

Abstract

On behalf of **Niko Beerenwinkel**, we kindly invite you to the Special Seminar:

Prof. Dr. Andreas Ziegler

Cardio-CARE Davos

Calibrating machine learning approaches for probability estimation: a comprehensive comparison

Prediction models have gained popularity in applied research. One challenge is the transfer of the prediction model to a different population which may be structurally different from the model for which it has been developed. An adaptation to the new population can be achieved by calibrating the model to the characteristics of the target population, for which numerous calibration techniques exist. In view of this diversity, we performed a systematic evaluation of various popular calibration approaches used by the statistical and the machine learning communities for estimating two-class probabilities. We present the results of a comprehensive simulation study. The calibration approaches are compared with respect to their empirical properties and relationships, their ability to generalize precise probability estimates to external populations and their availability in terms of easy-to-use software implementations. Calibration methods that estimated one or two slope parameters in addition to an intercept consistently showed the best results in the simulation studies. Calibration on logit transformed probability estimates generally outperformed calibration methods on non-transformed estimates. In case of structural differences between training and validation data, re-estimation of the entire prediction model should be outweighed against sample size of the validation data. We recommend regression-based calibration approaches using transformed probability estimates, where at least one slope is estimated in addition to an intercept for updating probability estimates in validation studies.