## Predicting Bank Loan Defaults:Logistic Regression in SAS

/\* Create your own library in SAS like here it is libref and mention the path \*/

libname libref “/home/aro1260/deep”;

/\* Importing loan default dataset in your assigned library \*/

PROC IMPORT DATAFILE= “/home/aroragaurav1260/data/loan\_default.csv”

DBMS=CSV Replace

OUT=libref.loan\_default;

GETNAMES=YES;

RUN;

/\*Checking the no of missing values in numeric var in data \*/

proc means data = libref.loan\_default NMISS N ;

run;

/\*Looking at the contents of the data \*/

PROC CONTENTS DATA=libref.loan\_default;

RUN;

/\*let’s look at some descriptive statistics of the data\*/

proc means data = libref.loan\_default;

var Term Saving\_amount ;

run;

/\* Applying Proc freq to see the frequency of the data \*/

proc freq data = libref.loan\_default ;

tables Emp\_status Default Default \* Emp\_status;

run;

/\* Applying proc univariate to get more detailed summary of the data \*/

proc univariate data = libref.loan\_default;

var Saving\_amount;

histogram Saving\_amount/normal;

run;

/\* Applying proc corr to find out the correlation between Variables in the data \*/

proc corr data = libref.loan\_default;

var Default Term;

run;

/\*Building logistic regression model on full data \*/

**Program1:**

proc logistic data = libref.loan\_default descending ;

class Gender Marital\_status Emp\_status

/ param=effect ref=first;

model default = Checking\_amount Term Credit\_score Car\_loan Personal\_loan Home\_loan Education\_loan Amount Saving\_amount Emp\_duration Gender Marital\_status Age No\_of\_credit\_acc Emp\_status / link=logit;

score out = Logistic\_result ;

run;

/\*Rebuilding logistic model on full data after knocking insignificant variables from the data \*/

**Program1.1**

proc logistic data = libref.loan\_default descending;

model default = Checking\_amount Term Credit\_score

Saving\_amount Age/ link=logit;

score out = Logistic\_result ;

run;

/\*splitting datasets into training (70%) and testing (30%) \*/

proc surveyselect data= libref.loan\_default

method=srs seed=2 outall

samprate=0.7 out=libref.credit\_subset;

proc print data=libref.credit\_subset;

run;

/\* Values of selected variable 1 means selected for training, 0 means selected for testing data set \*/

data libref.training;

set libref.credit\_subset;

if selected=1;

proc print;

data libref.testing;

set libref.credit\_subset;

if selected=0;

proc print;

/\*apply proc freq in training and testing data to Check for balanced data for target var \*/

proc freq data = libref.training;

tables Default;

run;

proc freq data = libref.testing;

tables Default;

run;

/\*Building logistic regression model on training data set \*/

**Program2**

ODS GRAPHICS ON;

proc logistic data = libref.training

descending PLOTS (ONLY) = ROC ;

class Gender Marital\_status Emp\_status

/param=effect ref=first;

model default = Checking\_amount Term Credit\_score Car\_loan Personal\_loan Home\_loan Education\_loan Amount Saving\_amount Emp\_duration Gender Marital\_status Age No\_of\_credit\_acc Emp\_status / link=logit;

score data = libref.testing out=WORK.Logistic\_output;

run;

ODS GRAPHICS OFF;

/\* Building logistic model with significant variables \*/

**Program2.1**

ODS GRAPHICS ON;

proc logistic data = libref.training

descending PLOTS (ONLY) = ROC ;

class Emp\_status/ param=ref ref=first;

model default = Checking\_amount Term Credit\_score Emp\_status

Amount Saving\_amount Age/ link=logit ;

score data = libref.testing out=WORK.Logistic\_output;

run;

ODS GRAPHICS OFF;

/\* Rebuilding Final Logistic Model with only significant variables \*/

**Program2.2**

ODS GRAPHICS ON;

proc logistic data = libref.training

descending PLOTS (ONLY) = ROC ;

class Emp\_status/ param=ref ref=first;

model default = Checking\_amount Term Credit\_score Emp\_status

Saving\_amount Age/ link=logit ;

score data = libref.testing out=WORK.Logistic\_output;

run;

ODS GRAPHICS OFF;