

SCM 651 Business Analytics

Team 4: Frederick B Lamarca III, Maria Ng, and Seth J Rissmiller

1. Perform a logit and probit analysis of the variables that affect whether a customer takes out a loan. Consider only main effects. Which variables are significant? How do the significant variables influence the likelihood of taking out a loan?

Logit

Logit					
Value	Estimate	Std..Error	z.value	Pr...z..	Significance
(Intercept)	-11.82686849	4.119433376	-2.870993995	0.004091833	**
Age	-0.053537867	0.061309898	-0.873233664	0.382535729	
Experience	0.063669466	0.060926887	1.04501425	0.296016383	
Income	0.054581665	0.002620227	20.83089122	2.27E-96	***
ZIP.Code	-3.95E-06	4.08E-05	-0.096830712	0.922860834	
Family	0.695839386	0.074310509	9.363943218	7.68E-21	***
CCAvg	0.124026985	0.039648998	3.128124069	0.001759259	**
Education	1.735968471	0.115076845	15.08529772	2.02E-51	***
Mortgage	0.000473575	0.000554289	0.854383517	0.392892526	
SecuritiesAccount	-0.93714453	0.285867076	-3.278252757	0.001044518	**
CDAccount	3.823990232	0.324308999	11.79119371	4.33E-32	***
Online	-0.675100462	0.157078748	-4.297847229	1.72E-05	***
CreditCard	-1.119738189	0.204972411	-5.462872711	4.68E-08	***

Probit

Value	Estimate	Std..Error	z.value	Pr...z..	Significance
(Intercept)	-5.663668389	2.097251598	-2.700519286	0.006923132	**
Age	-0.030352836	0.031284147	-0.970230581	0.331931571	
Experience	0.033757575	0.031131474	1.08435518	0.278207348	
Income	0.02773537	0.001270742	21.82611766	1.31E-105	***
ZIP.Code	-4.35E-06	2.08E-05	-0.209022633	0.83443057	
Family	0.341832473	0.03753506	9.107018131	8.47E-20	***
CCAvg	0.074403558	0.020928186	3.555184324	0.000377714	***
Education	0.851040933	0.056739351	14.99913058	7.44E-51	***
Mortgage	0.000221156	0.000295005	0.749670821	0.453452985	
SecuritiesAccount	-0.499463778	0.147054431	-3.396455147	0.000682647	***
CDAccount	2.005842506	0.164783138	12.17261993	4.35E-34	***
Online	-0.351613897	0.081073986	-4.336950906	1.44E-05	***
CreditCard	-0.582161077	0.104560378	-5.567702492	2.58E-08	***

After performing a logit and probit analysis, it was determined that Income, Family, CCAvg, Education, SecuritiesAccount, CDAccount, Online, and CreditCard are significant.

Income – As the income of a person increases, so will the likelihood the person will take out a loan.

Family – As the amount of family members increase, so will the likelihood the person will take out a loan.

CCAvg – As CCAvg increases, so will the likelihood of a person taking out a loan.

Education – The more educated a person is, the more likely they will take out a loan.

Securities Account – If the person has a securities account, they are less likely to take out a loan.

CDAccount – If the person has a CD account, they are more likely to take out a loan.

Online – If the person has internet banking facilities, they are less likely to take out a loan.

CreditCard – If the person has a credit card issued by the bank, they are less likely to take out a loan.

2. Add moderating effects (interactions of variables). Which interactions make sense conceptually? Which interactions are statistically significant? How do you interpret the coefficients on these variables?

Logit						
	Estimate	Std..Error	z.value	Pr...z..	Input	Multiply
(Intercept)	-0.280252439	0.920507364	-0.304454315	0.760781793	1	-0.28025
Income	-0.055547154	0.007795314	-7.125710789	1.04E-12	120	-6.66566
Family	0.823075237	0.089519873	9.194329785	3.77E-20	3	2.469226
Education	-7.478967729	0.752775811	-9.935186033	2.93E-23	2	-14.9579
CDAccount	3.396175905	0.319165686	10.64079272	1.92E-26	1	3.396176
Online	-0.800587069	0.194911023	-4.107448907	4.00E-05	1	-0.80059
CreditCard	-1.100190774	0.241326059	-4.558938961	5.14E-06	1	-1.10019
Income:Education	0.083957365	0.00703751	11.92998111	8.26E-33	240	20.14977
					Sum(u)	2.21
					Exp(u)	9.120687
					1+EXP(u)	10.12069
					Logit	90%
Income 120 family 3	Income 120 family 1					
90%	64%					

The interactions that make the most sense conceptually are family and credit card. For example if a person has a larger family, they might need more money to support their family. Therefore, they will be more likely to take out a loan. Additionally, if a person has a credit card they are less likely to take out a loan because they already have a credit line so they would not need additional funds to borrow. All of these interactions in the snippet above are statistically significant in the analysis. If a coefficient is positive, it is

directly proportional to the probability of a person taking out a loan. If the coefficient is negative it is inversely proportional to the probability of a person taking out a loan. The only exception for the snippet above is income and family and this is due to the moderating effects. As one can see the coefficient for Income:Education is positive, this means as income and education increase, so will the probability of this person deciding to take out a loan.

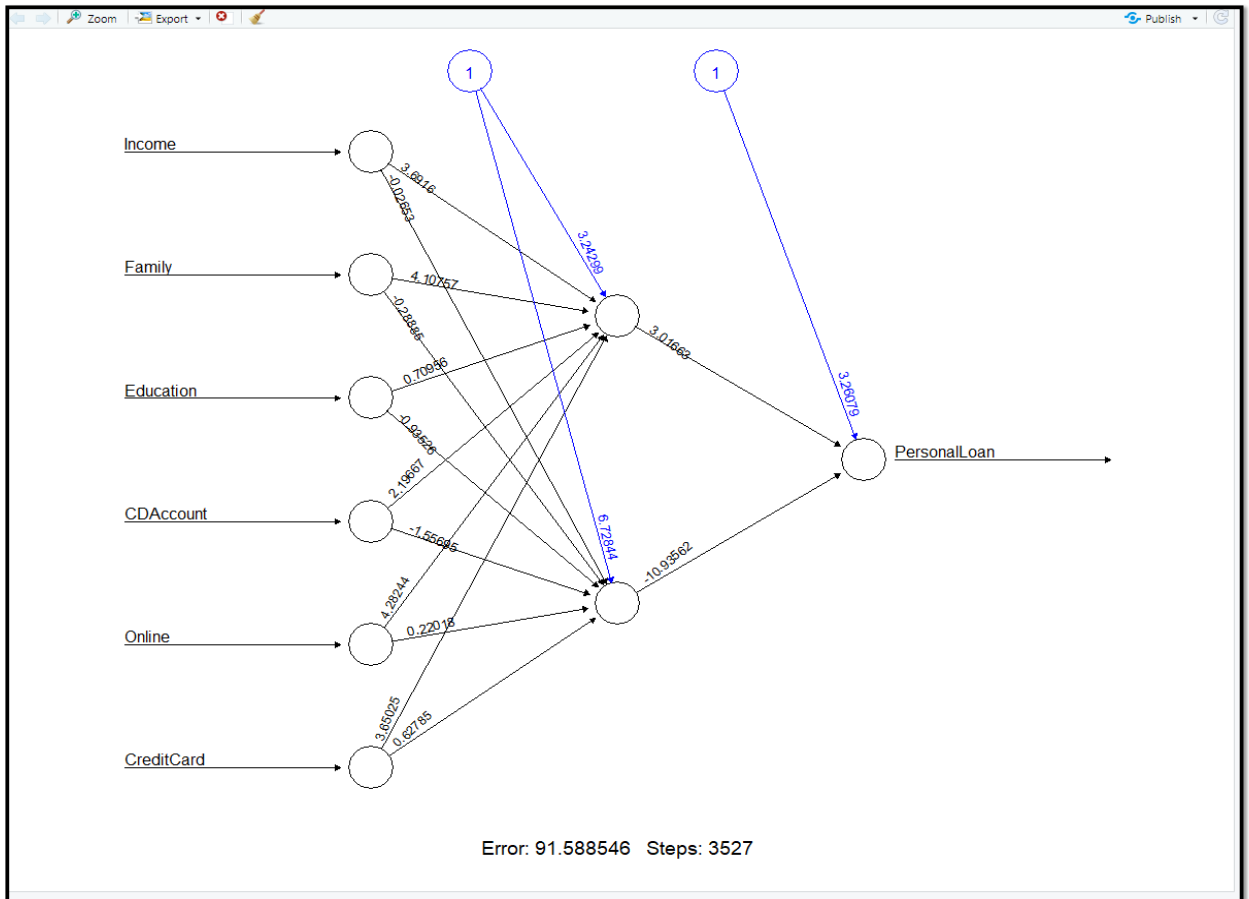
3. Create a final regression model with the variables that you feel are important (both main effects and interaction terms). Create a spreadsheet prediction of the model. Which variables have the greatest influence on the customers' loan behavior (combined main effects and interaction effects)? Perform a sensitivity analysis as seen earlier in the semester.

Logit	Estimate	Std..Error	z.value	Pr...z..	Input	Multiply						
(Intercept)	-13.26134074	0.555312733	-23.88085119	4.84E-126	1	-13.26134074						
Income	0.057628199	0.002429891	23.71637587	2.44E-124	120	6.915383877						
Family	0.693350561	0.073645804	9.414664891	4.75E-21	3	2.080051682						
Education	1.669689353	0.110110852	15.16371299	6.15E-52	2	3.339378705						
CDAccount	3.37123592	0.277327365	12.15616037	5.32E-34	1	3.37123592						
Online	-0.60495853	0.154769243	-3.908777483	9.28E-05	1	-0.60495853						
CreditCard	-1.015237769	0.199004398	-5.101584599	3.37E-07	1	-1.015237769						
					Sum(u)	0.82						
					Exp(u)	2.280770097						
					1+EXP(u)	3.280770097						
					Logit	0.695193515						
		Income										
	0.695193515	20	40	60	80	100	120	140	160	180	200	
CreditCard	0	2%	6%	17%	39%	67%	86%	95%	98%	100%	100%	
	1	1%	2%	7%	19%	42%	70%	88%	96%	99%	100%	
		Income										
	0.695193515	20	40	60	80	100	120	140	160	180	200	
Education	1	0%	0%	1%	4%	12%	30%	58%	81%	93%	98%	
	2	1%	2%	7%	19%	42%	70%	88%	96%	99%	100%	
	3	4%	11%	28%	55%	79%	92%	97%	99%	100%	100%	

The two variables that have the greatest influence on a customer deciding to take out a loan are income and education. As one can see in the sensitivity analysis above, when analyzing income and credit card, income has a great influence on that person deciding to take out a loan. For example, as income increases the probability of a person deciding to take out a loan is still increasing whether a person has a credit card or not. Additionally, as income gets higher the probabilities of a person taking out a loan whether they have a credit card or do not have a credit card get closer and closer until at income of \$200,000 the probability of that person taking out a loan is 100% whether they have a credit card or not. When analyzing income and education, one can see both of these inputs have a great influence. For example, we know income has a strong influence by itself based on looking at the probability increase from left to right in all three education categories. But education also has a strong influence because if one looks at the income of \$120,000, they will see the large differences in probability based on their education level.

- Perform a neural network analysis of the variables found to be significant in the logit and probit analysis above.

error	91.5885463
reached.threshold	0.09347697
steps	3527
Intercept.to.1layhid1	3.24299009
Income.to.1layhid1	3.69159872
Family.to.1layhid1	4.10756795
Education.to.1layhid1	0.70956281
CDAccount.to.1layhid1	2.19667238
Online.to.1layhid1	4.28243774
CreditCard.to.1layhid1	3.65025475
Intercept.to.1layhid2	6.72843656
Income.to.1layhid2	-0.02652504
Family.to.1layhid2	-0.28884698
Education.to.1layhid2	-0.93526381
CDAccount.to.1layhid2	-1.55695133
Online.to.1layhid2	0.2201826
CreditCard.to.1layhid2	0.62785415
Intercept.to.PersonalLoan	3.26078664
1layhid1.to.PersonalLoan	3.01663433
1layhid2.to.PersonalLoan	-10.93562299



5. Create a prediction model of the neural network. Using the prediction model, perform a sensitivity analysis for the neural network model similar to the logit and probit sensitivity analysis.

error	91.58855
reached.threshold	0.093477
steps	3527
Intercept.to.1layhid1	3.24299
Income.to.1layhid1	3.691599
Family.to.1layhid1	4.107568
Education.to.1layhid1	0.709563
CDAccount.to.1layhid1	2.196672
Online.to.1layhid1	4.282438
CreditCard.to.1layhid1	3.650255
Intercept.to.1layhid2	6.728437
Income.to.1layhid2	-0.02653
Family.to.1layhid2	-0.28885
Education.to.1layhid2	-0.93526
CDAccount.to.1layhid2	-1.55695
Online.to.1layhid2	0.220183
CreditCard.to.1layhid2	0.627854
Intercept.to.PersonalLoan	3.260787
1layhid1.to.PersonalLoan	3.016634
1layhid2.to.PersonalLoan	-10.9356

Input	
Intercept	1
Income	120
Family	3
Education	3
CDAccount	1
Online	1
CreditCard	1

Hidden Node 1		
Variable	Coefficient	Coef*Value
Intercept	3.24299009	3.24299009
Income	3.69159872	442.9918464
Family	4.10756795	12.32270385
Education	0.70956281	2.12868843
CDAccount	2.19667238	2.19667238
Online	4.28243774	4.28243774
CreditCard	3.65025475	3.65025475
	sum	470.8155936
	exp(sum)	2.969E+204
	Probability	100%

Hidden Node 2		
Variable	Coefficient	Coef*Value
Intercept	6.72843656	6.72843656
Income	-0.02652504	-3.1830048
Family	-0.28884698	-0.86654094
Education	-0.93526381	-2.80579143
CDAccount	-1.55695133	-1.55695133
Online	0.2201826	0.2201826
CreditCard	0.62785415	0.62785415
	sum	-0.83581519
	exp(sum)	0.433520935
	Probability	30%

Output			
Variables	Coefficient	Value	Coef*Value
Intercept	3.26078664	100%	3.26078664
Hidden1	3.01663433	100%	3.01663433
Hidden2	-10.93562299	30%	-3.307117037
		Sum	2.970303933
		exp(sum)	19.49784474
		Probability	95%

Sensativity Analysis										
	Income									
Education	95%	20	40	60	80	100	120	140	160	180
	1	1%	1%	2%	3%	6%	14%	37%	71%	91%
	2	2%	3%	5%	11%	30%	63%	88%	96%	99%
	3	4%	9%	24%	55%	84%	95%	98%	99%	100%
	Income									
CreditCard	95%	20	40	60	80	100	120	140	160	180
	1	4%	9%	24%	55%	84%	95%	98%	99%	100%
	0	11%	28%	61%	87%	96%	99%	99%	100%	100%