Probabilistic Scoring for Unbalanced Multi-Class Classifiers: Enhancing Calibration with Nested Dichotomies

Agathe Fernandes Machado

Université du Québec à Montréal (UQAM)
201 Av. du Président-Kennedy, Montréal, QC H2X 3Y7
Canada

Joint work with: Ana-Maria Patrón Piñerez, Arthur Charpentier, Ewen Gallic

In multi-class prediction, calibration is essential for accurately interpreting predicted scores as probabilities, thereby expanding the classification task beyond label prediction. Calibration metrics and "post-hoc" calibration techniques, originally designed for binary problems, have been adapted for multi-class contexts; however, they still face challenges, especially in unbalanced situations with underrepresented classes. To address these issues, we propose using nested dichotomy algorithms. The multi-class problems is simplified, decomposed into binary ones in a tree-like structure, allowing for a stronger emphasis on minority classes. Through experiments using simulated unbalanced data, we will compare the performance and calibration of nested dichotomies versus multi-class models, assessing the effects of post-calibration techniques. Finally, we will apply this methodology to real insurance data for race prediction, which includes minority groups, in accordance with Colorado legislation SB21-169.