Logistic Regresison Examples

Abhirup Sen

02/06/2021

load Data

```
library(ROCR)
data(mtcars)
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
mtcars$vs <- as.factor(mtcars$vs)
```

```
#step 1  split the data
library(caTools)
set.seed(100)

split <- sample.split(mtcars, SplitRatio = 0.8)
train <- subset(mtcars, split = TRUE)
test <- subset(mtcars, split = FALSE)


#step 2 : Train the dataset
logit_model <- glm(vs~wt+disp, data = train, family="binomial")
summary(logit_model)
```

```
##
## Call:
## glm(formula = vs ~ wt + disp, family = "binomial", data = train)
##
```

```
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.67506  -0.28444  -0.08401   0.57281   2.08234
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.60859    2.43903   0.660    0.510
## wt           1.62635    1.49068   1.091    0.275
## disp        -0.03443    0.01536  -2.241    0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 43.86  on 31  degrees of freedom
## Residual deviance: 21.40  on 29  degrees of freedom
## AIC: 27.4
##
## Number of Fisher Scoring iterations: 6
```

```
#step 3: Predict test
fitted.result <- predict(logit_model,test,type="response")
fitted.result
```

```
##           Mazda RX4       Mazda RX4 Wag          Datsun 710      Hornet 4 Drive
##         0.589098973         0.684593276         0.840625523         0.114398085
##   Hornet Sportabout             Valiant          Duster 360           Merc 240D
##         0.005525208         0.374768453         0.006817186         0.851350376
##            Merc 230            Merc 280           Merc 280C          Merc 450SE
##         0.867993906         0.807236894         0.807236894         0.219433362
##          Merc 450SL         Merc 450SLC  Cadillac Fleetwood Lincoln Continental
##         0.139202255         0.149234967         0.002224998         0.004453588
##   Chrysler Imperial            Fiat 128         Honda Civic      Toyota Corolla
##         0.007772280         0.922487560         0.835966790         0.895173677
##       Toyota Corona    Dodge Challenger          AMC Javelin           Camaro Z28
##         0.814883948         0.026171375         0.036518408         0.014802949
##    Pontiac Firebird           Fiat X1-9       Porsche 914-2        Lotus Europa
##         0.002700619         0.884456037         0.720433157         0.688821969
##      Ford Pantera L        Ferrari Dino       Maserati Bora          Volvo 142E
##         0.004858739         0.754118670         0.049742275         0.876897600
```

```
#step 4: change the outcome to probabilities
fitted.result <- ifelse(fitted.result > 0.5,1,0)

#step 5 :  model evaluation
table(fitted.result, test$vs)
```

```
##
## fitted.result  0  1
##             0 14  2
##             1  4 12
```

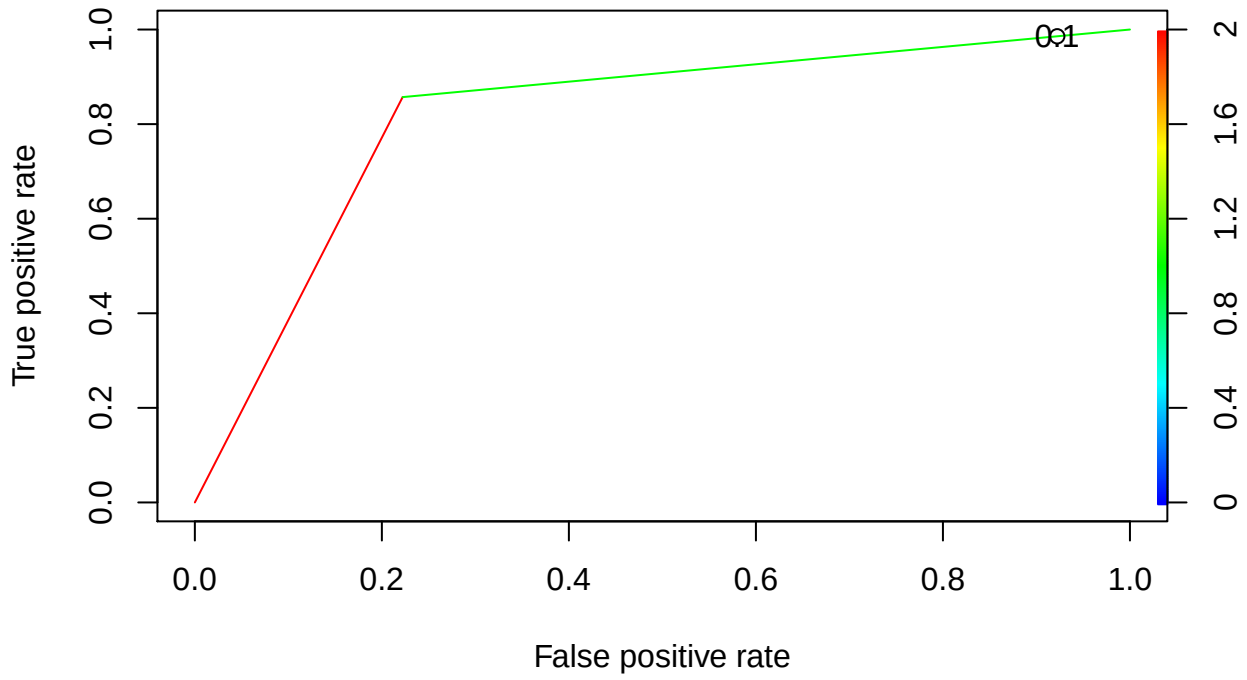
```
misClassError <- mean(fitted.result != test$vs)
print(paste('Accuracy =', 1 - misClassError))
```

```
## [1] "Accuracy = 0.8125"
```

```
#ROC AUC Curve

ROCRPred <- prediction(fitted.result, test$vs)
ROCRPerf <- performance(ROCRPred, measure ="tpr", x.measure = "fpr")
```

```
#plot(str(ROCRPred))
plot(ROCRPerf, colorize = TRUE)
plot(ROCRPerf, colorize = TRUE, print.cutoffs.at = seq(0.1,  0.1))
```




```
plot(ROCRPerf, colorize = TRUE, print.cutoffs.at = seq(0.1,  0.1),main = "ROC Curve")
abline(a=0, b =1)

auc <- performance(ROCRPred, measure = "auc")
auc <- auc@y.values[[1]]
auc
```

```
## [1] 0.8174603
```

```r
auc <- round(auc,4)
legend(.5,.4,auc, title = "AUC")
```