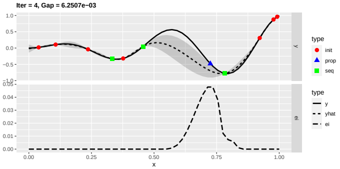
Introduction to Machine Learning

Hyperparameter Tuning - Advanced Tuning Techniques: MBO & Hyperband



Learning goals

- Understand the idea of model based optimization
- Be able to explein the terms 'surrogate model' and 'expected improvement'
- Understand the idea of hyperband

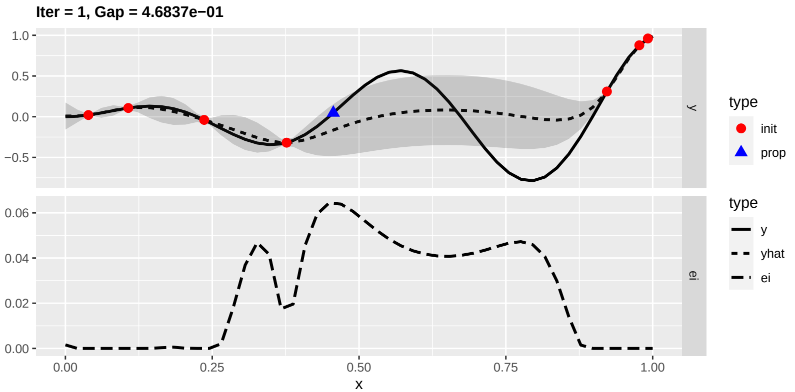
MODEL-BASED OPTIMIZATION

Model-based optimization (MBO) is a sequential optimization procedure. We start with an initial design, i.e., a set of configurations λ_i where we have evaluated the corresponding (resampling) performance.

Repeat:

1. From the available performance measurements, we build a **surrogate model** that models the relationship between model hyperparameters and estimated generalization error. It serves as a cheap approximation of the expensive objective.
2. Based on information provided by the surrogate model, a new configuration $\lambda^{(\text{new})}$ is proposed: we pick a value for which the surrogate model predicts a large potential improvement over the already evaluated configurations.
3. The resampling performance of the learner with hyperparameter setting $\lambda^{(\text{new})}$ is evaluated and added to the set of design points.

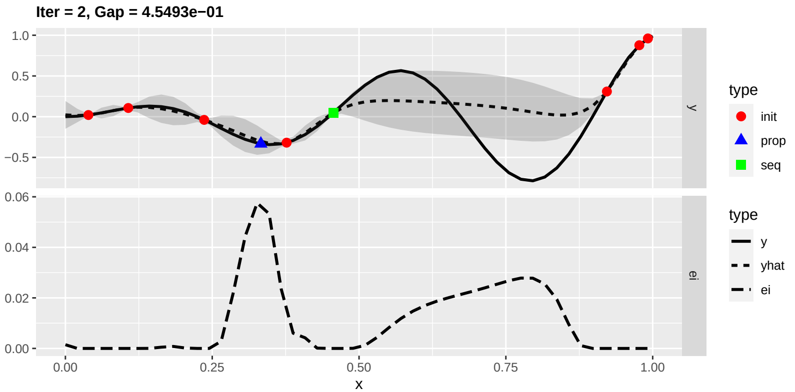
MODEL-BASED OPTIMIZATION



Upper plot: The surrogate model (black, dashed) models the *unknown* relationship between input and output (black, solid) based on the initial design (red points).

Lower plot: Mean and variance of the surrogate model are used to derive the expected improvement (EI) criterion. The point that maximizes the EI is proposed (blue point).

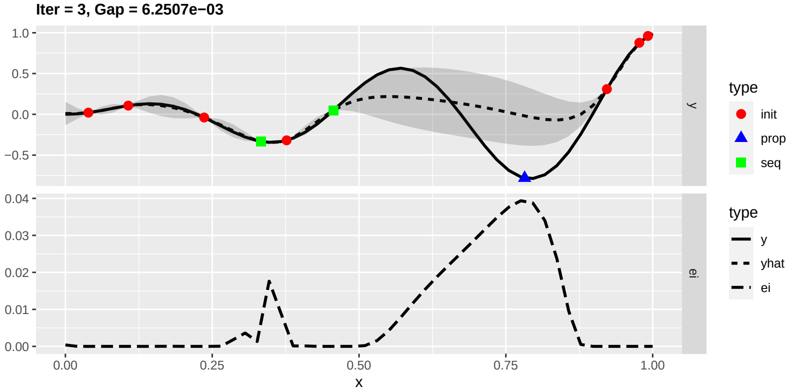
MODEL-BASED OPTIMIZATION



Upper plot: The surrogate model (black, dashed) models the *unknown* relationship between input and output (black, solid) based on the initial design (red points).

Lower plot: Mean and variance of the surrogate model are used to derive the expected improvement (EI) criterion. The point that maximizes the EI is proposed (blue point).

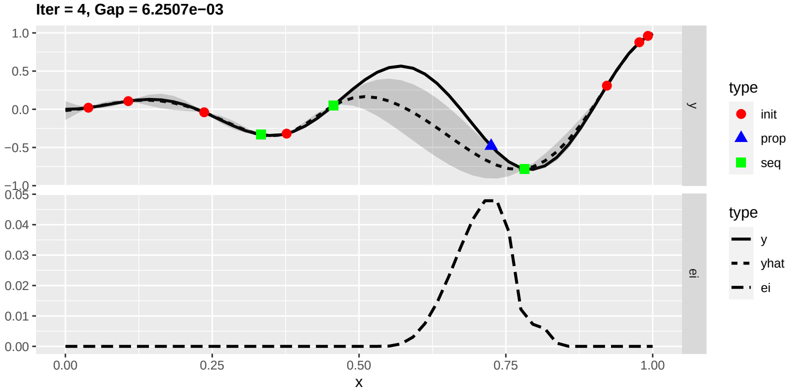
MODEL-BASED OPTIMIZATION



Upper plot: The surrogate model (black, dashed) models the *unknown* relationship between input and output (black, solid) based on the initial design (red points).

Lower plot: Mean and variance of the surrogate model are used to derive the expected improvement (EI) criterion. The point that maximizes the EI is proposed (blue point).

MODEL-BASED OPTIMIZATION



Upper plot: The surrogate model (black, dashed) models the *unknown* relationship between input and output (black, solid) based on the initial design (red points).

Lower plot: Mean and variance of the surrogate model are used to derive the expected improvement (EI) criterion. The point that maximizes the EI is proposed (blue point).

MODEL-BASED OPTIMIZATION

Since we use the sequentially updated surrogate model predictions of performance to propose new configurations, we are guided to “interesting” regions of $\mathbf{\Lambda}$ and avoid irrelevant evaluations:

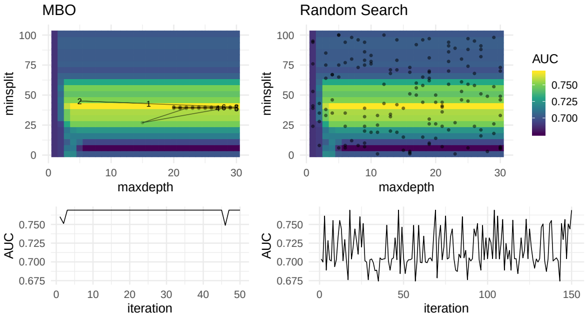


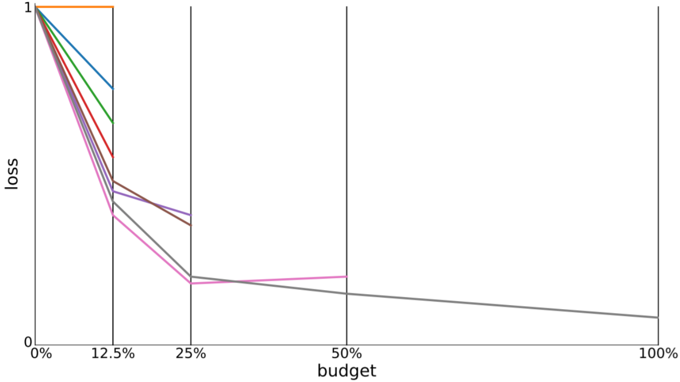
Figure: Tuning tree depth and minimal node size for splits for CART on the `sonar` data (10-fold CV maximizing AUC).
Left panel: random search, 150 configurations; right panel: MBO, 50 iterations.

HYPERBAND

- It is extremely expensive to train complex models on large data sets
- For many configurations, it might be clear early on that further training is not likely to significantly improve the performance
- More importantly, the relative ordering of configurations (for a given data set) can also become evident early on.
- **Idea:** "weed out" poor configurations early during training
- One approach is **successive halving**: Given an initial set of configurations, all trained for a small initial budget, repeat:
  - Remove the half that performed worst, double the budget
  - Continue until the new budget is exhausted
- Successful halving is performed several times with different trade-offs between the number of configurations considered and the budget that is spent on them.

HYPERBAND

Only the most promising configuration(s) are trained to completion:



taken from Hutter, Kotthoff, Vanschoren. *Automated Machine Learning – Methods, Systems, Challenges*. Springer, 2019. (Fig. 1.3)

MORE TUNING ALGORITHMS:

Other advanced techniques besides model-based optimization and the hyperband algorithm are:

- Stochastic local search, e.g., simulated annealing
- Genetic algorithms / CMAES
- Iterated F-Racing
- Many more . . .