7.785227e-02
##           13           14           15           16           17
18
## 9.555475e-03 3.042317e-01 1.701256e-02 8.928631e-03 4.047283e-01
9.389703e-02
##           19           20           21           22           23
24
## 5.566506e-02 4.379948e-01 1.421114e-01 4.784779e-01 2.057009e-02
3.381231e-02
##           25           26           27           28           29
30
## 3.214471e-08 3.991481e-02 2.249349e-02 4.129405e-01 6.791931e-01
7.938076e-02
##           31           32           33           34           35
36
## 3.214758e-02 5.060923e-02 2.867411e-01 8.509580e-02 1.202598e-01
1.028478e-01
##           37           38           39           40           41
42
## 3.396471e-01 2.626931e-02 3.916565e-01 5.117851e-02 3.378583e-01
6.021318e-02
##           43           44           45           46           47
48
## 7.365360e-01 4.175585e-02 9.433669e-03 7.960637e-04 3.684306e-02
4.991763e-02
##           49           50           51           52           53
54
## 4.119912e-03 5.541478e-02 4.819726e-02 7.673766e-03 6.313838e-02
6.264179e-03
##           55           56           57           58           59
60
## 4.704367e-02 2.570623e-01 1.554808e-01 5.326460e-01 5.160662e-03
```

```
1.174608e-01
##           61           62           63           64           65
66
## 9.357054e-03 6.710255e-03 9.351471e-02 1.219135e-01 2.409801e-02
2.501798e-01
##           67           68           69           70           71
72
## 9.888769e-02 1.404597e-02 1.508893e-02 3.470187e-02 1.917560e-02
1.375075e-01
##           73           74           75           76           77
78
## 8.662167e-02 2.143217e-01 4.998304e-03 5.725074e-01 3.983686e-03
1.791414e-02
##           79           80           81           82           83
84
## 3.252800e-01 5.153524e-03 1.188987e-01 1.249005e-01 8.961190e-03
5.903851e-02
##           85           86           87           88           89
90
## 2.008660e-03 4.098870e-02 7.583019e-01 3.469624e-01 4.601827e-01
2.269919e-02
##           91           92           93           94           95
96
## 1.785220e-02 1.192321e-01 1.689908e-02 7.753706e-03 8.511920e-02
5.876820e-03
##           97           98           99          100          101
102
## 7.188444e-02 2.274694e-01 1.163923e-01 6.576043e-02 5.852396e-03
4.696956e-01
##          103          104          105          106          107
108
## 3.972061e-02 1.783544e-01 1.682313e-01 3.274212e-01 6.811577e-01
1.628573e-02
##          109          110          111          112          113
114
## 1.211451e-01 1.453674e-01 2.279103e-01 1.108452e-02 3.813180e-02
1.703899e-02
##          115          116          117          118          119
120
## 1.070331e-01 1.624354e-01 5.653804e-03 1.213312e-01 1.352661e-02
5.117664e-01
##          121          122          123          124          125
126
## 7.637758e-02 6.853137e-01 8.463897e-02 1.015945e-01 6.666160e-02
4.576728e-02
##          127          128          129          130          131
132
## 5.283717e-02 1.140613e-01 5.864981e-01 8.649474e-02 2.208575e-02
7.450971e-03
##          133          134          135          136          137
```

```
138
## 8.253257e-02 9.108629e-02 3.266319e-01 3.865223e-02 4.255479e-02
2.576089e-01
##          139          140          141          142          143
144
## 8.743418e-01 6.919122e-02 5.217708e-02 2.525506e-01 1.923488e-01
1.453881e-01
##          145          146          147          148          149
150
## 6.956300e-01 1.028439e-01 7.323294e-02 4.232023e-03 6.471320e-01
4.825716e-02
##          151          152          153          154          155
156
## 4.653044e-01 8.446710e-01 7.133982e-02 5.500552e-02 1.053069e-02
8.183185e-01
##          157          158          159          160          161
162
## 1.026368e-01 1.030027e-01 3.014940e-02 4.739658e-01 7.757665e-03
2.315346e-03
##          163          164          165          166          167
168
## 1.108945e-01 2.838702e-02 2.564742e-01 1.192756e-01 1.544312e-02
4.504492e-02
##          169          170          171          172          173
174
## 3.404136e-02 1.326255e-02 1.751769e-01 9.420997e-02 7.096038e-02
1.091941e-01
##          175          176          177          178          179
180
## 1.661494e-01 2.036006e-02 9.957361e-03 7.217772e-02 1.007065e-01
3.947591e-01
##          181          182          183          184          185
186
## 4.768431e-03 1.167714e-02 1.387057e-03 3.233793e-02 1.921303e-02
1.359356e-02
##          187          188          189          190          191
192
## 4.418864e-01 8.489871e-01 4.069940e-01 7.844585e-02 6.256148e-01
7.724598e-02
##          193          194          195          196          197
198
## 1.156128e-02 1.640975e-01 3.458973e-02 6.324389e-02 7.370505e-02
2.725376e-02
##          199          200          201          202          203
204
## 4.895363e-01 4.525209e-02 6.164411e-01 2.880220e-01 2.000014e-01
4.766072e-01
##          205          206          207          208          209
210
## 3.499571e-02 3.667813e-02 2.665505e-02 7.799304e-02 8.683063e-02
```

```
1.985738e-01
##          211          212          213          214          215
216
## 2.145597e-02 7.692436e-02 1.209671e-01 6.808270e-01 6.062561e-02
8.166097e-02
##          217          218          219          220          221
222
## 8.359876e-01 1.430178e-01 4.721838e-02 9.509808e-02 3.351981e-02
9.690810e-02
##          223          224          225          226          227
228
## 2.640395e-01 7.876597e-03 2.314058e-01 6.684532e-02 2.272447e-02
9.013169e-02
##          229          230          231          232          233
234
## 9.167267e-02 2.119067e-01 2.819757e-02 6.455799e-01 1.182633e-01
1.981893e-01
##          235          236          237          238          239
240
## 4.001281e-01 7.617717e-03 1.533568e-01 3.005239e-01 4.384578e-01
6.266640e-02
##          241          242          243          244          245
246
## 3.795825e-02 5.435288e-03 5.082718e-02 1.190417e-02 6.909313e-01
2.376894e-01
##          247          248          249          250          251
252
## 4.518628e-02 1.236215e-01 8.415152e-02 4.535745e-02 3.063447e-01
3.143914e-03
##          253          254          255          256          257
258
## 2.129022e-01 4.804031e-02 3.405040e-02 1.210218e-02 6.671248e-02
2.508650e-01
##          259          260          261          262          263
264
## 4.766138e-01 2.865852e-03 8.750447e-01 2.580668e-03 7.248715e-01
5.754999e-01
##          265          266          267          268          269
270
## 1.488204e-01 3.670788e-01 5.834742e-01 7.999044e-02 4.174123e-02
1.224643e-01
##          271          272          273          274          275
276
## 1.153238e-01 1.679260e-03 2.433304e-02 2.423117e-02 2.067021e-02
3.457511e-03
##          277          278          279          280          281
282
## 2.063727e-01 2.139014e-02 1.141001e-02 1.589813e-01 2.581922e-02
8.799937e-02
##          283          284          285          286          287
```

```
288
## 1.898113e-03 1.156024e-02 2.044734e-01 2.348620e-01 4.630752e-03
2.058743e-02
##          289          290          291          292          293
294
## 7.576197e-02 2.386428e-02 2.007438e-01 8.677358e-02 2.818839e-02
1.072993e-02
```

```
confusion_matrix <- table(empchurn_test$Attrition, logreg_pred > 0.5)
confusion_matrix
```

```
##
##       FALSE TRUE
##   No    240    6
##   Yes    24   24
```

```
#so we have 240 true positive and true negative 24, 6 false negatives and 24
false positives
#Now we will see the accuracy of our classification.

accuracy <- (sum(diag(confusion_matrix)) / sum(confusion_matrix))
accuracy
```

```
## [1] 0.8979592
```

```
# So our model gives 89.79% of accuracy which is good for our data.

#install.packages("pROC",
lib="/Library/Frameworks/R.framework/Versions/3.5/Resources/library")
library(pROC)
```

```
## Warning: package 'pROC' was built under R version 3.6.3
```

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

##
## Attaching package: 'pROC'
```

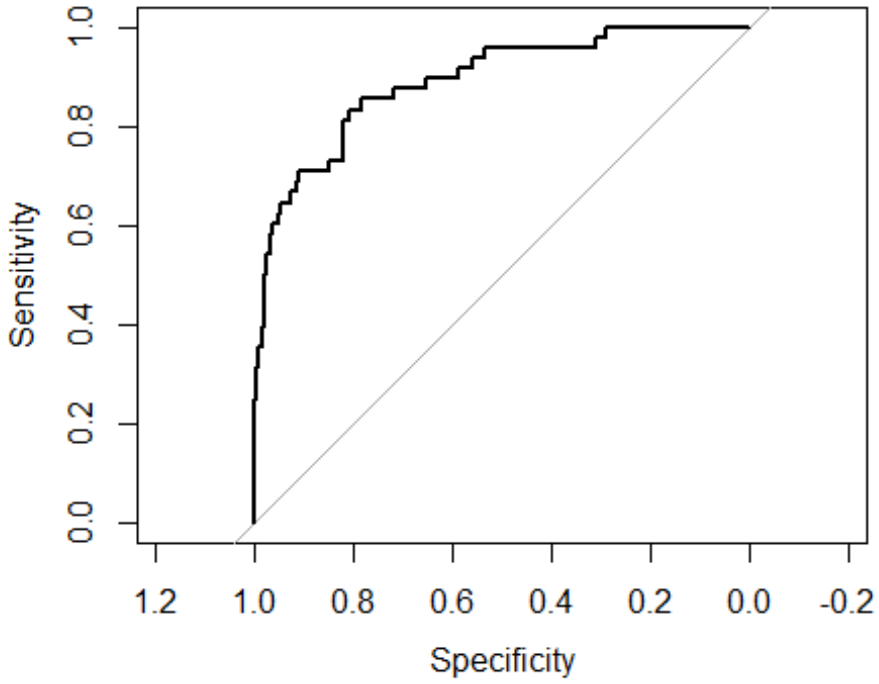
```
## The following objects are masked from 'package:stats':
##
##     cov, smooth, var
```

```
logreg_roc <- roc(empchurn_test$Attrition, logreg_pred)
```

```
## Setting levels: control = No, case = Yes
```

```
## Setting direction: controls < cases
```

```
plot.roc(logreg_roc)
```




```
##it is a plot of true positive rate against false positive rate. The graph
shows that we have pretty good accuracy for our data.
```