Abstract

Censored data frequently occurs in fields such as medical research and survival analysis, posing unique challenges for reliable variance estimation in predictive models. This study addresses these challenges by developing and evaluating a novel variance estimation method tailored to predictions from IPC-weighted classification models.

In this thesis, we develop the Infinitesimal-Jackknife-after-weighted-Bootstrap-unbiased (IJK-AWB-U) estimator. Building upon Wager's Infinitesimal Jackknife approach for unweighted bagged learners, the IJK-AWB-U estimator extends this methodology to IPC-weighted resampling and incorporates an effective bias correction to adjust for finite bootstrap samples. This novel estimator provides unbiased variance estimates for bagged learners, particularly those based on decision trees, when dealing with censored data.

An extensive simulation study was conducted following the ADEMP framework to compare the performance of the IJK-AWB-U estimator with traditional methods, including the nonparametric Bootstrap and the Jackknife-after-Bootstrap. The results demonstrated that the IJK-AWB-U estimator offers reliable and accurate variance estimates, especially under low to moderate censoring proportions and with larger training sample sizes. It effectively corrects the bias present in the original IJK-AWB estimator and achieves a favorable balance between accuracy and computational efficiency. Compared to the computationally intensive Bootstrap estimator, the IJK-AWB-U estimator provides similar accuracy with significantly reduced computational time. In contrast, while the Jackknife-after-Bootstrap estimator is unbiased under unweighted resampling, it consistently overestimated variance in IPC-weighted contexts, making it less reliable in such settings. Given these findings, the IJK-AWB-U estimator emerges as the preferred method for variance estimation in our context

The practical application of the IJK-AWB-U estimator to the TxReg dataset further validated its reliability and utility. The estimator produced confidence intervals closely aligned with those generated by the Bootstrap estimator, demonstrating its effectiveness in real-world scenarios where computational efficiency is crucial.

Future research directions include extending the IJK-AWB-U estimator to other machine learning architectures such as gradient-boosting models and neural networks, and applying it to diverse fields beyond medical research, including finance, engineering, and epidemiology. These extensions would further validate the estimator's versatility and adaptability.

Keywords: IPC-weighted Classification Models, Censored Data, Infinitesimal Jackknife, Jackknife, Jackknife-after-Bootstrap, Nonparametric Bootstrap, Survival Analysis, Bagged Learner, Decision Trees, Uncertainty Estimation