Area under ROC curve – review and efficient computation in R

by Błażej Kochański, Przemysław Pepliński, Miriam Nieslona, Wiktor Galewski, and Piotr Geremek

Abstract The AUC (Area Under the Curve) measure is widely used in statistical classification and machine learning, including credit scoring, where it is employed to assess the quality of predictive models. The goal of this paper is to review methods for calculating the AUC measure, followed by an analysis of the efficiency of computing this measure in R.

1 Introduction

ROC curves...

2 Background

3 AUC – alternative names and formulas

–> ** O czym trzeba wspomnieć: ** <–

AUC is referred to as:

- C-statistic
- a version of estimator of the common language effect size statistic
- probability of superiority
- "Vargha & Delaney A" ("measure of stochastic superiority") (Vargha and Delaney, 2000)
- "relative treatment effect" / "stochastic superiority statistic" in the Brunner-Munzel test

Gini coefficient is also referred to as:

- pseudo Gini (Idczak, 2019)

- a version of the rank-biserial correlation coefficient

- accuracy ratio based on the Cumulative Accuracy Profile

- a special case of the Somers' D statistic

- Cliff's delta (Cliff, 1993)

- Relation to the U statistic in the Mann-Whitney.

- Relation to the Somers' D.

- Cumulative Accuracy profile.

4 The need for computational efficiency in AUC estimation

Opisać:

- AUC / Gini bootstrapping / permutation tests

    - including DALEX (?)

- AUC optimization algorithms
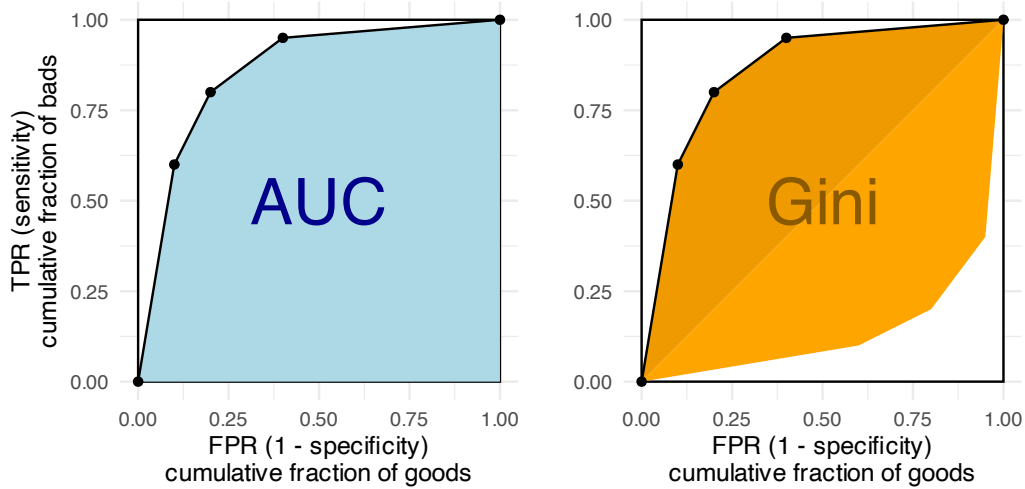
- credit market simulation

(Kochanski, 2021)

- inne ...

Figure 1: Geometric interpretation: AUC (AUROC) is the area under the ROC curve, Gini is twice the area between the diagonal y=x and the ROC curve.

5 R packages

–> Tutaj należy wstawić tabelkę pana Wiktora <–

We identified three main types of algorithms for AUC computation: (1) trapezoidal integration over the ROC Curve, (2) (optimized) pairwise comparison, (3) and rank-based (Mann–Whitney U statistic formulation).

Let (s_i, y_i) for $i = 1, \ldots, n$ denote the score assigned to an account and its corresponding true label. In line with standard practice in credit scoring, we assume that $y_i = 1$ indicates a "bad" account (e.g., one that defaults, doesn't repay the loan), while $y_i = 0$ represents a good account. A properly functioning scoring should assign lower scores to accounts with a higher predicted probability of being bad, and higher scores to those likely to be good.

Let $n_1 = \sum_{i=1}^{n} \mathbb{I}(y_i = 1)$ denote the number of bad accounts, and $n_0 = \sum_{i=1}^{n} \mathbb{I}(y_i = 0)$ – number of good accounts.

Trapezoidal Integration over the ROC Curve

$$\text{AUC} = \sum_{k=1}^{m-1} (\text{FPR}_{k+1} - \text{FPR}_k) \cdot \frac{\text{TPR}_{k+1} + \text{TPR}_k}{2}$$

where m is the number of distinct score thresholds from lowest to the highest score, TPR –> dokończyć <–

Optimized Pairwise Comparison

AUC as the probability that a randomly chosen positive instance receives a higher score than a randomly chosen negative instance:

$$\text{AUC} = \frac{1}{n_1 n_0} \sum_{i:y_i=1} \sum_{j:y_j=0} \left[\mathbb{I}(s_i < s_j) + \frac{1}{2}\mathbb{I}(s_i = s_j) \right]$$

Generally, naive pairwise comparison is not efficient (?), but . . .

Rank-Based (Mann–Whitney U Statistic Formulation)

$$\text{AUC} = \frac{\bar{R}_1 - n_1(n_1 + 1)/2}{n_0}$$

where $\bar{R}_1$ is the mean rank for s_i where $y_i = 1$:

$$\bar{R}_1 = \frac{1}{n_1} \sum_{i:y_i=1} \text{Rank}(s_i)$$

6 Efficiency study

7 Case studies

–> studieS, jeżeli będzie więcej <–

DALEX package

–> Wykres generuje się długo, a możemy go przyśpieszyć <–

8 Summary

Bibliography

N. Cliff. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 114 (3):494–509, 1993. ISSN 1939-1455. doi: 10.1037/0033-2909.114.3.494. [p1]

A. P. Idczak. Remarks on statistical measures for assessing quality of scoring models. *Acta Universitatis Lodziensis. Folia Oeconomica*, 4(343343):21–38, 2019. ISSN 2353-7663. doi: 10.18778/0208-6018.343.02. [p1]

B. Kochanski. A simulation model for risk and pricing competition in the retail lending market. *Czech Journal of Economics and Finance*, 71(2):96–118, Oct. 2021. ISSN 2464-7683. doi: 10.32065/CJEF.2021. 02.01. [p1]

A. Vargha and H. D. Delaney. A critique and improvement of the "cl" common language effect size statistics of mcgraw and wong. *Journal of Educational and Behavioral Statistics*, 25(2):101–132, 2000. ISSN 1076-9986. doi: 10.2307/1165329. [p1]

Błażej Kochański

Przemysław Pepliński
Gdańsk University of Technology
Faculty of Management and Economics
Gdańsk

Miriam Nieslona

Wiktor Galewski

Piotr Geremek