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

The gROC curve and the optimal classification

Pablo Martínez-Camblor[1]

[1]Department of Anesthesiology and Biomedical Data Science Department, Geisel School of Medicine at Dartmouth, NH, USA

Abstract

The binary classification problem (BCP) aims to correctly allocate subjects in one of two possible groups. The groups are frequently defined as having or not one characteristic of interest. With this goal, we are allowed to use different types of information. There is a huge number of methods dealing with this problem; including standard binary regression models, or complex machine learning techniques such as support vector machine, boosting, or perceptron, among others. When this information is summarized in a continuous score, we have to define classification regions (or subsets) which will determine whether the subjects are classified as positive, with the characteristic under study, or as negative, otherwise. The standard (or regular) receiver-operating characteristic (ROC) curve considers classification subsets in the way $[c, \infty)$ ($c \in \mathbb{R}$), and plots the true- against the false- positive rates (sensitivity against one minus specificity). The so-called generalized ROC curve, gROC, allows that both higher and lower values of the score were associated with higher probabilities of being positive. Besides, the efficient ROC curve, eROC, considers the optimal use of the scores without considering the potential impact on the associated classification subsets. In this document, we are interested in studying, comparing and approximating the transformations leading to the eROC and to the gROC curves. We will prove that, when the optimal transformation does have no relative maximum, both curves are equivalent. Besides, we investigate the use of the gROC curve on some theoretical models, explore the relationship between the gROC and the eROC curves, and propose two non-parametric procedures for approximating the transformation leading to the gROC curve. The finite-sample behavior of the proposed estimators is explored through Monte Carlo simulations. Two real-data sets illustrate the practical use of the proposed methods.