Report: Classification Metrics and Model Evaluation for Homework 2

Authors: Warner Alexis, Saloua Daouki, Souleymane Doumbia, Fomba Kassoh, Lewris Mota Sanchez
Course: DATA 621 - Business Analytics and Data Mining
Date: 2024-10-13

1. Introduction

In this homework assignment, we analyze a binary classification model's performance using a dataset containing actual and predicted class labels, as well as predicted probabilities. We compute various classification metrics and compare the results between custom functions and those from the caret and pROC packages.

2. Data Overview

The dataset includes three key columns:

1. **class**: Actual class for the observation.
2. **scored.class**: Predicted class based on a threshold of 0.5.
3. **scored.probability**: Predicted probability of success.

Confusion Matrix

The confusion matrix provides an overview of the model's predictions:

Predicted\Actual	**0**	**1**
0 (Negative)	119	30
1 (Positive)	5	27

- **True Negatives (TN)** = 119
- **False Negatives (FN)** = 30
- **False Positives (FP)** = 5
- **True Positives (TP)** = 27

Classification Metrics

3. Accuracy

The accuracy measures the proportion of correct predictions

The accuracy for this model is 0.8066 (80.66%).

4. Classification Error Rate

The classification error rate is the proportion of incorrect predictions

The error rate is 0.1934 (19.34%), and the sum of accuracy and error rate equals 1.

5. Precision

Precision represents the proportion of positive identifications that were actually correct

The precision is 0.8438 (84.38%).

6. Sensitivity (Recall)

Sensitivity (or Recall) represents the proportion of actual positives that were correctly identified

The sensitivity is 0.4737 (47.37%).

7. Specificity

Specificity represents the proportion of actual negatives that were correctly identified

The specificity is 0.9597 (95.97%).

8. F1 Score

The F1 score is the harmonic mean of Precision and Sensitivity

The F1 score is 0.6067 (60.67%).

9. Bounds on the F1 score

The F1 score for boundary case of (0, 1), (0.5, 0.5), and (1, 1) and for intermediate values for Precision and Recall were all shown to be between 0 and 1.

10. ROC Curve and AUC

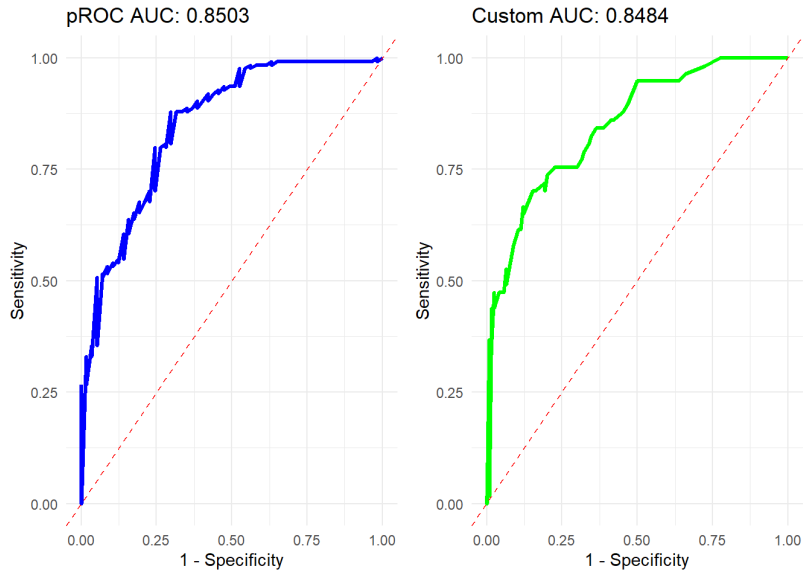
The Receiver Operating Characteristic (ROC) curve plots the True Positive Rate (Sensitivity) against the False Positive Rate (1 - Specificity). The Area Under the Curve (AUC) quantifies the overall performance to be stron.

Custom Function

Using a custom ROC function, we calculated the AUC to be 0.8489, which reflects a strong model performance.

11. Comparison with pROC

Using the pROC package, the AUC was also 0.8503, confirming that the custom function aligns closely with pROC's results.



Gain and Lift Charts

Gain Chart

The Gain Chart shows how well the model identifies true positives as more of the population is sampled. The chart indicates that the top 25% of the sample captures about 75% of the positives, showing that the model is more effective than random guessing.

Lift Chart

The Lift Chart shows the improvement over random selection. A lift of 3 at the start means the model is three times better at identifying positives than random selection, and it converges toward 1 as the sample size increases.

Kolmogorov-Smirnov (K-S) Chart

The K-S Chart shows the K-S statistic of 0.58, indicating the model effectively separates positive and negative classes. The sensitivity curve rises sharply, meaning the model captures positives early, while the false positive rate rises more slowly.

Caret Package Results

Using the caret package's confusionMatrix, sensitivity, and specificity functions:

- **Sensitivity** (using caret): 0.4737
- **Specificity** (using caret): 0.9597

The results are consistent with the custom functions when the class levels are aligned.

Conclusion

This report presents a comprehensive evaluation of the classification model using various metrics and graphical analysis techniques. Both custom functions and external packages (caret, pROC) were used to confirm the results. The model demonstrates strong performance, particularly in specificity and overall accuracy, with areas for improvement in recall and F1 score.