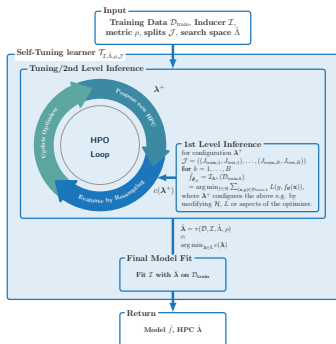
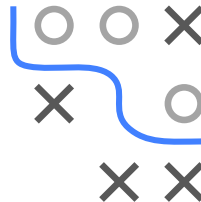


Introduction to Machine Learning

Hyperparameter Tuning Practical Aspects



Learning goals

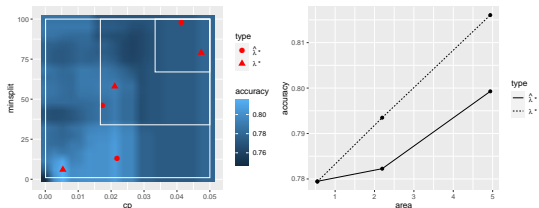
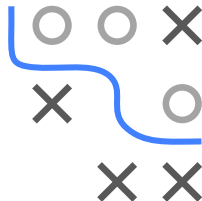
- Understand the possible design choices for HPT
- Know termination criteria of HPT

A 3x3 grid with a blue path starting at the top-left cell (0,0) and ending at the bottom-right cell (2,2). The path consists of the following cells: (0,0), (0,1), (1,1), (1,2), and (2,2). The cells (0,2), (1,0), and (2,0) are empty. The cells (0,1), (1,1), and (1,2) contain a grey 'X'. The cells (1,0) and (2,0) contain a grey circle.

-
- Figure 1 displays the performance of different resampling methods, comparing Mean Coverage Error (MCE) and GE true values across various parameters.
- The top row shows MCE for **CV-2** (left) and **GE_true** (right). The y-axis represents **msplit** (0 to 100) and the x-axis represents **cp** (0.000 to 0.100). The color scale for MCE ranges from 0.45 to 0.60 for CV-2 and 0.36 to 0.44 for GE_true.
- The bottom-left panel shows the difference in MCE, **CV-2 - GE_true**, with a color scale ranging from 0.00 to 0.20.
- The bottom-right panel is a boxplot showing the distribution of **GE true** values for different numbers of resampling folds (1 to 7). The y-axis ranges from 0.32 to 0.44.

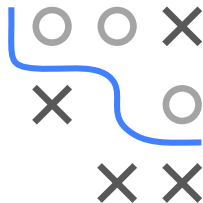
PRACTICAL ASPECTS OF HPO

- Choosing performance measure
 - Desired implications when applying the model in practice
- Choosing a pipeline and search space
 - Numeric HPs of arbitrary size should be tuned on log scale
 - Size of search space results in different trade-offs:
 - too small may miss out well performing HPCs;
 - too large makes optimization more difficult



PRACTICAL ASPECTS OF HPO

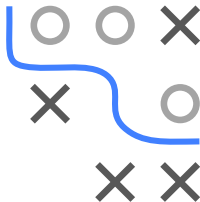
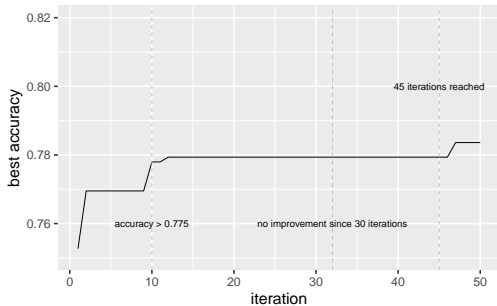
- Choosing HPO algorithm
 - For few HPS (1-3), grid search can be used
 - BO with GPs for upto 10 numeric HPs
 - BO with RFs handle mixed HP spaces
 - Random search and Hyperband work well as long as the “effective” dimension is low
 - EAs are somewhat in-between BO and RS, can handle very complex spaces, but less sample efficient than BO
 - **Also: use something that’s stable and robust! More an aspect of the implementation than the algo!**



PRACTICAL ASPECTS OF HPO

When to terminate HPO

- Specify a certain amount of runtime/budget beforehand
- Set a lower bound regarding \widehat{GE}
- Terminate if performance improvement stagnates



Different stopping points while tuning CART on the `titanic` data depending on which termination criterion is used

PRACTICAL ASPECTS OF HPO

- Warm starts
 - Evaluations (e.g., weight sharing of neural networks)
 - Optimization (initializing with HPCs that worked well before)
- Control of execution
 - Parallelizability of HPO algorithms differs strongly
 - HPO execution can be parallelized at different levels (outer resampling, iteration, evaluation, inner resampling, model fit)

