ROC2025 - International Workshop on ROC Analysis and Related Topics
Vigo, January 23 and 24, 2025

AUC optimism correction with missing data

Susana Rafaela Martins[1], María del Carmen Iglesias-Pérez[2] and Jacobo de Uña-Álvarez[2]

[1]Escola Superior de Desporto e Lazer, Instituto Politécnico de Viana do Castelo, Portugal

[2]CITMAga, SiDOR Research Group and Department of Statistics and Operations Research, Universidade de Vigo, Spain

Abstract

The Area Under the ROC Curve (AUC) plays an important role in the study of the predictive capacity of regression models. It is well known that an inflated AUC may result when the same data are used for training and testing the model. In this paper the correction for the optimism of the AUC in the presence of missing data is investigated. More specifically, split-sample, K-fold cross-validation and leave-one-out cross-validation are adapted to missing data under MCAR and MAR assumptions to introduce optimism corrections. Complete case analysis, inverse probability weighting and multiple imputation are employed to address the issue of missing data. The methods are compared through intensive Monte Carlo simulations in the particular setting of logistic regression. Results suggest that all the methods perform well, leave-one-out cross-validation being generally the best. Among the several strategies to cope with missing data, multiple imputation is recommended. A real data illustration is provided.