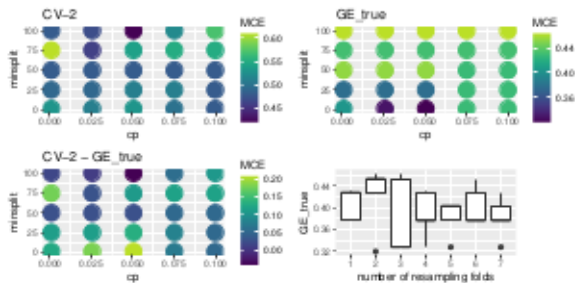


PRACTICAL ASPECTS OF HPO

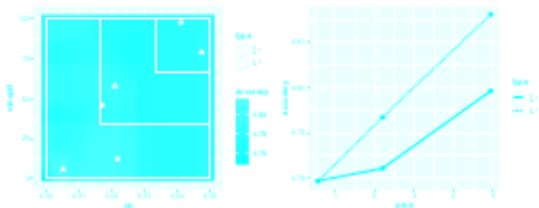
- Choosing resampling
 - Nr of observations, i.i.d assumption for data sampling process
 - Higher resampling rates likely result in a better model; however they are computationally more expensive



Tuning a CART on the spirals data with a k-fold CV (k=1 means here a 2/3 holdout split) using grid search and estimating the true GE with a very large test set (5 repetitions)

PRACTICAL ASPECTS OF HPO / 2

- Tuning a CART on the spirals data with a k-fold CV (k=1 means here a 2/3 holdout split) using grid search and estimating the true GE with a very large test set (5 repetitions)
- 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



Tuning cp and minsplit for a CART on the titanic data over 3 increasing rectangular search spaces with random search (candidates number fixed) and comparing the result with the optimal model (found with exhaustive grid search)

PRACTICAL ASPECTS OF HPO

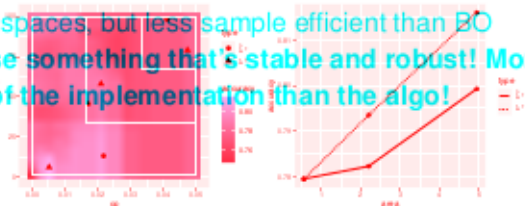
- Choosing performance measure

- 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
- 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!

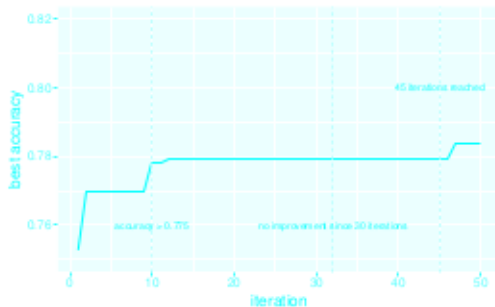


PRACTICAL ASPECTS OF HPO / 2

When to terminate HPO

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

Tuning `cp` and `minsplit` for a CART on the `titanic` data over 3 increasing rectangular search spaces with random search (candidates number fixed) and comparing the result with the optimal model (found with exhaustive grid search)



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



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 (not worked well before)
 - BO with RFs handle mixed HP spaces
- Control of execution
 - Random search and Hyperband work well as long as the "effective" dimension is low
 - Parallelizability of HPO algorithms differs strongly
 - HPO execution can be parallelized at different levels (outer resampling, iteration, evaluation, inner resampling, model fit)
 - 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!**

