Aalto-yliopisto
Perustieteiden korkeakoulu

Training decision trees using mixed-integer optimisation

Presentation of the BSc thesis topic

Joel Vääräniemi

17.4.2023

Instructor: *Nikita Belyak*

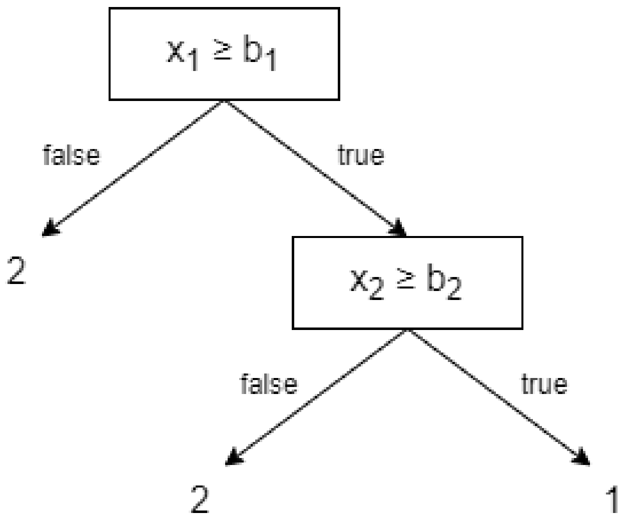
Supervisor: *Fabricio Oliveira*

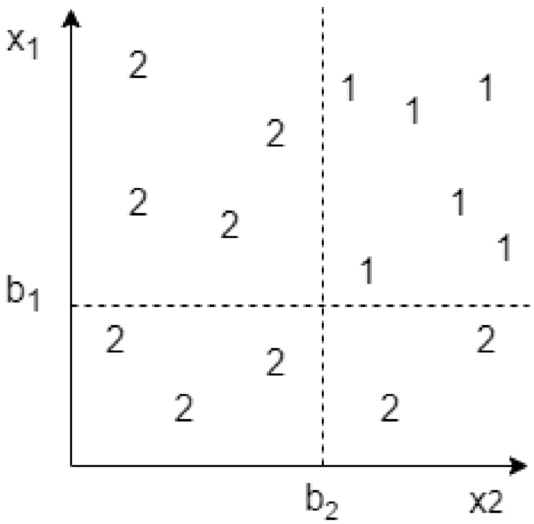
Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

Systeemianalyysin laboratorio

Decision trees

- Branch and leaf nodes
- Popular choice for classification problems
- Easy to interpret





Top-down method: Classification and regression trees (CART)

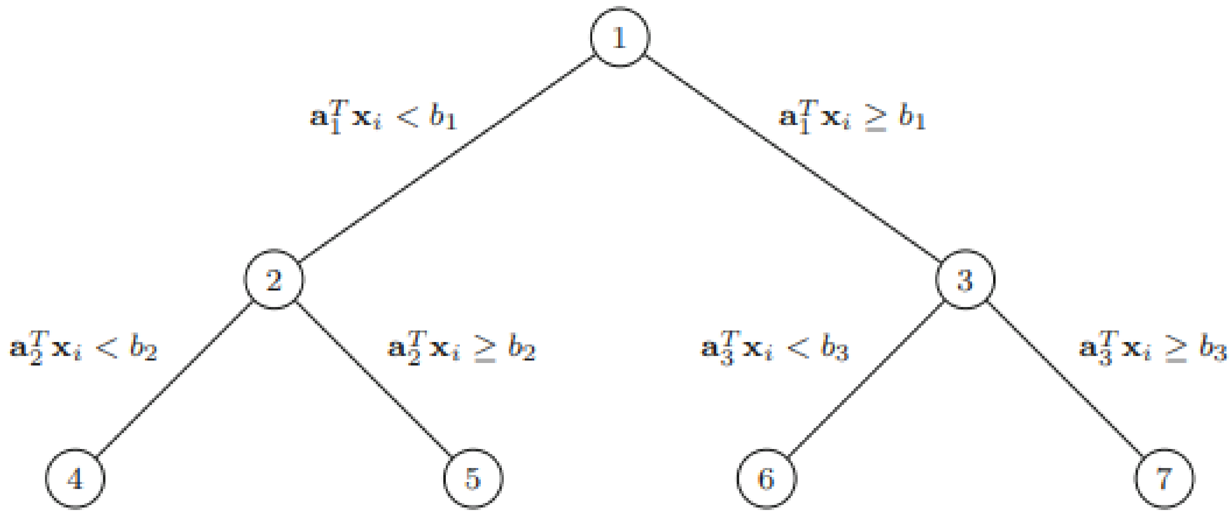
- Top-down approach, starting from the root node
- Every split is made in isolation without information about future splits
- Split determined by optimisation problem (minimising impurity score)
- Creates a binary tree (splits into two groups in every branch node)

MIP method: Optimal classification trees (OCT) model

- Creates the whole tree once with full knowledge of all possible splits
- Results in a globally optimal solution
- Construction is an NP-hard problem
  - However, reasonable with modern computing power and solvers

MIP formulation for OCT model

- Restriction: Only axis-aligned splits (taking only one dimension into account)
- 3 hyper-parameters: maximum tree depth, minimum leaf size, complexity parameter



Aims

- The goal is to implement the MIP decision tree formulation in Julia, using Gurobi as an optimiser
- The model will be tested and analysed on various datasets with different hyper-parameters
  - Focus on speed and accuracy
  - Datasets for example from Kaggle*
- Results will be compared to the top-down approach (CART)

*www.kaggle.com/datasets

Schedule

- 23.4. Code is implemented
- 30.4. Testing is completed and results are analysed
- 1.5. Start of the thesis writing
- 16.5 Thesis done

References

- Bertsimas, D., & Dunn, J. (2017). Optimal classification trees. Machine Learning, 106(7), 1039–1082.

Appendix: MIP formulation

$$\min \quad \frac{1}{\hat{L}} \sum_{t \in \mathcal{T}_L} L_t + \alpha \cdot C$$

s.t.

$$
\begin{aligned}
& L_t \geq N_t - N_{kt} - n(1 - c_{kt}), && \forall t \in \mathcal{T}_L, k \in [K], \\
& L_t \leq N_t - N_{kt} + n c_{kt}, && \forall t \in \mathcal{T}_L, k \in [K], \\
& L_t \geq 0, && \forall t \in \mathcal{T}_L, \\
& N_{kt} = \sum_{i : y_i = k} z_{it}, && \forall t \in \mathcal{T}_L, k \in [K], \\
& N_t = \sum_{i=1}^{n} z_{it}, && \forall t \in \mathcal{T}_L, \\
& \sum_{k=1}^{K} c_{kt} = l_t, && \forall t \in \mathcal{T}_L, \\
& C = \sum_{t \in \mathcal{T}_B} d_t, &&
\end{aligned}
$$

$$
\begin{aligned}
& \mathbf{a}_m^{\mathsf{T}} \mathbf{x}_i \geq b_m - (1 - z_{it}), && \forall i \in [n], t \in \mathcal{T}_L, m \in A_R(t), \\
& \mathbf{a}_m^{\mathsf{T}} (\mathbf{x}_i + \boldsymbol{\epsilon} - \epsilon_{min}) + \epsilon_{min} \\
& \qquad \leq b_m + (1 + \epsilon_{max})(1 - z_{it}), && \forall i \in [n], t \in \mathcal{T}_L, m \in A_L(t), \\
& \sum_{t \in \mathcal{T}_L} z_{it} = 1, && \forall i \in [n], \\
& z_{it} \leq l_t, && \forall t \in \mathcal{T}_L, \\
& \sum_{i=1}^{n} z_{it} \geq N_{min} l_t, && \forall t \in \mathcal{T}_L, \\
& \sum_{j=1}^{p} a_{jt} = d_t, && \forall t \in \mathcal{T}_B, \\
& 0 \leq b_t \leq d_t, && \forall t \in \mathcal{T}_B, \\
& d_t \leq d_{p(t)}, && \forall t \in \mathcal{T}_B \setminus \{1\}, \\
& z_{it}, l_t, c_{kt} \in \{0, 1\}, && \forall i \in [n], k \in [K], t \in \mathcal{T}_L, \\
& a_{jt}, d_t \in \{0, 1\}, && \forall j \in [p], t \in \mathcal{T}_B,
\end{aligned}
$$