Due Dec 9, 5:59 AM BRST

Try again once you are ready

TO PASS 70% or higher

GRADE 69.23%

Week 3 Application Assignment

LATEST SUBMISSION GRADE

69.23%

Submit your assignment

DUE Dec 9, 5:59 AM BRST **ATTEMPTS** 3 every 8 hours

Receive grade

TO PASS 70% or higher

Grade 69%

We keep your highest score

1. **Let's reconsider the customer reward program dataset. In this exercise, you will complete a predictive modeling task where the target variable is binary. Using the following data file for this exercise:** — 1 / 1 point

crp_cleandata.xlsx

The dataset also contains a column IndustryType, which is created based on the column Industry in the raw data. Note that Industry has many categories. The analyst who prepared the data chose to combine some categories, which resulted in the column IndustryType. IndustryType has five categories: Department, Discount, Grocery, Restaurants, Specialty. You can create a set of dummy variables based on IndustryType in XLMiner by using the Transform functions.

Part I.

Consider logistic regression models with Reward column as the target variable. Fit the model with two indicator variables, one indicating whether a retailer is a discount store (i.e., IndustryType is Discount), and the other indicating whether a retailer is a grocery store (i.e., IndustryType is Grocery). Report the coefficient estimates in the next three questions. [Hint: After you create the dummy variables, use them as Selected Variables (instead of Categorical Variables) in the first step of Logistic Regression.]

What is the estimated intercept coefficient?

○ 10
● 0.5108
○ 133.47
○ 0.03023

✓ **Correct**
Bravo!

2. **What is the estimated coefficient for IndustryType_Discount (round the answer to 4 decimal places i.e. x.xxxx) ?** — 1 / 1 point

-0.9628

✓ **Correct**
Bravo!

3. **What is the estimated coefficient for IndustryType_Grocery (round the answer to four decimal places i.e. x.xxxx) ?** — 1 / 1 point

-0.7340

✓ **Correct**
Bravo

4. **What is the number of true positives? (Specify a whole number.)** — 1 / 1 point

40

✓ **Correct**
Bravo!

5. **What is the number of true negatives? (Specify a whole number.)** — 1 / 1 point

21

✓ **Correct**
Bravo!

6. **Part II.** — 1 / 1 point

Split the dataset into training and validation sets using a 60:40 split (set the seed for partitioning to 12345; this should be the default value if you have not changed it). [Hint: note that there two Partition buttons in XLMiner ribbon. You should use the Partition->Standard Partition in the Data Mining group.] Report the new coefficient estimates in the next three questions. Use the same two predictor variables as in Part I.

What is the estimated intercept coefficient (round the answer to 4 decimal places i.e. x.xxxx) ?

0.6567

✓ **Correct**
Bravo!

7. **What is the estimated coefficient for IndustryType_Discount (round the answer up to 4 decimal places i.e. x.xxxx) ?** — 1 / 1 point

-0.8390

✓ **Correct**
Bravo!

8. **What is the estimated coefficient for IndustryType_Grocery (round the answer to 4 decimal places i.e. x.xxxx) ?** — 0 / 1 point

-1,1675

! **Incorrect**
The answer you gave is not a number.

9. **How many observations are in the training set?** — 1 / 1 point

60

✓ **Correct**
Bravo!

10. **What is the number of true positives on the validation data? (Specify a whole number.)** — 0 / 1 point

27

! **Incorrect**
The result is in the Confusion Matrix under Validation Data Scoring - Summary Report. If your answer is incorrect, also check to make sure that the success class is chosen to be 1 in Logistic Regression - Step 1 of 3.

11. **What is the number of true negatives on the validation data? (Specify a whole number.)** — 0 / 1 point

11

! **Incorrect**
The result is in the Confusion Matrix under Validation Data Scoring - Summary Report. If your answer is incorrect, also check to make sure that the success class is chosen to be 1 in Logistic Regression - Step 1 of 3.

12. **(Part 3) By default, XLMiner uses the cutoff threshold 0.5. Repeat Part II with a cutoff threshold 0.3. What are the numbers of true positives and true negatives on the validation data?** — 0 / 1 point

Report the number of true positives:

35

! **Incorrect**
The result is in the Confusion Matrix under Validation Data Scoring - Summary Report. You can change the cutoff in Logistic Regression - Step 1 of 3. If your answer is incorrect, also check to make sure that the success class is chosen to be 1 in Logistic Regression - Step 1 of 3.

13. **Report the number of true negatives:** — 1 / 1 point

0

✓ **Correct**
Bravo!