Practical No. 07

AIM - Practical of Logistics Regression.

Source Code - *(CSV → shorturl.at/fBIS0)

```
loan <- read.csv(file.choose(), header=TRUE, sep=",")
head(loan)
summary(loan)
str(loan)
loan$AGE <- as.factor(loan$AGE)
str(loan)
names(loan)

"Creating Model"
model1 <- glm(DEFAULTER~., family=binomial, data=loan)
summary(model1)

"Global Testing for the acceptance of the model"
null <- glm(DEFAULTER~1, family=binomial, data=loan)
anova(null, model1, test="Chisq")

"Predicting the Probabilities"
loan$predprob <- round(fitted(model1),2)
head(loan)

"Classification and Misclassification Analysis"
library(gmodels)
table(loan$DEFAULTER, fitted(model1)>0.5)
sens <- 95/(88+95)*100
sens
spc <- 478/(478+39)*100
spc

"Check Trade Off between sensitivity and specificity using different cut off values"
table(loan$DEFAULTER, fitted(model1)>0.1)
table(loan$DEFAULTER, fitted(model1)>0.2)
table(loan$DEFAULTER, fitted(model1)>0.3)
table(loan$DEFAULTER, fitted(model1)>0.4)
table(loan$DEFAULTER, fitted(model1)>0.5)

"Goodness of fit using receiver Operating Curve"
```

```
pred <- predict(model1, loan, type="response")
install.packages("ROCR")
library(ROCR)

rocrpred <- prediction(pred,loan$DEFAULTER)
rocrperf <- performance(rocrpred, "tpr", "fpr")

"To Check Proper Cut Off Point"
plot(rocrperf, colorize=TRUE, print.cutoffs.at=seq(0.1, by=0.1))

"To Check Coefficients"
coef(model1)
exp(coef(model1))

"As credit to debt ratio of person increases by 1 unit,”
“Odds of the event increases by 77%"
"Model validation same as Linear Regression"
"Variable selection same as Linear Regression"
```

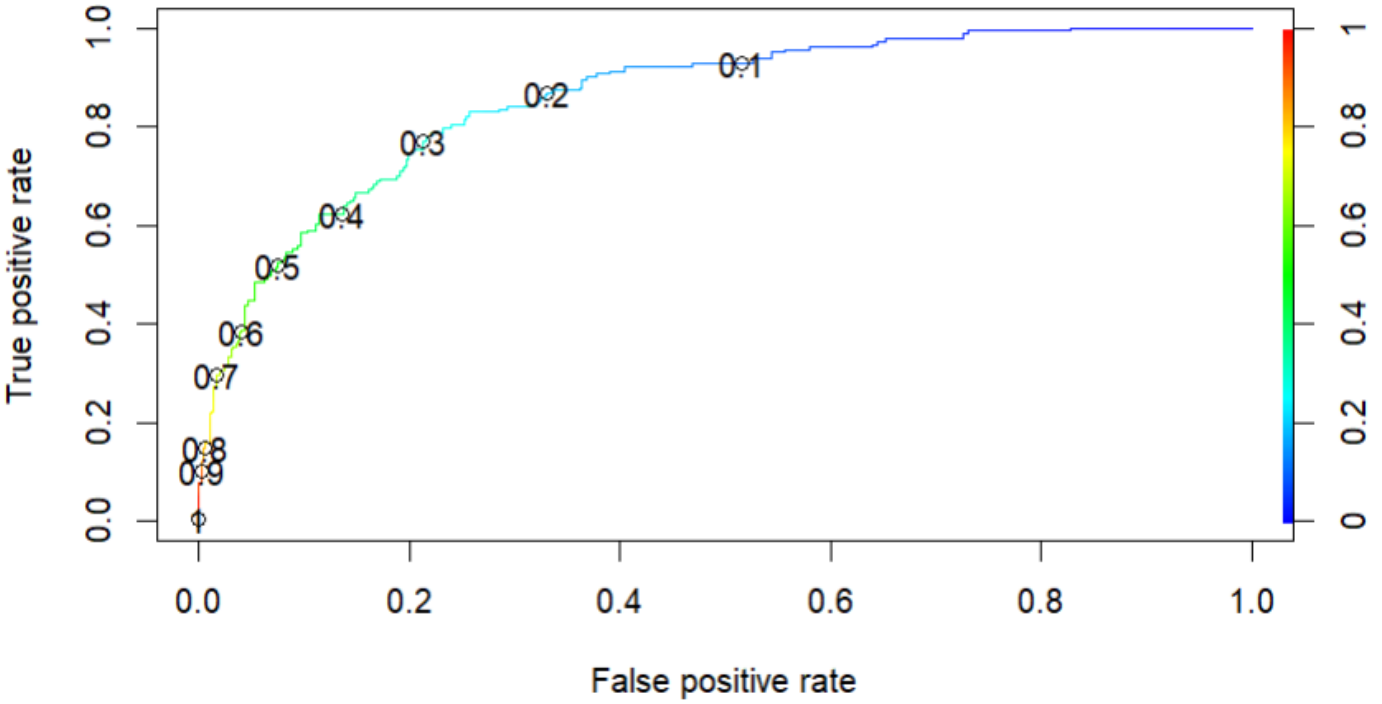
OUTPUT -

```
> loan <- read.csv(file.choose(), header=TRUE, sep=",")
> head(loan)
  SN AGE EMPLOY ADDRESS DEBTINC CREDDEBT OTHDEBT DEFAULTER
1  1   3     17      12     9.3    11.36    5.01         1
2  2   1     10       6    17.3     1.36    4.00         0
3  3   2     15      14     5.5     0.86    2.17         0
4  4   3     15      14     2.9     2.66    0.82         0
5  5   1      2       0    17.3     1.79    3.06         1
6  6   3      5       5    10.2     0.39    2.16         0
> summary(loan)
       SN              AGE            EMPLOY          ADDRESS          DEBTINC          CREDDEBT
 Min.   :  1.0   Min.   :1.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.40   Min.   : 0.010
 1st Qu.:175.8   1st Qu.:1.000   1st Qu.: 3.000   1st Qu.: 3.000   1st Qu.: 5.00   1st Qu.: 0.370
 Median :350.5   Median :2.000   Median : 7.000   Median : 7.000   Median : 8.60   Median : 0.855
 Mean   :350.5   Mean   :1.903   Mean   : 8.389   Mean   : 8.279   Mean   :10.26   Mean   : 1.553
 3rd Qu.:525.2   3rd Qu.:2.000   3rd Qu.:12.000   3rd Qu.:12.000   3rd Qu.:14.12   3rd Qu.: 1.905
 Max.   :700.0   Max.   :3.000   Max.   :31.000   Max.   :34.000   Max.   :41.30   Max.   :20.560
    OTHDEBT         DEFAULTER
 Min.   : 0.050   Min.   :0.0000
 1st Qu.: 1.048   1st Qu.:0.0000
 Median : 1.985   Median :0.0000
 Mean   : 3.058   Mean   :0.2614
 3rd Qu.: 3.928   3rd Qu.:1.0000
 Max.   :27.030   Max.   :1.0000
> str(loan)
'data.frame':   700 obs. of  8 variables:
```

```
 $ SN       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ AGE      : int  3 1 2 3 1 3 2 3 1 2 ...
 $ EMPLOY   : int  17 10 15 15 2 5 20 12 3 0 ...
 $ ADDRESS  : int  12 6 14 14 0 5 9 11 4 13 ...
 $ DEBTINC  : num  9.3 17.3 5.5 2.9 17.3 10.2 30.6 3.6 24.4 19.7 ...
 $ CREDDEBT : num  11.36 1.36 0.86 2.66 1.79 ...
 $ OTHDEBT  : num  5.01 4 2.17 0.82 3.06 ...
 $ DEFAULTER: int  1 0 0 0 1 0 0 0 1 0 ...
> loan$AGE <- as.factor(loan$AGE)
> str(loan)
'data.frame':   700 obs. of  8 variables:
 $ SN       : int  1 2 3 4 5 6 7 8 9 10 ...
 $ AGE      : Factor w/ 3 levels "1","2","3": 3 1 2 3 1 3 2 3 1 2 ...
 $ EMPLOY   : int  17 10 15 15 2 5 20 12 3 0 ...
 $ ADDRESS  : int  12 6 14 14 0 5 9 11 4 13 ...
 $ DEBTINC  : num  9.3 17.3 5.5 2.9 17.3 10.2 30.6 3.6 24.4 19.7 ...
 $ CREDDEBT : num  11.36 1.36 0.86 2.66 1.79 ...
 $ OTHDEBT  : num  5.01 4 2.17 0.82 3.06 ...
 $ DEFAULTER: int  1 0 0 0 1 0 0 0 1 0 ...
> names(loan)
[1] "SN"        "AGE"       "EMPLOY"    "ADDRESS"   "DEBTINC"   "CREDDEBT"  "OTHDEBT"   "DEFAULTER"
> "creating model"
[1] "creating model"
> model1 <- glm(DEFAULTER~., family=binomial, data=loan)
> summary(model1)

Call:
glm(formula = DEFAULTER ~ ., family = binomial, data = loan)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2903  -0.6562  -0.3092   0.2481   2.8942
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.9571221  0.3267254  -2.929 0.003396 **
SN           0.0004689  0.0005275   0.889 0.374064
AGE2         0.2523596  0.2667267   0.946 0.344080
AGE3         0.6089838  0.3612509   1.686 0.091841 .
EMPLOY      -0.2607294  0.0318825  -8.178 2.89e-16 ***
ADDRESS     -0.0995857  0.0223934  -4.447 8.70e-06 ***
DEBTINC      0.0857756  0.0221648   3.870 0.000109 ***
CREDDEBT     0.5618315  0.0885848   6.342 2.26e-10 ***
OTHDEBT      0.0212219  0.0570848   0.372 0.710071
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 804.36  on 699  degrees of freedom
Residual deviance: 552.62  on 691  degrees of freedom
AIC: 570.62
```

```
Number of Fisher Scoring iterations: 6

> "global testing for the acceptance of the model"
[1] "global testing for the acceptance of the model"
> null <- glm(DEFAULTER~1, family=binomial, data=loan)
> anova(null, model1, test="Chisq")
Analysis of Deviance Table

Model 1: DEFAULTER ~ 1
Model 2: DEFAULTER ~ SN + AGE + EMPLOY + ADDRESS + DEBTINC + CREDDEBT +
    OTHDEBT
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       699     804.36
2       691     552.62  8   251.75 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> "predicting the probilities"
[1] "predicting the probilities"
> loan$predprob <- round(fitted(model1),2)
> head(loan)
  SN AGE EMPLOY ADDRESS DEBTINC CREDDEBT OTHDEBT DEFAULTER predprob
1  1   3     17      12     9.3    11.36    5.01         1     0.79
2  2   1     10       6    17.3     1.36    4.00         0     0.14
3  3   2     15      14     5.5     0.86    2.17         0     0.01
4  4   3     15      14     2.9     2.66    0.82         0     0.02
5  5   1      2       0    17.3     1.79    3.06         1     0.75
6  6   3      5       5    10.2     0.39    2.16         0     0.27
> "classification and misclassification analysis "
[1] "classification and misclassification analysis "
> library(gmodels)
> table(loan$DEFAULTER, fitted(model1)>0.5)

    FALSE TRUE
  0   478   39
  1    88   95
> sens <- 95/(88+95)*100
> sens
[1] 51.91257
> spc <- 478/(478+39)*100
```

```
> spc
[1] 92.45648
> "check the trade off between sensivity and specificity using different cut off values"
[1] "check the trade off between sensivity and specificity using different cut off values"
> table(loan$DEFAULTER, fitted(model1)>0.1)

    FALSE TRUE
  0   250  267
  1    13  170
> table(loan$DEFAULTER, fitted(model1)>0.2)

    FALSE TRUE
  0   346  171
  1    25  158
> table(loan$DEFAULTER, fitted(model1)>0.3)

    FALSE TRUE
  0   407  110
  1    43  140
> table(loan$DEFAULTER, fitted(model1)>0.4)

    FALSE TRUE
  0   448   69
  1    69  114
> table(loan$DEFAULTER, fitted(model1)>0.5)

    FALSE TRUE
  0   478   39
  1    88   95
> "goodness of fit using receiver Operational Curve "
[1] "goodness of fit using receiver Operational Curve "
> pred <- predict(model1, loan, type="response")
> install.packages("ROCR")

package 'ROCR' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\Kunal\AppData\Local\Temp\RtmpmKwA7v\downloaded_packages
> library(ROCR)
> rocrpred <- prediction(pred,loan$DEFAULTER)
> rocrperf <- performance(rocrpred, "tpr", "fpr")
> "to check proper cut off point"
[1] "to check proper cut off point"
> plot(rocrperf, colorize=TRUE, print.cutoffs.at=seq(0.1, by=0.1))
```




```
> "to check coeficients"
[1] "to check coeficients"
> coef(model1)
  (Intercept)            SN          AGE2          AGE3        EMPLOY       ADDRESS
DEBTINC
-0.9571221261  0.0004689067  0.2523595660  0.6089837712 -0.2607293672 -0.0995856709
0.0857755990
     CREDDEBT       OTHDEBT
 0.5618315081  0.0212219271
> exp(coef(model1))
(Intercept)          SN        AGE2        AGE3      EMPLOY     ADDRESS     DEBTINC
CREDDEBT     OTHDEBT
  0.3839964   1.0004690   1.2870587   1.8385620   0.7704894   0.9052124   1.0895618
1.7538818   1.0214487
> "as credit to debit ratio of person incerases by 1 unit ,odds of the event increses
by 77%"
[1] "as credit to debit ratio of person incerases by 1 unit ,odds of the event incres
es by 77%"
> "model validation same as linear regression"
[1] "model validation same as linear regression"
> "variable selection same as linear regresiion"
[1] "variable selection same as linear regresiion"
>
```