Logistic Regressions

This exercise builds on the previous exercises of Data Preprocessing & Exploration and Decision Trees.

1. **Adding New Data Partition**
   Steps: Sample ➔ Data Partition ➔ Change values for Data Set Allocations in Property panel
   - 60% Training data
   - 40% Validation data
   - Connect it with the data source ➔ Run

2. **Adding a Regression node**
   Steps: Model ➔ Regression
   - Choose Logistic Regression in Property panel
   - Connect it with the new data partition
   - Run and view the Results
   - **Is there anything you are concerned? Why or why not?**

   - **Can you see the same problem with Decision Trees? (Note: To answer this question, you can add another decision tree node, connect it with the new data partition, run it, and see the results)**

   - **To fix the above issue with regression node, how do you want to modify your process flow?**

3. **Re-run the Regression node and see the Results.**
   - **How many parameters were estimated?**

   - **How many observations were used to build the full model?**

   - **What is the model's misclassification rate (on the validation)?**

   - **Is this regression model optimized? Why or why not?**

   - Renaming the node: Right-Click ➔ Select Rename ➔ Type "Regression: Full"

4. **Selecting Input Variables**
   - Add another Regression node
   - Change Selection Model (under Model Selection) from None to Backward
   - Run Regression and see the Results
   - **How many variables are included in the final model?**

   - **What is the model's misclassification rate (on the validation)?**

   - Select View ➔ Model ➔ Iteration Plot
   - **Are you concerned with the iteration plot? Why or why not?**

5. Creating Optimal Regression Model

Steps: Click Regression

- Change Selection Criterion from Default to Validation Misclassification
- Run and See the Results
- Renaming the node: Right-Click ➔ Select Rename ➔ Type "Regression: Optimal"

Misclassification Rate

Confusion Matrix

		Predicted (Output in SASEM output)	
		1	**0**
Actual (Target in SASEM output)	**1**		
	0		

Error Report

Class	# of Cases	# of Errors	% of Errors
1			
0			
Overall			

Model Implementation

6. **Adding Score Data Set**
   Steps: File ➔ New ➔ Data Source ➔ Source: SAS Table ➔ Click Browse ➔ Click Padata library ➔ Click data file (bank_personal_loan_score)

7. **Adding Score Dataset in Diagram Workspace**
   Steps: Data Sources in Project panel ➔ Move data source to Diagram panel
   - Change Role of the data set from Raw to Score

8. **Adding Score Node**
   Steps: Assess ➔ Score
   - Connect it with the score data set and Regression: Optimal
   - Run and see the Results

9. **Exploring Scored Data**
   Steps: Select Score node ➔ Click the ellipsis next to Exported Data in Property Panel ➔ Select Score ➔ Click Explore
   - **Which customers are classified as "Accept"?**

   - **What are predicted probabilities for them?**

   - **What was the cut-off value for classification?**