Bayesian Optimization for Wrapper Feature Selection

Feature selection is an important step in machine learning pipelines. We consider feature selection for classification tasks. Training classifiers with fewer features is faster, and the models are more comprehensible. Furthermore, classifier performance might improve if unnecessary features are excluded. There is a rich variety of existing feature selection algorithms. An important category are wrapper feature selection approaches, which search the space of feature subsets and evaluate each subset with a classifier. An appropriate search strategy is vital to obtain good results, as the search space grows exponentially: For \( n \) features, there are \( 2^n \) combinations. Existing search strategies include simple approaches like greedy forward selection, but also general-purpose optimization algorithms like genetic algorithms or simulated annealing.

Bayesian optimization is a technique that has gained momentum in the machine learning community over the last few years. It allows optimizing so-called black-box functions whose overall shape might be unknown, and where computation of single values is expensive. Classifiers in machine learning are such black-box functions regarding their parametrization. Researchers and practitioners have already applied Bayesian optimization to hyperparameter tuning and model selection for classification. This raises the question if one could also use Bayesian optimization as search strategy for feature selection.

The goal of this thesis is to study the use of Bayesian optimization for wrapper feature selection. The following questions are particularly interesting:

- Which adaptations are necessary or beneficial to use Bayesian optimization for feature selection?
- How does Bayesian optimization compare to existing wrapper feature selection methods?
- Is there a benefit compared to classification approaches with built-in feature selection?
- Can we provide guidelines under which settings one should use Bayesian optimization for feature selection?

The following steps are part of your thesis:

- Review literature about (wrapper) feature selection and Bayesian optimization.
- Design and implement an approach to use Bayesian optimization for wrapper feature selection. Depending on your choice of programming language, you might build on existing libraries for general-purpose Bayesian optimization.
- Design and implement experiments to evaluate wrapper feature selection for classification. You can use existing machine learning libraries as well as datasets from public repositories like the UCI Machine Learning Repository.
- Evaluate your approach experimentally. You can use the server infrastructure of our chair.

During your work on this thesis, you will acquire practical knowledge about state-of-the-art machine learning libraries. You will get familiar with Bayesian optimization as well as feature selection and gain an understanding of their usefulness as well as limitations for different kinds of datasets. You will gain experience in running and evaluating large-scale scientific experiments.

The scope of the topic can be adapted to a Bachelor as well as a Master thesis. You can write the thesis in English or German. Prior experience with classification in any programming language (e.g. Python, R) is beneficial, but not necessary.

Contact
Jakob Bach jakob.bach@kit.edu +49 721 608-47339 Room: 351
Am Fasanengarten 5 76131 Karlsruhe Building: 50.34