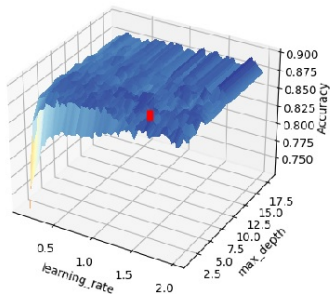


Optimization in Machine Learning

Optimization Problems

Other optimization problems



(a) vehicle

Learning goals

- Discrete / feature selection
- Black-box / hyperparameter optimization
- Noisy
- Multi-objective

OTHER CLASSES OF OPTIMIZATION PROBLEMS

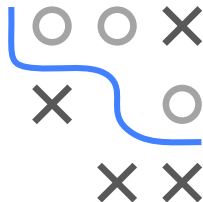
So far: “nice” (un)constrained problems:

- Problem defined on continuous domain \mathcal{S}
- Analytical objectives (and constraints)

Other characteristics:

- Discrete domain \mathcal{S}
- f **black-box**: Objective not available in analytical form; computer program to evaluate
- f **noisy**: Objective can be queried but evaluations are noisy
 $f(\mathbf{x}) = f_{\text{true}}(\mathbf{x}) + \epsilon, \quad \epsilon \sim F$
- f **expensive**: Single query takes time / resources
- f multi-objective: $f(\mathbf{x}) : \mathcal{S} \rightarrow \mathbb{R}^m, f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$

These make the problem typically much harder to solve!



EXAMPLE 2: WRAPPER FEATURE SELECTION

Subset sel. can be generalized to any learner \mathcal{I} using only features \mathbf{s} :

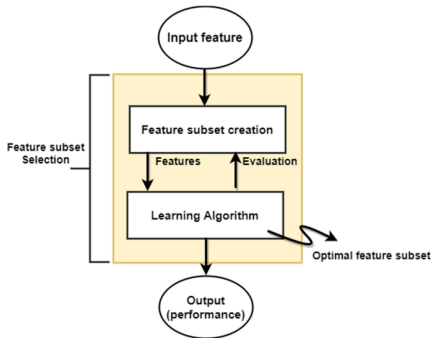
$$\min_{\mathbf{s} \in \{0,1\}^p} \widehat{\text{GE}}(\mathcal{I}, \mathcal{J}, \rho, \mathbf{s}),$$

$\widehat{\text{GE}}$ general. err. with metric ρ and estim. with resampling splits \mathcal{J}



Problem characteristics:

- black box
eval by program
- \mathcal{S} is discrete / binary
- expensive
1 eval: 1 or multiple ERM(s)!
- noisy
uses data / resampling
- NB: Less features can be better in prediction (overfitting)

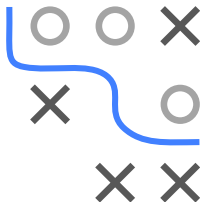


EXAMPLE 3: FEATURE SEL. (MULTIOBJECTIVE)

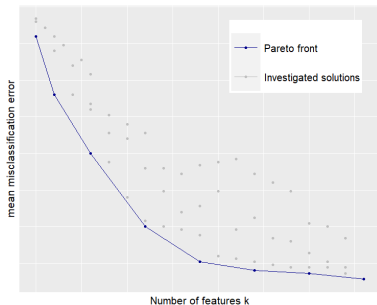
Feature selection is usually inherently multi-objective, with model sparsity as a 2nd trade-off target:

$$\min_{\mathbf{s} \in \{0,1\}^p} \left(\widehat{\text{GE}}(\mathcal{I}, \mathcal{J}, \rho, \mathbf{s}), \sum_{i=1}^p s_i \right).$$

$\widehat{\text{GE}}$ general. err. with metric ρ and estim. with resampling splits \mathcal{J}



- Multiobjective
- black box
eval by program
- S is discrete / binary
- expensive
1 eval: 1 or multiple ERM(s)!
- noisy
uses data / resampling

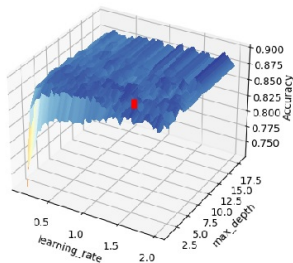
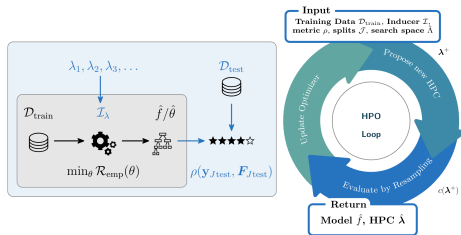
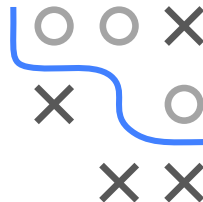


EXAMPLE 4: HYPERPARAMETER OPTIMIZATION

- Learner \mathcal{I} usually configurable by hyperparameters $\lambda \in \Lambda$.
- Find best HP configuration λ^*

$$\lambda^* \in \arg \min_{\lambda \in \Lambda} c(\lambda) = \arg \min \widehat{GE}(\mathcal{I}, \mathcal{J}, \rho, \lambda)$$

\widehat{GE} general. err. with metric ρ and estim. with resampling splits \mathcal{J}



(a) vehicle

XGBoost HP landscape; source:

ceur-ws.org/Vol-2846/paper22.pdf

EXAMPLE 4: HYPERPARAMETER OPTIMIZATION

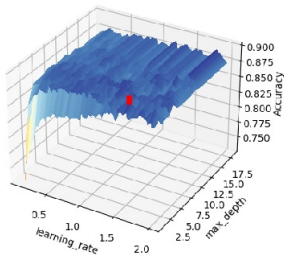
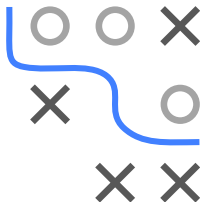
/ 2

Solving

$$\lambda^* \in \arg \min_{\lambda \in \Lambda} c(\lambda)$$

is very challenging:

- c black box
eval by program
- expensive
1 eval: 1 or multiple ERM(s)!
- noisy
uses data / resampling
- the search space Λ might be mixed
continuous, integer, categ. or hierarchical



(a) vehicle

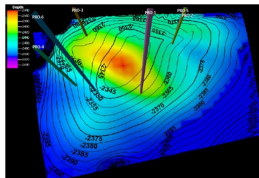
XGBoost HP landscape; source:

ceur-ws.org/Vol-2846/paper22.pdf

MORE BLACK-BOX PROBLEMS

Black-box problems from engineering: **oil well placement**

- The goal is to determine the optimal locations and operation parameters for wells in oil reservoirs
- Basic premise: achieving maximum revenue from oil while minimizing operating costs
- In addition, the objective function is subject to complex combinations of geological, economical, petrophysical and fluid dynamical constraints
- Each function evaluation requires several computationally expensive reservoir simulations while taking uncertainty in the reservoir description into account



Oil saturation at various depths with possible location of wells.

Source: [https://doi.org/10.1007/](https://doi.org/10.1007/s13202-019-0710-1)

s13202-019-0710-1

