Abstract

Annual conference of the Canadian Econometrics Study Group

"Calibration of Probabilistic Scores of Classifiers" - December 2024

In binary classification tasks, accurate probabilistic predictions are essential for applications like credit default prediction and medical risk assessment. Traditional performance metrics, such as accuracy, often overlook model uncertainty and can misalign predicted scores with actual probabilities, especially in sensitive domains like finance or healthcare. Our study emphasizes the importance of calibration, highlighting the limitations of traditional metrics when score distributions deviate from the underlying data. We propose optimizing alignment between predicted and true probability distributions, focusing on minimizing Kullback-Leibler (KL) divergence. Through extensive empirical analysis using Random Forest and XGBoost models across 10 UCI datasets, we demonstrate that KL-based optimization achieves superior calibration without significant performance loss. Additionally, we introduce the Local Calibration Score and advocate for local regressions as effective recalibration methods that improve interpretability and visualization. These insights are applied in real-world scenarios, underscoring the importance of calibration in enhancing model reliability for critical decision-making.