

# White-box optimization from historical data

Sicco Verwer

S.VERWER@CS.RU.NL

Radboud University Nijmegen, P.O. Box 9010, 6500 GL, Nijmegen, The Netherlands

Qing Chuan Ye, Yingqian Zhang

{YE,YQZHANG}@ESE.EUR.NL

Erasmus University Rotterdam, Burg. Oudlaan 50, 3062 PA, Rotterdam, The Netherlands

**Keywords:** optimization, mathematical modeling, machine learning, auction design

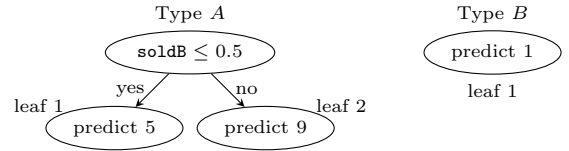
It is challenging to construct a mathematical model describing the properties of a system, especially when the structure of the system cannot be fully determined from the hypotheses at hand. In such cases, machine learning techniques can be used to replace (parts of) a mathematical decision model. However, the models produced by machine learning have so far only been used in a black-box fashion, e.g. as fitness functions or parameters. We propose a *white-box* optimization by mapping a learned regression tree model to a mixed integer linear program that can be used for optimization. Consequently, the learned model's properties are visible as constraints to a mathematical problem solver, that can then use sophisticated branching and cutting techniques on these constraints when finding solutions, which are impossible in black-box optimization.

We illustrate our approach using a sequential auction design problem. The objective is to maximize the expected revenue of the auctioneer with multiple bidders (agents) who have complementary preferences over items. We try to find the optimal *ordering of items* to maximize the expected revenue. This problem is proven to be NP-complete. We learn the overall preferences of the group of bidders from historical data by viewing the prediction of the revenue of an auction as a regression problem. We split this problem into the subproblems of predicting the revenue of the auctioned items, and then sum these up to obtain the overall objective function.

We use the following simple example to demonstrate the use of our method. Two agents  $a_1$  and  $a_2$  partake in a sequential auction of items  $A$  and  $B$ . Their valuations are given by  $v_{a_1}(A) = 1, v_{a_1}(B) = 1, v_{a_1}(\{A, B\}) = 10, v_{a_2}(A) = 5$ . Two past auctions are known:  $A$  was sold first to  $a_2$  and then  $B$  to  $a_1$  with a total revenue of 6, whereas in the second auction,  $B$  was sold first and then  $A$ , both to  $a_1$  with a total revenue of 10. We construct feature values from

two auctions, as shown in the table.  $\text{sold}(\cdot)$  represents how many items of type  $(\cdot)$  have been sold prior to the current item. For each of item types, we learn a regression tree (see the figure).

type	value	soldA	soldB	type	value	soldA	soldB
A	5	0	0	B	1	0	0
B	1	1	0	A	9	0	1



Suppose we are given the set of items  $\{A, B, B\}$  for which we need to find an optimal ordering. We can then use the learned trees to formulate the problem of finding an optimal ordering for this set of items as an integer linear program (ILP), which can then be solved by one of many ILP-solvers:

$$\begin{aligned}
 & \max \quad \sum_{1 \leq i \leq 3} 5z_{i,1,A} + 9z_{i,2,A} + 1z_{i,1,B} \\
 & \text{subject to} \\
 & \quad x_{i,A} + x_{i,B} = 1 \quad \text{for all } 1 \leq i \leq n \\
 & \quad x_{1,A} + x_{2,A} + x_{3,A} = 1 \\
 & \quad x_{1,B} + x_{2,B} + x_{3,B} = 2 \\
 & \quad \text{sold}_{1,B} = 0 \quad \text{sold}_{2,B} = x_{1,B} \quad \text{sold}_{3,B} = x_{1,B} + x_{2,B} \\
 & \quad y_{i,\text{sold}B > 0.5} < 1.0 + \frac{\text{sold}_{1,B}^{-0.5}}{100} \quad \text{for all } 1 \leq i \leq n \\
 & \quad y_{i,\text{sold}B > 0.5} \geq \frac{\text{sold}_{1,B}^{-0.5}}{100} \quad \text{for all } 1 \leq i \leq n \\
 & \quad y_{i,\text{sold}B \leq 0.5} = 1.0 - y_{i,\text{sold}B > 0.5} \quad \text{for all } 1 \leq i \leq n \\
 & \quad z_{i,1,A} \leq \frac{x_{i,A} + y_{i,\text{sold}B \leq 0.5}}{1 + 0.5} \quad \text{for all } 1 \leq i \leq n \\
 & \quad z_{i,2,A} \leq \frac{x_{i,A} + y_{i,\text{sold}B > 0.5}}{1 + 0.5} \quad \text{for all } 1 \leq i \leq n \\
 & \quad z_{i,1,B} \leq \frac{x_{i,B}}{0.5} \quad \text{for all } 1 \leq i \leq n
 \end{aligned}$$

Although optimizing the orderings in sequential auctions is a well-known hard problem, our method obtains very high revenues, significantly outperforming the greedy and random methods proposed in the literature. Our constructions are general and can be applied to any settings where regression trees can be learned from data, and their feature values can be computed as linear functions from solutions.