SCM 651

Homework 4

Loan Analysis

Using the Universal Bank data, determine the factors which influence whether a customer takes out a loan

Resources

Use the dataset SCM 651 Homework 4 Universal Bank.csv.

Outline and grading criteria:

1. **Perform a logit and probit analysis of the variables that affect whether a customer takes out a loan. Consider only main effects.**

    a. *Which variables are significant?*

    Using logit, when considering age, income, credit card average, family size, education, mortgage, and securities account, all of the variables are statistically significant (when *alpha* = 0.05) **except** age (just barely, at *p-value* = 0.055) and mortgage (*p-value* = 0.2).

    Using probit, we see the same variables are statistically significant (except age at *p-value* = 0.15, and mortgage at *p-value* = 0.14), but with slightly different *p-values*.

    b. *How do the significant variables influence the likelihood of taking out a loan?*

    Using logit, all of the significant variables have a positive effect. When income, family size, credit card average, education, and securities increases, the likelihood of taking out a loan also increases. No statistically significant variables had a negative effect that would decrease the likelihood of taking out a loan.

    Using probit, we see the same positive effect for all of the statistically significant variables. No statistically significant variables had a negative effect that would decrease the likelihood of taking out a loan.

    c. *Copy screen snapshots of your analysis in R to your report.*

    Logit

```
Coefficients:
                    Estimate  Std. Error z value Pr(>|z|)
(Intercept)       -14.1116316   0.6216320 -22.701  < 2e-16 ***
Age                 0.0117206   0.0060994   1.922 0.054656 .
Income              0.0544416   0.0024091  22.599  < 2e-16 ***
CCAvg               0.1263517   0.0364847   3.463 0.000534 ***
Family              0.6918473   0.0690324  10.022  < 2e-16 ***
Education           1.6597624   0.1044880  15.885  < 2e-16 ***
Mortgage            0.0006615   0.0005213   1.269 0.204465
SecuritiesAccount   0.5173873   0.2111582   2.450 0.014276 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Probit

```
Coefficients:
                    Estimate Std. Error z value  Pr(>|z|)
(Intercept)       -7.1134039  0.2915124 -24.402   < 2e-16 ***
Age                0.0045947  0.0031577   1.455   0.14565
Income             0.0277167  0.0011750  23.588   < 2e-16 ***
CCAvg              0.0792593  0.0194629   4.072 0.0000465 ***
Family             0.3398349  0.0351090   9.679   < 2e-16 ***
Education          0.8074032  0.0512303  15.760   < 2e-16 ***
Mortgage           0.0004028  0.0002762   1.458   0.14474
SecuritiesAccount  0.2941015  0.1094193   2.688   0.00719 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2. Add moderating effects (interactions of variables).

a. Which interactions make sense conceptually?

Income and *CCAvg* make sense to compare, because the data roughly shows that as income increases, the average monthly spending on credit cards also increases. People with higher amounts of credit card debt often look to consolidate their credit card balances using a person loan to lower the monthly interest charges and pay one monthly amount compared to all their credit cards separately. The monthly loan payment is also potentially less money than their separate credit card payments combined, which frees up more cash for other expenses or savings each month. Those with a lower income and a higher credit card debt would be more inclined to try and consolidate debt with a personal loan in order to decrease their monthly payments and increase their monthly cash flow. However, people with a higher income and little to no credit card debt would not be inclined to take out a personal loan.

b. Which interactions are statistically significant?

All three variables (*income, CCAvg, income*CCAvg*) were statistically significant using *alpha* = 0.05.

c. How do you interpret the coefficients on these variables?

When *income* and *CCAvg* increase, the likelihood of taking out a loan also increases, since both of these coefficients are positive. However, *income*CCAvg* has a negative effect. This is because *income* and *CCAvg* are including all the other variables, include age, mortgage, experience, and family size.

d. Copy screen snapshots of your analysis in R to your report.

```
Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)  -9.5556476  0.4309632  -22.17   <2e-16 ***
Income        0.0600142  0.0031175   19.25   <2e-16 ***
CCAvg         1.3398815  0.1215377   11.02   <2e-16 ***
Income:CCAvg -0.0086107  0.0008001  -10.76   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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.

a. Which variables have the greatest influence on the customers' loan behavior (combined main effects and interaction effects)?

We chose *income*, *CCAvg*, *family*, *education*, and a moderating effect of *Income*CCAvg* for a final regression model. All of the coefficients were statistically significant. In the model we created, the variables that seemed to effect the biggest changes in the probability were *CCAvg*, *family*, and *Education*. The effect of *income* was not as large of a probably change as the other variables.

Inputs		
Variable	Value	
Income	90	Annual income in $
CCAvg	4	Avg spending on CC/month
Family	4	Family size
Education	2	Education level
Income*CCAvg	360	Interaction

Output			
Variable	Coefficient	Value	Coeff*Value
Intercept	-20.14606	1	-20.14606
Income	0.098543	90	8.86887
CCAvg	2.2201	4	8.8804
Family	0.770788	4	3.083152
Education	1.818407	2	3.636814
Income*CCAvg	-0.014008	360	-5.04288
		Sum	-0.719704
		Exp(sum)	0.48689636
		Probability	33%

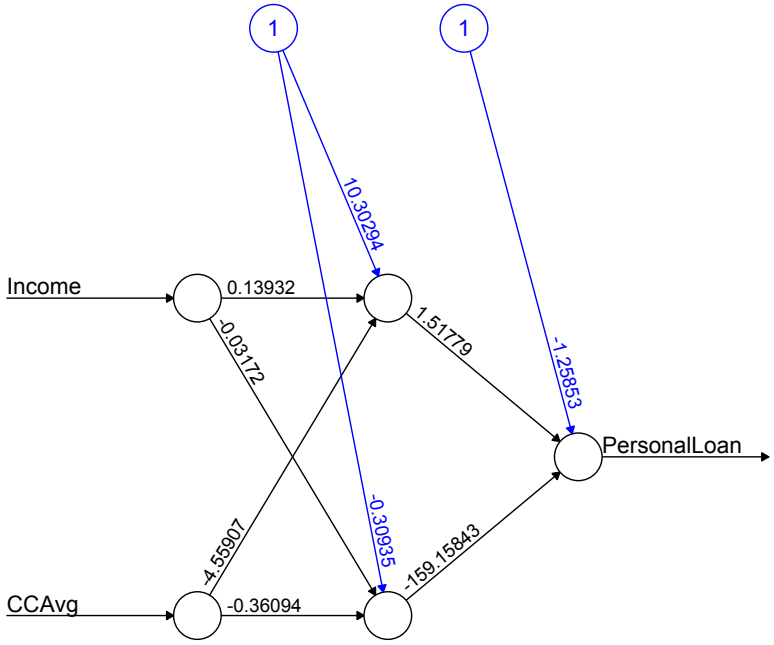
b. Perform a sensitivity analysis as seen earlier in the semester.

Sensitivity Analysis							
		CCAvg					
	33%	0	2	4	6	8	10
Income	8	0%	0%	1%	50%	99%	100%
	25	0%	0%	3%	56%	98%	100%
	42	0%	0%	6%	62%	98%	100%
	59	0%	1%	12%	68%	97%	100%
	76	0%	3%	21%	73%	96%	100%
	93	1%	8%	36%	78%	96%	99%
	110	7%	23%	53%	82%	95%	99%
	127	29%	49%	70%	85%	93%	97%
	144	68%	76%	83%	88%	92%	94%
	161	92%	91%	91%	90%	90%	89%
	178	98%	97%	95%	92%	87%	80%
	195	100%	99%	98%	94%	85%	66%
	212	100%	100%	99%	95%	81%	49%
	224	100%	100%	99%	96%	79%	37%

c. *Copy screen snapshots of your analysis in R to your report.*

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -20.146060   0.883313  -22.81   <2e-16 ***
Income         0.098543   0.004896   20.13   <2e-16 ***
CCAvg          2.220100   0.167626   13.24   <2e-16 ***
Family         0.770788   0.074187   10.39   <2e-16 ***
Education      1.818407   0.109227   16.65   <2e-16 ***
Income:CCAvg  -0.014008   0.001091  -12.84   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

4. **Perform a neural network analysis of the variables found to be significant in the logit and probit analysis above. Copy screen snapshots of your final neural network model in R to your report.**



Error: 144.281108 Steps: 21291

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.

Inputs		
Variable	Value	
Income	80	Annual income in $
CCAvg	6	Avg spending on CC/month

Hidden Nodes			
Hidden Node 1			
Variable	Coefficient	Value	Coeff*Value
Intercept	10.30293778	1	10.30293778
Income	0.139318425	80	11.145474
CCAvg	-4.559067935	6	-27.35440761
		Sum	-5.905995828
		Exp(sum)	0.002723069
		Probability	0.00272
Hidden Node			
Variable	Coefficient	Value	Coeff*Value
Intercept	-0.30934774	1	-0.30934774
Income	-0.031719444	80	-2.53755552
CCAvg	-0.360938344	6	-2.165630064
		Sum	-5.012533324
		Exp(sum)	0.006654025
		Probability	0.00661

Output			
Variable	Coefficient	Value	Coeff*Value
Intercept	-1.258527321	1	-1.258527321
Hidden 1	1.517794771	0.00271567	0.004121835
Hidden 2	-159.1584312	0.00661004	-1.052043875
		Sum	-2.306449361
		Exp(sum)	0.099614319
		Probability	9%

Sensitivity Analysis							
		CCAvg					
	9%	0	2	4	6	8	10
Income	8	0%	0%	0%	0%	0%	2%
	25	0%	0%	0%	0%	2%	6%
	42	0%	0%	0%	1%	5%	11%
	59	0%	0%	1%	4%	10%	15%
	76	0%	1%	9%	8%	14%	18%
	93	0%	7%	23%	13%	17%	19%
	110	4%	19%	36%	19%	19%	21%
	127	14%	32%	44%	38%	20%	21%
	144	28%	42%	49%	51%	21%	22%
	161	39%	48%	52%	54%	22%	22%
	178	46%	51%	54%	55%	27%	22%
	195	50%	54%	55%	56%	46%	22%
	212	53%	55%	56%	56%	55%	22%
	224	54%	55%	56%	56%	56%	23%