Nested dichotomies (NDs) are used as classifiers for multi-class problems in machine learning [1, 2]. Based on a recursive binary partitioning of the complete set of classes, they reduce the original multi-class problem to a set of binary problems, for which any (probabilistic) binary classifier can be used. In practice, the performance of a dichotomy may strongly depend on its structure, i.e., the way in which classes are partitioned. Therefore, NDs are often used as an ensemble technique, taking advantage of averaging effects: Several structures are randomly sampled from the space of all ND topologies, the corresponding models are trained, and their predictions are combined using techniques such as majority voting.

The goal of this thesis is to elaborate on optimal sampling strategies, and to compare simple uniform sampling with biased variants. In fact, there is evidence that uniform sampling might not be optimal, especially as it often produces quite imbalanced structures. Therefore, alternative strategies should be developed, analyzed, and experimentally evaluated in terms of their performance (predictive accuracy) on benchmark data sets.

Requirements: Implementation of nested dichotomies; design of sampling strategies; empirical evaluation of such strategies.

Prerequisites: Basic knowledge in machine learning; programming skills.

For more information, please contact Vitalik Melnikov (melnikov@mail.upb.de) or Prof. Eyke Hüllermeier (eyke@upb.de).

References
