BAN 5733 – Individual Exercise 10 (10 Points) Solution

A telecom operator which has successfully launched a fourth generation (4G) mobile telecommunications network would like to make use of existing customer usage and demographic data to identify which customers are likely to switch to using their 4G network. The target categorical variable is "Customer_Type" (3G/4G). A 4G customer is defined as a customer who has a 4G Subscriber Identity Module (SIM) card and is currently using a 4G network compatible mobile phone.

Variable Description:
Make sure you are using the Telecom_Ex10.jmp dataset for this assignment.

Make sure the model roles and measurement levels of the variables are as shown below.

Attribute Name	Role	Level	Description
Age_Impute (and flag)	INPUT	INTERVAL	Customer age in years
Contract_flag_impute (and flag)	INPUT	NOMINAL	Contract ownership flag (Y/N)
CUSTOMER_CLASS	INPUT	NOMINAL	Codes indicating VIP, Individual, Corporate, Government, Under 21, Foreigner, etc.
CUSTOMER_TYPE	TARGET	Binary	Target Field: 4G Customer Flag (3G/ 4G)
Gender_Impute (and flag)	INPUT	NOMINAL	Male or Female
Log(HS_AGE)	INPUT	INTERVAL	Handset age in months
ID_CHANGE_FLAG	REJECTED	NOMINAL	1 if the customer changed the ID in the last 6 months
LINE_TENURE	INPUT	INTERVAL	Line tenure in days
Marital_Status_Impute (and flag)	INPUT	NOMINAL	Marital status
NATIONALITY	REJECTED	NOMINAL	Nationality
Occup_CD_Impute (and flag)	INPUT	NOMINAL	Occupation code
Pay_Meth_Impute (and flag)	INPUT	NOMINAL	Payment method code (Credit Card, Cash, etc.) as of last billing cycle in the 6 months
SERIEL_NUMBER	ID	NOMINAL	Record Index (ID)
Log(SUBPLAN)	INPUT	INTERVAL	Current subscription plan type
TOP1_INT_CD	REJECTED	NOMINAL	Top 1 international country
TOP2_INT_CD	REJECTED	NOMINAL	Top 2 international country
TOP3_INT_CD	REJECTED	NOMINAL	Top 3 international country
VALIDATION	VALIDATION	NOMINAL	Validation or training indicator
Prob(DT_CT==3G)	REJECTED	INTERVAL	Probability of 3G customer type from decision tree
Prob(DT_CT==4G)	REJECTED	INTERVAL	Probability of 4G customer type from decision tree
DT_PRED_CT	REJECTED	NOMINAL	Predicted outcome of customer type from decision tree
Reg1 - Lin[4g] - Prob[3G] - Prob[4G] - Most Likely C_T	REJECTED	Various	Regression results from Regression1 analysis – no transformations
Reg2 Same variables as Reg1	REJECTED	Various	Regression results from Regression1 analysis – transformations
Reg3 Same variables as Reg1	REJECTED	Various	Regression results from Regression1 analysis – optimized regression

1. You will work with **`Telecom_Ex10.jmp`** data set that contains 48 variables and 18,000 observations. The variables in the data set are shown above with the appropriate roles and levels.
   - ✓ Open the **Telecom_Ex10.jmp** dataset provided. Make sure the model roles and measurement levels are shown as above.
   - ✓ Save the dataset as Telecom_Ex10_[yourinitials}.jmp. You will save scripts to the data table instead of a project.

   a. Is the imputation of missing values needed for a neural network model? Why or why not? (1 Point)

   Solution: Yes, imputation of missing values is needed for a neural network model because just like regression model, the neural network model requires a complete record for estimation and scoring. It does not handle the missing values.

   b. Is data transformation generally needed for a neural network model? Why or why not? (1 Point)

   Solution: Yes, transformation is generally needed for a neural network model. Transformation is required to achieve normality and thus improve the performance of the model just like in regression.

2. Neural Network 1
   - ✓ Install the Random Seed Reset add-in and run it
   - ✓ Set the seed to 12345
   - ✓ Analyze >>Predictive Modeling>>Neural
   - ✓ Enter the target variable in the Y section
   - ✓ Enter the input variables for the model in the X section
   - ✓ Enter Validation into the Validation section
   - ✓ Select OK
   - ✓ Leave all defaults in place and hit Go
   - ✓ Save the script to the data table and name "Neural Net 1"

   a. **How does the validation misclassification rate of this neural net model compare to the three regression models built earlier?** For this exercise, you may use the regression scripts in the data table to obtain base information for the misclassification rate of each of the regression runs. **(0.5 Points)**

   Solution:

Placed in X column in order of JMP data table

Model	Regression Model 1	Transformed Regression Model 2	Optimized Regression Model 3	Neural Network Model 1a
Misclassification Rate	0.1543	.1459	.1531	0.1498
Sensitivity	19.2%	36.7%	20.0%	36.4%

Placed in X column in order of data dictionary above

Model	Regression Model 1	Transformed Regression Model 2	Optimized Regression Model 3	Neural Network Model 1b
Misclassification Rate	0.1543	.1459	.1531	0.1426
Sensitivity	19.2%	36.7%	20.0%	44.2%

b. **Which model would be selected at this point? (0.5 Points)**

Solution: From the above table, the Misclassification Rate is lowest for the Transformed Regression Model if items are entered in order of the JMP data table (1a). If they were entered into X variable section in order of the data dictionary above (1b), the misclassification rate is smallest for NN1. The sensitivity rates showed a similar pattern.

3. Neural Network 2
   - ✓ Install the Random Seed Reset add-in and run it
   - ✓ Set the seed to 12345
   - ✓ Analyze >>Predictive Modeling>>Neural
   - ✓ Enter the target variable in the Y section
   - ✓ Enter the following variables in the X section in order: Log[HS_Age], Contract_Flag_Impute, Age_Impute, Line_Tenure, Gender_Impute, Log[Subplan], Pay_Meth_Impute
   - ✓ Enter Validation into the Validation section
   - ✓ Select OK
   - ✓ Leave all defaults in place and hit Go
   - ✓ Save the script to the data table and name "Neural Net 7 vars ordered"

a. **What is the validation misclassification rate? Does it change from the previous model (Neural Net 1)? (0.5 Points)**

Model	Regression Model 1	Transformed Regression Model 2	Optimized Regression Model 3	Neural Network Model 1a	Neural Network Model 1b	Neural Network Model 2
Misclassification Rate	0.1543	.1459	.1531	0.1498	0.1426	.1444
Sensitivity	19.2%	36.7%	20.0%	36.4%	44.2%	37.6%

The NN2 model does improve over the NN1a or the regression models. However, it does not improve over the NN1b.

4. Neural Network 3
   - ✓ Install the Random Seed Reset add-in and run it
   - ✓ Set the seed to 12345
   - ✓ Analyze >>Predictive Modeling>>Neural
   - ✓ Enter the target variable in the Y section
   - ✓ Enter the following variables in the X section in order: Log[HS_Age], Contract_Flag_Impute, Age_Impute, Line_Tenure, Gender_Impute
   - ✓ Enter Validation into the Validation section
   - ✓ Select OK
   - ✓ Leave all defaults in place and hit Go
   - ✓ Save the script to the data table and name "Neural Net 5 vars"

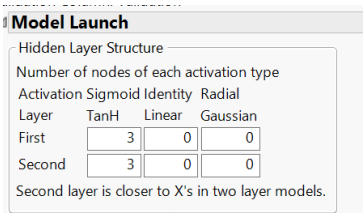
   a. **What is the validation misclassification rate? Does it change from the previous models? (0.5 Points)**

Model	Regression Model 1	Transformed Regression Model 2	Optimized Regression Model 3	Neural Network Model 1a	Neural Network Model 1b	Neural Network Model 2	Neural Network Model 3
Misclassification Rate	0.1543	.1459	.1531	0.1498	0.1426	.1444	.1404
Sensitivity	19.2%	36.7%	20.0%	36.4%	44.2%	37.6%	41.4%

Again, this new model does better than the previous models when you consider both misclassification rate and the sensitivity is better than all models except 1b. We are better able to predict customer type with a reduced number of variables.

5. Neural Network 4
   - ✓ Install the Random Seed Reset add-in and run it
   - ✓ Set the seed to 12345
   - ✓ Analyze >>Predictive Modeling>>Neural

- ✓ Enter the target variable in the Y section
- ✓ Enter the following variables in the X section in order: Log[HS_Age], Contract_Flag_Impute, Age_Impute, Line_Tenure, Gender_Impute
- ✓ Enter Validation into the Validation section
- ✓ Select OK
- ✓ Add a second set of 3 hidden layers
- ✓ Save the script to the data table and name "Neural Net 5 vars 3X3"



a. What is the validation misclassification rate? Does it change from the previous models? (0.5 Points)

Model	Regression Model 1	Transformed Regression Model 2	Optimized Regression Model 3	Neural Net Model 1a	Neural Net Model 1b	Neural Net Model 2	Neural Net Model 3	Neural Net Model 4
Misclassification Rate	0.1543	.1459	.1531	0.1498	0.1426	.1443	.1404	.145
Sensitivity	19.2%	36.7%	20.0%	36.4%	44.2%	37.6%	41.4%	36.3%

This more complex neural net model did not improve on our ability to predict customer type. We would not select this model with respect to misclassification rate.

b. Which parameter estimate is the largest? Report the value. (0.5 Points)

The Customer Type (3G) to H1_3 layer has the largest positive parameter estimate but the Customer_Type (3G) to H1_2 parameter actually has the largest value at -18.007.

While these are not too informative when viewed alone, the weights can be combined to create a prediction formula that can be saved to the data table and applied to other data. In addition, you can examine the relationship of all the variables through the profiler from inside the neural network.

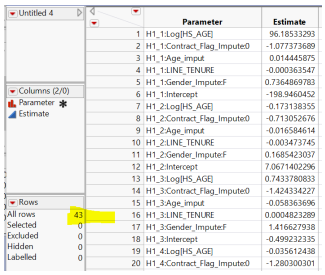
6. Neural Network 5
   - ✓ Install the Random Seed Reset add-in and run it
   - ✓ Set the seed to 12345
   - ✓ Analyze >>Predictive Modeling>>Neural
   - ✓ Enter the target variable in the Y section
   - ✓ Enter the following variables in the X section in order: Log[HS_Age], Contract_Flag_Impute, Age_Impute, Line_Tenure, Gender_Impute
   - ✓ Enter Validation into the Validation section
   - ✓ Select OK
   - ✓ Change the number of hidden nodes to read 6 in the first layer only
   - ✓ Save the script to the data table and name "Neural Net 5 vars 6 nodes"

a. What is the validation misclassification rate? Does it change from the previous models? (0.5 Points)

Model	Regression Model 1	Transformed Regression Model 2	Optimized Regression Model 3	Neural Net Model 1a	Neural Net Model 1b	Neural Net Model 2	Neural Net Model 3	Neural Net Model 4	Neural Net Model 5
Misclassification Rate	0.1543	.1459	.1531	0.1498	0.1426	.1443	.1404	.145	.1420
Sensitivity	19.2%	36.7%	20.0%	36.4%	44.2%	37.6%	41.4%	36.3%	40.9%

b. How many parameters are created by this model? (0.5 Points)

Solution: Neural Net 5 estimated 43 parameters. This can be quickly seen when you make a table out of the estimates.



7. Model Selection

a. Create a table that shows the number of parameters estimated, the training and validation misclassification rates, and the training and validation sensitivity rates for the decision tree, the three regression models, and the neural networks built in this exercise. (1.5 Points)

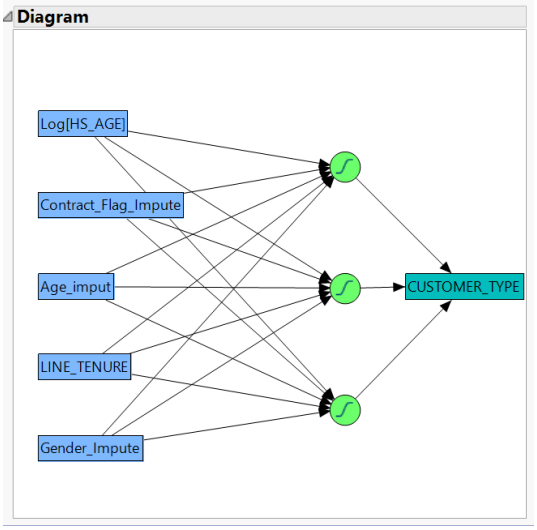
Model	Optimized Decision Tree*	Regression Model 1	Transformed Regression Model 2	Optimized Regression Model 3	Neural Net Model 1a	Neural Net Model 1b	Neural Net Model 2	Neural Net Model 3	Neural Net Model 4	Neural Net Model 5
Parameters	8 vars/22 leaves	15	14	19	121	121	37	22	34	43
Training Misclassification	.1416	.1532	.1460	.1523	.1497	.1459	.1448	.1444	.1451	.1444
Validation Misclassification	.1398	0.1543	.1459	.1531	.1498	.1426	.1444	.1404	.1450	.1420
Training Sensitivity	31.8%	17.6%	33.3%	18.2%	33.9%	40.2%	35.1%	37.5%	33.0%	37.9%
Validation Sensitivity	33.2%	19.2%	36.7%	20.0%	36.4%	44.2%	37.6%	41.4%	36.3%	40.9%

*Optimized Decision Tree numbers were not required because they were not part of the exercise 10 data table.

b. Select the best fitting neural network model according to Validation Misclassification Rate and report the following information.

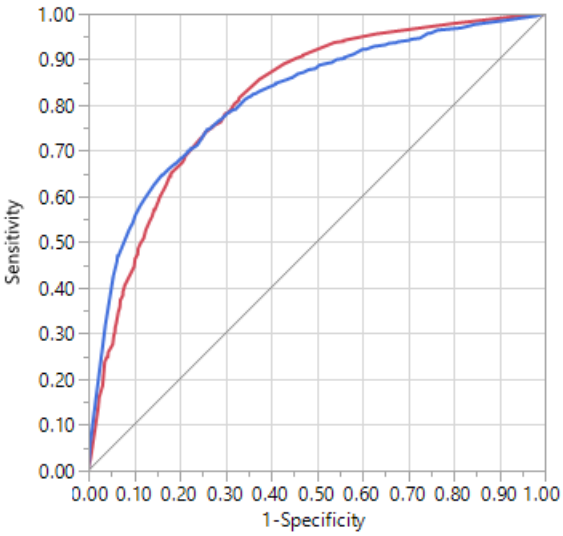
i. Report the Neural Network Diagram (0.5 Points)

Solution: Neural Net 3 has the best validation misclassification rate.



ii. Report the Validation ROC curve (0.5 Points)

Solution: This neural network produces an ROC Index of .8148.



iii. Report the Validation Confusion Matrix and report the Sensitivity and Specificity. (0.5 Points)

Solution: The validation sensitivity rate is .414 or this model is accurately predicting the 4G customers 41.4% of the time. While the validation specificity is 0.051 or 5.1% of the time the model is inaccurately predicting 4G customer status when they were 3G.

Confusion Matrix

Actual	Predicted Count	
CUSTOMER_TYPE	3G	4G
3G	4269	231
4G	527	373

Confusion Rates

Actual	Predicted Rate	
CUSTOMER_TYPE	3G	4G
3G	0.949	0.051
4G	0.586	0.414

c. How does this model compare to other models you have built so far (including the three Regression models and all neural networks)? (1 Points)

Solution: From the above table, you can see that neural network 3 is the best of all the neural net models according to misclassification rate but the decision tree produces the lowest misclassification rate overall (.1398). However, it does have a lower validation sensitivity than the neural net (33.2% compared to 41.4%) but it also has a lower validation specificity (3.4% compared to 5.1%). The regression models did not really come close to being the best.