**BAN 5733 – Individual Exercise 11 (10 Points)**

**Neural Networks**

Exercise Description:

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. You are asked to assist them in examining the data.

You completed a previous report for this same company using a decision tree for analyzing service. The telecom company would like you to re-evaluate the analysis using some other methods and answer some specific questions they have about the data and the results.

Variable Description:

*Make sure you are using the Telecom\_Ex11.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 |

Questions:

1. You will work with **Telecom\_Ex11.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\_Ex11.jmp** dataset provided. Make sure the model roles and measurement levels are shown as above.
* Save the dataset as Telecom\_Ex11\_[yourinitials}.jmp. You will save scripts to the data table instead of a project.

1. **Is the imputation of missing values needed for a neural network model? Why or why not? (1 Point)**
2. **Is data transformation generally needed for a neural network model? Why or why not? (1 Point)**
3. 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”

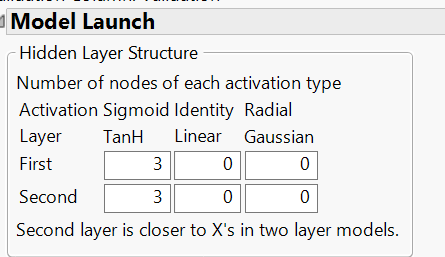
1. **How does the validation misclassification rate of this neural net model compare to the three regression models built by the regression scripts?** For this exercise, you may use the regression scripts in the data table to obtain base information for the misclassification rate of each of the regression runs. **(0.5 Points)**
2. **Which model would be selected at this point? (0.5 Points)**
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 8 vars”

1. **What is the validation misclassification rate? Does it change from the previous model (Neural Net 1)?** **(0.5 Points)**
2. 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”

1. **What is the validation misclassification rate? Does it change from the previous models?** **(0.5 Points)**
2. 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”

1. **What is the validation misclassification rate? Does it change from the previous models?** **(0.5 Points)**
2. **Which parameter estimate is the largest? Report the value. (0.5 Points)**
3. 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”

1. **What is the validation misclassification rate? Does it change from the previous models?** **(0.5 Points)**
2. **How many parameters are created by this model? (0.5 Points)**
3. Model Selection
4. **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 all the neural networks built in this exercise.** Note: decision tree and regression models have saved as variables in the data table and you can use Fit Y by X to recreate their confusion matrixes. **(1 Point)**
5. **Select the best fitting neural network model according to Validation Misclassification Rate and report the following information.** 
   * 1. **Report the Neural Network Diagram (0.5 Points)**
     2. **Report the Validation ROC curve (0.5 Points)**
     3. **Report the Validation Confusion Matrix and report the Sensitivity and Specificity. (1 Point)**
6. **How does this model compare to other models you have built so far (including the three Regression models and all neural networks)? (1 Point)**

Deliverables (please follow these instructions):

* As you complete the exercise, create a report in Microsoft Word. In this report, answer the questions in the exercise description.
* Make sure you comment or explain and not just provide snapshots of data.
* Limit your report to no more than **7 pages** including tables and diagrams.
* Copy and paste or screen shot supporting tables/diagrams as needed to justify any of your answer. You may need to shrink your table/ diagrams but please ensure they are readable.
* Include any required data sets or codes/projects as requested as separate files.
* Make sure you print your name, student ID#, student email on the cover page of the report and turn-in the report as communicated by your instructor.
* Please also put a running header/footer with your name, on each page of your exercise solution report.

Failure to follow these instructions will result in deduction of points