scm651\_hw4.R

tamtam

2020-12-09

library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

library(tidyr)  
library(car)

## Loading required package: carData

bankData<-read.csv("~/OneDrive - Syracuse University/651 Business Analytics/HW4/Homework 4 Data Set - Universal Bank.csv")  
  
summary(bankData)

## CustomerID PersonalLoan Age Experience Income   
## Min. : 1 Min. :0.000 Min. :23.00 Min. :-3.0 Min. : 8.00   
## 1st Qu.:1251 1st Qu.:0.000 1st Qu.:35.00 1st Qu.:10.0 1st Qu.: 39.00   
## Median :2500 Median :0.000 Median :45.00 Median :20.0 Median : 64.00   
## Mean :2500 Mean :0.096 Mean :45.34 Mean :20.1 Mean : 73.77   
## 3rd Qu.:3750 3rd Qu.:0.000 3rd Qu.:55.00 3rd Qu.:30.0 3rd Qu.: 98.00   
## Max. :5000 Max. :1.000 Max. :67.00 Max. :43.0 Max. :224.00   
## ZIP.Code Family CCAvg Education   
## Min. : 9307 Min. :1.000 Min. : 0.000 Min. :1.000   
## 1st Qu.:91911 1st Qu.:1.000 1st Qu.: 0.700 1st Qu.:1.000   
## Median :93437 Median :2.000 Median : 1.500 Median :2.000   
## Mean :93152 Mean :2.396 Mean : 1.938 Mean :1.881   
## 3rd Qu.:94608 3rd Qu.:3.000 3rd Qu.: 2.500 3rd Qu.:3.000   
## Max. :96651 Max. :4.000 Max. :10.000 Max. :3.000   
## Mortgage SecuritiesAccount CDAccount Online   
## Min. : 0.0 Min. :0.0000 Min. :0.0000 Min. :0.0000   
## 1st Qu.: 0.0 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000   
## Median : 0.0 Median :0.0000 Median :0.0000 Median :1.0000   
## Mean : 56.5 Mean :0.1044 Mean :0.0604 Mean :0.5968   
## 3rd Qu.:101.0 3rd Qu.:0.0000 3rd Qu.:0.0000 3rd Qu.:1.0000   
## Max. :635.0 Max. :1.0000 Max. :1.0000 Max. :1.0000   
## CreditCard   
## Min. :0.000   
## 1st Qu.:0.000   
## Median :0.000   
## Mean :0.294   
## 3rd Qu.:1.000   
## Max. :1.000

model.logit<-glm(formula = PersonalLoan ~ .,data=bankData,family=binomial(logit))  
summary(model.logit)

##   
## Call:  
## glm(formula = PersonalLoan ~ ., family = binomial(logit), data = bankData)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1334 -0.2014 -0.0800 -0.0307 3.9183   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -1.172e+01 4.115e+00 -2.847 0.004409 \*\*   
## CustomerID -5.231e-05 5.142e-05 -1.017 0.309040   
## Age -5.382e-02 6.136e-02 -0.877 0.380388   
## Experience 6.393e-02 6.098e-02 1.048 0.294502   
## Income 5.466e-02 2.625e-03 20.820 < 2e-16 \*\*\*  
## ZIP.Code -3.745e-06 4.072e-05 -0.092 0.926717   
## Family 6.952e-01 7.432e-02 9.353 < 2e-16 \*\*\*  
## CCAvg 1.218e-01 3.968e-02 3.070 0.002142 \*\*   
## Education 1.740e+00 1.153e-01 15.098 < 2e-16 \*\*\*  
## Mortgage 4.639e-04 5.549e-04 0.836 0.403228   
## SecuritiesAccount -9.453e-01 2.860e-01 -3.305 0.000951 \*\*\*  
## CDAccount 3.823e+00 3.242e-01 11.792 < 2e-16 \*\*\*  
## Online -6.717e-01 1.572e-01 -4.272 1.94e-05 \*\*\*  
## CreditCard -1.114e+00 2.051e-01 -5.430 5.62e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3162.0 on 4999 degrees of freedom  
## Residual deviance: 1283.3 on 4986 degrees of freedom  
## AIC: 1311.3  
##   
## Number of Fisher Scoring iterations: 8

vif(model.logit,bankData)

## CustomerID Age Experience Income   
## 1.008579 91.636420 91.502950 2.436080   
## ZIP.Code Family CCAvg Education   
## 1.005893 1.316732 1.442160 1.913245   
## Mortgage SecuritiesAccount CDAccount Online   
## 1.037874 1.364434 2.069393 1.130820   
## CreditCard   
## 1.400506

model.probit<-glm(formula = PersonalLoan ~ .,data=bankData,family=binomial(probit))  
summary(model.probit)

##   
## Call:  
## glm(formula = PersonalLoan ~ ., family = binomial(probit), data = bankData)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.2281 -0.2079 -0.0526 -0.0087 4.4415   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -5.537e+00 2.079e+00 -2.664 0.007721 \*\*   
## CustomerID -4.033e-05 2.671e-05 -1.510 0.131090   
## Age -3.047e-02 3.133e-02 -0.973 0.330764   
## Experience 3.382e-02 3.118e-02 1.085 0.278095   
## Income 2.783e-02 1.275e-03 21.824 < 2e-16 \*\*\*  
## ZIP.Code -4.777e-06 2.056e-05 -0.232 0.816295   
## Family 3.413e-01 3.757e-02 9.084 < 2e-16 \*\*\*  
## CCAvg 7.300e-02 2.096e-02 3.483 0.000495 \*\*\*  
## Education 8.562e-01 5.694e-02 15.036 < 2e-16 \*\*\*  
## Mortgage 2.151e-04 2.953e-04 0.728 0.466367   
## SecuritiesAccount -5.035e-01 1.473e-01 -3.419 0.000629 \*\*\*  
## CDAccount 2.006e+00 1.650e-01 12.157 < 2e-16 \*\*\*  
## Online -3.492e-01 8.119e-02 -4.300 1.70e-05 \*\*\*  
## CreditCard -5.779e-01 1.046e-01 -5.527 3.27e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3162.0 on 4999 degrees of freedom  
## Residual deviance: 1300.9 on 4986 degrees of freedom  
## AIC: 1328.9  
##   
## Number of Fisher Scoring iterations: 8

vif(model.probit,bankData)

## CustomerID Age Experience Income   
## 1.006506 88.221350 88.136703 2.156826   
## ZIP.Code Family CCAvg Education   
## 1.004307 1.253515 1.439103 1.711942   
## Mortgage SecuritiesAccount CDAccount Online   
## 1.032420 1.329893 1.906736 1.116457   
## CreditCard   
## 1.352404

model.logit\_2<-glm(formula = PersonalLoan ~ .^2,data=bankData,family=binomial(logit))

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(model.logit\_2)

##   
## Call:  
## glm(formula = PersonalLoan ~ .^2, family = binomial(logit), data = bankData)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5176 -0.0555 -0.0042 -0.0001 4.9753   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.014e+02 2.058e+02 -1.464 0.14314   
## CustomerID 6.238e-03 5.658e-03 1.102 0.27030   
## Age 1.316e+01 8.091e+00 1.627 0.10376   
## Experience -1.271e+01 8.099e+00 -1.569 0.11671   
## Income -2.531e-01 2.638e-01 -0.960 0.33722   
## ZIP.Code 3.062e-03 2.184e-03 1.402 0.16081   
## Family 2.875e+00 7.840e+00 0.367 0.71378   
## CCAvg 3.170e-01 4.827e+00 0.066 0.94763   
## Education -1.562e+01 1.070e+01 -1.460 0.14423   
## Mortgage -2.568e-02 7.021e-02 -0.366 0.71458   
## SecuritiesAccount 2.374e+01 4.201e+01 0.565 0.57199   
## CDAccount 1.316e+01 4.015e+01 0.328 0.74306   
## Online -1.800e+01 1.899e+01 -0.948 0.34323   
## CreditCard -2.912e+01 2.549e+01 -1.143 0.25320   
## CustomerID:Age 1.070e-05 1.053e-04 0.102 0.91904   
## CustomerID:Experience -2.278e-05 1.056e-04 -0.216 0.82920   
## CustomerID:Income 5.770e-06 3.041e-06 1.897 0.05777 .   
## CustomerID:ZIP.Code -7.769e-08 5.464e-08 -1.422 0.15510   
## CustomerID:Family 2.658e-05 9.180e-05 0.290 0.77212   
## CustomerID:CCAvg -2.628e-05 5.610e-05 -0.469 0.63943   
## CustomerID:Education 1.877e-04 1.243e-04 1.510 0.13097   
## CustomerID:Mortgage -6.283e-07 8.494e-07 -0.740 0.45949   
## CustomerID:SecuritiesAccount 8.531e-04 4.639e-04 1.839 0.06593 .   
## CustomerID:CDAccount -4.054e-04 4.411e-04 -0.919 0.35807   
## CustomerID:Online -1.133e-05 2.249e-04 -0.050 0.95981   
## CustomerID:CreditCard 9.868e-05 2.966e-04 0.333 0.73935   
## Age:Experience -2.337e-03 1.177e-03 -1.986 0.04702 \*   
## Age:Income 2.160e-03 5.367e-03 0.402 0.68733   
## Age:ZIP.Code -1.305e-04 8.580e-05 -1.521 0.12815   
## Age:Family -2.122e-01 1.538e-01 -1.379 0.16776   
## Age:CCAvg -4.436e-02 8.168e-02 -0.543 0.58705   
## Age:Education -5.739e-01 1.757e-01 -3.267 0.00109 \*\*   
## Age:Mortgage 9.310e-04 1.275e-03 0.730 0.46540   
## Age:SecuritiesAccount -2.663e-01 6.063e-01 -0.439 0.66054   
## Age:CDAccount -1.691e-01 6.410e-01 -0.264 0.79196   
## Age:Online 2.423e-01 3.336e-01 0.726 0.46773   
## Age:CreditCard 2.245e-01 4.373e-01 0.513 0.60772   
## Experience:Income -1.617e-03 5.353e-03 -0.302 0.76263   
## Experience:ZIP.Code 1.266e-04 8.591e-05 1.474 0.14052   
## Experience:Family 1.921e-01 1.541e-01 1.246 0.21264   
## Experience:CCAvg 4.180e-02 8.135e-02 0.514 0.60735   
## Experience:Education 6.196e-01 1.772e-01 3.497 0.00047 \*\*\*  
## Experience:Mortgage -8.201e-04 1.272e-03 -0.645 0.51914   
## Experience:SecuritiesAccount 1.251e-01 5.899e-01 0.212 0.83202   
## Experience:CDAccount 3.134e-01 6.359e-01 0.493 0.62214   
## Experience:Online -2.609e-01 3.350e-01 -0.779 0.43614   
## Experience:CreditCard -2.371e-01 4.369e-01 -0.543 0.58728   
## Income:ZIP.Code 2.521e-07 2.527e-06 0.100 0.92053   
## Income:Family 4.971e-02 6.077e-03 8.179 2.87e-16 \*\*\*  
## Income:CCAvg -1.583e-02 2.505e-03 -6.320 2.61e-10 \*\*\*  
## Income:Education 1.077e-01 1.096e-02 9.825 < 2e-16 \*\*\*  
## Income:Mortgage 1.028e-05 3.995e-05 0.257 0.79700   
## Income:SecuritiesAccount 2.760e-02 3.209e-02 0.860 0.38963   
## Income:CDAccount 9.678e-03 2.240e-02 0.432 0.66573   
## Income:Online 1.010e-02 1.011e-02 0.999 0.31795   
## Income:CreditCard -3.580e-03 1.361e-02 -0.263 0.79254   
## ZIP.Code:Family -6.816e-06 7.044e-05 -0.097 0.92292   
## ZIP.Code:CCAvg 2.311e-05 4.593e-05 0.503 0.61482   
## ZIP.Code:Education 2.088e-04 1.002e-04 2.085 0.03711 \*   
## ZIP.Code:Mortgage -9.282e-08 6.393e-07 -0.145 0.88457   
## ZIP.Code:SecuritiesAccount -2.246e-04 3.872e-04 -0.580 0.56189   
## ZIP.Code:CDAccount -9.004e-05 3.759e-04 -0.240 0.81068   
## ZIP.Code:Online 1.420e-04 1.762e-04 0.806 0.42012   
## ZIP.Code:CreditCard 2.597e-04 2.338e-04 1.111 0.26673   
## Family:CCAvg 2.109e-01 9.608e-02 2.195 0.02813 \*   
## Family:Education -7.490e-01 1.676e-01 -4.469 7.86e-06 \*\*\*  
## Family:Mortgage 2.857e-03 1.291e-03 2.214 0.02686 \*   
## Family:SecuritiesAccount -1.882e+00 7.861e-01 -2.394 0.01666 \*   
## Family:CDAccount 1.181e+00 6.977e-01 1.693 0.09043 .   
## Family:Online -3.417e-01 3.002e-01 -1.138 0.25500   
## Family:CreditCard 4.268e-02 3.851e-01 0.111 0.91176   
## CCAvg:Education 6.078e-01 1.288e-01 4.721 2.35e-06 \*\*\*  
## CCAvg:Mortgage 4.156e-04 6.194e-04 0.671 0.50225   
## CCAvg:SecuritiesAccount 3.060e-01 4.329e-01 0.707 0.47956   
## CCAvg:CDAccount -4.847e-01 3.665e-01 -1.323 0.18599   
## CCAvg:Online -2.892e-01 1.789e-01 -1.617 0.10590   
## CCAvg:CreditCard 1.839e-01 2.388e-01 0.770 0.44136   
## Education:Mortgage -1.856e-04 1.622e-03 -0.114 0.90893   
## Education:SecuritiesAccount 7.217e-01 1.006e+00 0.717 0.47309   
## Education:CDAccount 3.024e-01 9.230e-01 0.328 0.74316   
## Education:Online -2.435e-01 4.100e-01 -0.594 0.55265   
## Education:CreditCard -5.167e-01 5.348e-01 -0.966 0.33396   
## Mortgage:SecuritiesAccount 9.240e-03 5.927e-03 1.559 0.11903   
## Mortgage:CDAccount -3.430e-03 6.373e-03 -0.538 0.59040   
## Mortgage:Online 1.006e-03 2.629e-03 0.383 0.70195   
## Mortgage:CreditCard -1.582e-03 4.111e-03 -0.385 0.70038   
## SecuritiesAccount:CDAccount 5.099e+00 3.687e+00 1.383 0.16662   
## SecuritiesAccount:Online -4.140e+00 2.536e+00 -1.632 0.10260   
## SecuritiesAccount:CreditCard -1.128e+00 2.605e+00 -0.433 0.66512   
## CDAccount:Online 2.529e+00 2.908e+00 0.870 0.38440   
## CDAccount:CreditCard -2.007e+00 3.210e+00 -0.625 0.53196   
## Online:CreditCard -3.988e+00 1.594e+00 -2.501 0.01238 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3162.04 on 4999 degrees of freedom  
## Residual deviance: 472.35 on 4908 degrees of freedom  
## AIC: 656.35  
##   
## Number of Fisher Scoring iterations: 11

model.logit\_significant<-glm(formula = PersonalLoan ~ Age+Income+Family+CCAvg+Education+SecuritiesAccount+CDAccount+Online+CreditCard,data=bankData,family=binomial(logit))  
  
summary(model.logit\_significant)

##   
## Call:  
## glm(formula = PersonalLoan ~ Age + Income + Family + CCAvg +   
## Education + SecuritiesAccount + CDAccount + Online + CreditCard,   
## family = binomial(logit), data = bankData)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -3.1639 -0.2033 -0.0795 -0.0308 3.9351   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -13.757158 0.667122 -20.622 < 2e-16 \*\*\*  
## Age 0.010193 0.006492 1.570 0.11637   
## Income 0.054956 0.002602 21.121 < 2e-16 \*\*\*  
## Family 0.697989 0.074350 9.388 < 2e-16 \*\*\*  
## CCAvg 0.120317 0.039433 3.051 0.00228 \*\*   
## Education 1.710092 0.112884 15.149 < 2e-16 \*\*\*  
## SecuritiesAccount -0.936061 0.285341 -3.280 0.00104 \*\*   
## CDAccount 3.842066 0.323607 11.873 < 2e-16 \*\*\*  
## Online -0.672729 0.156881 -4.288 1.80e-05 \*\*\*  
## CreditCard -1.121904 0.204916 -5.475 4.38e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3162.0 on 4999 degrees of freedom  
## Residual deviance: 1286.1 on 4990 degrees of freedom  
## AIC: 1306.1  
##   
## Number of Fisher Scoring iterations: 8

#Second Question  
formula\_sig <- PersonalLoan ~ Age+Income+Family+CCAvg+Education+SecuritiesAccount+CDAccount+Online+CreditCard  
model.logit\_significant\_2<-glm(formula = PersonalLoan ~ (Age+Income+Family+CCAvg+Education+SecuritiesAccount+CDAccount+Online+CreditCard)^2,data=bankData,family=binomial(logit))

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(model.logit\_significant\_2)

##   
## Call:  
## glm(formula = PersonalLoan ~ (Age + Income + Family + CCAvg +   
## Education + SecuritiesAccount + CDAccount + Online + CreditCard)^2,   
## family = binomial(logit), data = bankData)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2459 -0.0760 -0.0069 -0.0001 5.1254   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 8.343e+00 3.661e+00 2.279 0.02265 \*   
## Age -5.577e-02 6.892e-02 -0.809 0.41838   
## Income -1.682e-01 2.575e-02 -6.531 6.54e-11 \*\*\*  
## Family -2.381e+00 7.340e-01 -3.244 0.00118 \*\*   
## CCAvg 1.406e+00 5.099e-01 2.758 0.00582 \*\*   
## Education -1.005e+01 1.362e+00 -7.381 1.57e-13 \*\*\*  
## SecuritiesAccount 1.399e+00 4.486e+00 0.312 0.75513   
## CDAccount -3.874e+00 5.621e+00 -0.689 0.49069   
## Online 1.565e+00 1.916e+00 0.817 0.41399   
## CreditCard 1.964e+00 2.298e+00 0.855 0.39269   
## Age:Income 4.747e-04 4.257e-04 1.115 0.26485   
## Age:Family -9.874e-03 9.804e-03 -1.007 0.31385   
## Age:CCAvg -6.171e-03 6.889e-03 -0.896 0.37031   
## Age:Education 2.671e-02 1.299e-02 2.056 0.03978 \*   
## Age:SecuritiesAccount -1.026e-01 5.988e-02 -1.713 0.08673 .   
## Age:CDAccount 1.105e-01 5.589e-02 1.977 0.04804 \*   
## Age:Online -1.097e-02 2.504e-02 -0.438 0.66126   
## Age:CreditCard -1.518e-02 3.016e-02 -0.503 0.61469   
## Income:Family 4.629e-02 5.449e-03 8.495 < 2e-16 \*\*\*  
## Income:CCAvg -1.390e-02 2.209e-03 -6.292 3.13e-10 \*\*\*  
## Income:Education 1.041e-01 9.981e-03 10.432 < 2e-16 \*\*\*  
## Income:SecuritiesAccount 3.105e-02 2.757e-02 1.127 0.25994   
## Income:CDAccount 6.296e-04 2.060e-02 0.031 0.97562   
## Income:Online 6.684e-03 9.157e-03 0.730 0.46540   
## Income:CreditCard -4.191e-03 1.231e-02 -0.340 0.73353   
## Family:CCAvg 1.954e-01 8.496e-02 2.299 0.02149 \*   
## Family:Education -8.339e-01 1.501e-01 -5.555 2.77e-08 \*\*\*  
## Family:SecuritiesAccount -1.322e+00 6.906e-01 -1.914 0.05563 .   
## Family:CDAccount 1.008e+00 6.294e-01 1.602 0.10907   
## Family:Online -3.697e-01 2.780e-01 -1.330 0.18363   
## Family:CreditCard -2.158e-01 3.398e-01 -0.635 0.52544   
## CCAvg:Education 4.960e-01 1.076e-01 4.612 4.00e-06 \*\*\*  
## CCAvg:SecuritiesAccount 3.806e-02 3.457e-01 0.110 0.91235   
## CCAvg:CDAccount -2.572e-01 3.335e-01 -0.771 0.44052   
## CCAvg:Online -2.268e-01 1.570e-01 -1.444 0.14869   
## CCAvg:CreditCard 1.825e-01 2.084e-01 0.876 0.38107   
## Education:SecuritiesAccount 3.517e-01 7.815e-01 0.450 0.65275   
## Education:CDAccount 3.169e-01 7.810e-01 0.406 0.68492   
## Education:Online -1.516e-01 3.550e-01 -0.427 0.66939   
## Education:CreditCard -5.282e-01 4.449e-01 -1.187 0.23508   
## SecuritiesAccount:CDAccount 4.830e+00 3.492e+00 1.383 0.16664   
## SecuritiesAccount:Online -3.575e+00 2.413e+00 -1.482 0.13836   
## SecuritiesAccount:CreditCard -2.437e+00 2.260e+00 -1.078 0.28088   
## CDAccount:Online 2.256e+00 2.774e+00 0.813 0.41605   
## CDAccount:CreditCard 5.049e-01 2.686e+00 0.188 0.85091   
## Online:CreditCard -4.139e+00 1.643e+00 -2.519 0.01176 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3162.04 on 4999 degrees of freedom  
## Residual deviance: 527.21 on 4954 degrees of freedom  
## AIC: 619.21  
##   
## Number of Fisher Scoring iterations: 10

#3rd Question   
interaction\_vars<- c("Income","Family","CCAvg","Education","Income:Family","Income:CCAvg","Income:Education","Family:CCAvg","Family:Education","Family:SecuritiesAccount","Family:CDAccount","CCAvg:Education","Online:CreditCard")  
interaction\_vars\_vif<-c(interaction\_vars,"Age:Education","Age:SecuritiesAccount","Age:CDAccount")  
  
formula\_int<-paste("PersonalLoan", paste(interaction\_vars\_vif, collapse=" + "), sep=" ~ ")  
model.logit\_significant\_interaction<-glm(formula =formula\_int,data=bankData,family=binomial(logit))

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(model.logit\_significant\_interaction)

##   
## Call:  
## glm(formula = formula\_int, family = binomial(logit), data = bankData)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.2222 -0.0913 -0.0102 -0.0002 5.0347   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 6.5985943 1.9349670 3.410 0.000649 \*\*\*  
## Income -0.1425678 0.0170050 -8.384 < 2e-16 \*\*\*  
## Family -2.8728071 0.6080344 -4.725 2.30e-06 \*\*\*  
## CCAvg 0.8841448 0.3781676 2.338 0.019389 \*   
## Education -8.6969967 1.0895529 -7.982 1.44e-15 \*\*\*  
## Income:Family 0.0444700 0.0052230 8.514 < 2e-16 \*\*\*  
## Income:CCAvg -0.0123248 0.0019621 -6.282 3.35e-10 \*\*\*  
## Income:Education 0.1022398 0.0092372 11.068 < 2e-16 \*\*\*  
## Family:CCAvg 0.1831904 0.0801107 2.287 0.022213 \*   
## Family:Education -0.8567226 0.1425634 -6.009 1.86e-09 \*\*\*  
## Family:SecuritiesAccount -0.4760582 0.4604861 -1.034 0.301222   
## Family:CDAccount 0.8021589 0.4315022 1.859 0.063028 .   
## CCAvg:Education 0.4776256 0.0976015 4.894 9.90e-07 \*\*\*  
## Online:CreditCard -4.0474169 0.7391907 -5.475 4.36e-08 \*\*\*  
## Education:Age -0.0006435 0.0050418 -0.128 0.898443   
## SecuritiesAccount:Age -0.0192758 0.0232358 -0.830 0.406781   
## CDAccount:Age 0.0831058 0.0229520 3.621 0.000294 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3162.04 on 4999 degrees of freedom  
## Residual deviance: 554.98 on 4983 degrees of freedom  
## AIC: 588.98  
##   
## Number of Fisher Scoring iterations: 10

#3rd Question 2nd attempt , lower AIC that before , skip it  
interaction\_vars2<- c("Income","Family","CCAvg","Education","Income:Family","Income:CCAvg","Income:Education","Family:CCAvg","Family:Education","CCAvg:Education","Online:CreditCard")  
interaction\_vars2<-c(interaction\_vars2,"Age:CDAccount")  
formula\_int2<-paste("PersonalLoan", paste(interaction\_vars2, collapse=" + "), sep=" ~ ")  
model.logit\_significant\_interaction\_2<-glm(formula =formula\_int2,data=bankData,family=binomial(logit))

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(model.logit\_significant\_interaction\_2)

##   
## Call:  
## glm(formula = formula\_int2, family = binomial(logit), data = bankData)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5172 -0.0902 -0.0125 -0.0002 5.0288   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 6.118731 1.903314 3.215 0.00131 \*\*   
## Income -0.140107 0.016707 -8.386 < 2e-16 \*\*\*  
## Family -2.634872 0.566484 -4.651 3.30e-06 \*\*\*  
## CCAvg 0.828824 0.376668 2.200 0.02778 \*   
## Education -8.721487 1.058377 -8.240 < 2e-16 \*\*\*  
## Income:Family 0.043407 0.004992 8.696 < 2e-16 \*\*\*  
## Income:CCAvg -0.012198 0.001937 -6.298 3.01e-10 \*\*\*  
## Income:Education 0.101839 0.009135 11.149 < 2e-16 \*\*\*  
## Family:CCAvg 0.183251 0.078435 2.336 0.01947 \*   
## Family:Education -0.886670 0.138193 -6.416 1.40e-10 \*\*\*  
## CCAvg:Education 0.504297 0.097114 5.193 2.07e-07 \*\*\*  
## Online:CreditCard -2.917616 0.554268 -5.264 1.41e-07 \*\*\*  
## Age:CDAccount 0.083705 0.009867 8.484 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3162.04 on 4999 degrees of freedom  
## Residual deviance: 573.21 on 4987 degrees of freedom  
## AIC: 599.21  
##   
## Number of Fisher Scoring iterations: 10

model.probit\_significant\_interaction\_2<-glm(formula =formula\_int2,data=bankData,family=binomial(probit))

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(model.probit\_significant\_interaction\_2)

##   
## Call:  
## glm(formula = formula\_int2, family = binomial(probit), data = bankData)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.3866 -0.0753 -0.0018 0.0000 6.0370   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.4342863 0.8858879 2.748 0.00600 \*\*   
## Income -0.0621875 0.0074922 -8.300 < 2e-16 \*\*\*  
## Family -1.2702470 0.2636629 -4.818 1.45e-06 \*\*\*  
## CCAvg 0.3473844 0.1814888 1.914 0.05561 .   
## Education -4.1195633 0.4789197 -8.602 < 2e-16 \*\*\*  
## Income:Family 0.0196499 0.0022025 8.922 < 2e-16 \*\*\*  
## Income:CCAvg -0.0060881 0.0009342 -6.517 7.18e-11 \*\*\*  
## Income:Education 0.0474602 0.0039747 11.941 < 2e-16 \*\*\*  
## Family:CCAvg 0.1166737 0.0387039 3.015 0.00257 \*\*   
## Family:Education -0.3683908 0.0665842 -5.533 3.15e-08 \*\*\*  
## CCAvg:Education 0.2680427 0.0491816 5.450 5.04e-08 \*\*\*  
## Online:CreditCard -1.5184837 0.2821444 -5.382 7.37e-08 \*\*\*  
## Age:CDAccount 0.0442889 0.0051465 8.606 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3162.04 on 4999 degrees of freedom  
## Residual deviance: 593.04 on 4987 degrees of freedom  
## AIC: 619.04  
##   
## Number of Fisher Scoring iterations: 11

#Logit  
#Without VIF  
# Everything - AIC: 1311.3  
# With Significant Vars - AIC: 1306.6  
# Significant Vars + moderating effect - AIC: 611.62  
# Significant Vars + significant moderating effect - AIC: 604.4  
  
#With VIF  
#Significant vars + vif + moderating : AIC: 588.98  
  
#Question 4  
library(neuralnet)  
model\_nn <- neuralnet(formula\_sig,bankData,hidden = 3,lifesign = "minimal", linear.output = FALSE,threshold = 0.1)

## hidden: 3 thresh: 0.1 rep: 1/1 steps:

## 54634 error: 31.00646 time: 35.09 secs

#summary(model\_nn)  
model\_nn$result.matrix

## [,1]  
## error 3.100646e+01  
## reached.threshold 5.645331e-02  
## steps 5.463400e+04  
## Intercept.to.1layhid1 -4.203176e-01  
## Age.to.1layhid1 1.126217e+00  
## Income.to.1layhid1 1.237078e+00  
## Family.to.1layhid1 1.067809e+00  
## CCAvg.to.1layhid1 -5.428843e-01  
## Education.to.1layhid1 -6.259698e-01  
## SecuritiesAccount.to.1layhid1 1.477235e+00  
## CDAccount.to.1layhid1 4.877242e-01  
## Online.to.1layhid1 1.118280e+00  
## CreditCard.to.1layhid1 -1.158957e+00  
## Intercept.to.1layhid2 3.067958e+01  
## Age.to.1layhid2 1.240318e-02  
## Income.to.1layhid2 -7.425630e-03  
## Family.to.1layhid2 -7.926552e+00  
## CCAvg.to.1layhid2 4.432331e-01  
## Education.to.1layhid2 -1.868476e+01  
## SecuritiesAccount.to.1layhid2 4.129906e+00  
## CDAccount.to.1layhid2 -4.504779e+00  
## Online.to.1layhid2 3.921182e+00  
## CreditCard.to.1layhid2 4.438690e+00  
## Intercept.to.1layhid3 6.007292e+00  
## Age.to.1layhid3 -4.195148e-03  
## Income.to.1layhid3 -4.267037e-02  
## Family.to.1layhid3 7.979065e-03  
## CCAvg.to.1layhid3 -2.421117e-01  
## Education.to.1layhid3 8.480188e-02  
## SecuritiesAccount.to.1layhid3 1.446173e-01  
## CDAccount.to.1layhid3 -1.104669e+00  
## Online.to.1layhid3 1.908262e-01  
## CreditCard.to.1layhid3 3.141464e-01  
## Intercept.to.PersonalLoan 6.971519e+00  
## 1layhid1.to.PersonalLoan 7.780281e+00  
## 1layhid2.to.PersonalLoan -2.028349e+03  
## 1layhid3.to.PersonalLoan -2.100582e+01

#plot(model\_nn)  
  
formula\_nn\_sig<-PersonalLoan ~ Age+Income+Family+CCAvg+Education+CDAccount+Online+CreditCard  
model\_nn2 <- neuralnet(formula\_nn\_sig,bankData,hidden = 3,lifesign = "minimal", linear.output = FALSE,threshold = 0.1)

## hidden: 3 thresh: 0.1 rep: 1/1 steps: 39617 error: 29.30171 time: 25.47 secs

#summary(model\_nn)  
plot(model\_nn2)  
model\_nn2$result.matrix

## [,1]  
## error 2.930171e+01  
## reached.threshold 9.388488e-02  
## steps 3.961700e+04  
## Intercept.to.1layhid1 9.370283e+01  
## Age.to.1layhid1 2.978172e-01  
## Income.to.1layhid1 -1.078286e+00  
## Family.to.1layhid1 1.125358e+00  
## CCAvg.to.1layhid1 -8.120338e+00  
## Education.to.1layhid1 2.808691e+00  
## CDAccount.to.1layhid1 -1.994706e+01  
## Online.to.1layhid1 2.680193e+00  
## CreditCard.to.1layhid1 5.944964e+00  
## Intercept.to.1layhid2 1.294589e+02  
## Age.to.1layhid2 1.038823e-01  
## Income.to.1layhid2 -1.125930e+00  
## Family.to.1layhid2 -1.404031e+00  
## CCAvg.to.1layhid2 -2.173786e+00  
## Education.to.1layhid2 2.933736e+00  
## CDAccount.to.1layhid2 -8.366185e+00  
## Online.to.1layhid2 1.673130e+00  
## CreditCard.to.1layhid2 5.708395e+00  
## Intercept.to.1layhid3 4.912575e+01  
## Age.to.1layhid3 -4.327730e-02  
## Income.to.1layhid3 -1.339990e-02  
## Family.to.1layhid3 -9.307213e+00  
## CCAvg.to.1layhid3 -1.289076e-01  
## Education.to.1layhid3 -2.195659e+01  
## CDAccount.to.1layhid3 -7.655566e+00  
## Online.to.1layhid3 1.365832e+00  
## CreditCard.to.1layhid3 1.210919e+00  
## Intercept.to.PersonalLoan 1.907279e+01  
## 1layhid1.to.PersonalLoan -2.263273e+03  
## 1layhid2.to.PersonalLoan -1.950838e+01  
## 1layhid3.to.PersonalLoan -3.009481e+02

library(ggplot2)  
plot <- ggplot(bankData,aes(x=Income,y=Family))+ geom\_point()   
plot  
  
  
plot1 <- ggplot(bankData,aes(x=CCAvg,y=Education,color=Education,size=PersonalLoan,alpha=I(0.1))) + geom\_point()  
plot1  
  
#unique(bankData$Education)  
library(sqldf)

## Loading required package: gsubfn  
## Loading required package: proto  
## Loading required package: RSQLite

groupData<-sqldf("select avg(PersonalLoan) as approval,Education,CCAvg as CCAvg from bankData group by CCAvg,Education ")  
groupData$Education<-as.factor(groupData$Education)  
#library(reshape2)  
#melted<-melt(groupData,id=c("CCAvg"))  
plot2 <- ggplot(groupData,aes(x=CCAvg,y=approval,group=Education,color=Education)) + geom\_smooth()  
plot2

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'