Devising Multivariate Splits for Optimal Decision Trees

1. Decision Tree Learning

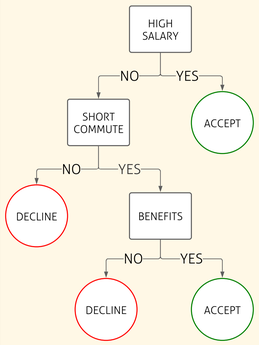
Uses a tree structure to map observations about an item to conclusions about its target value.

Internal nodes represent conditions on data features, leaves are final decisions.

Favoured for its easiness to interpret and understand.

DT = Decision Tree

e.g.: Job Offer DT



2. Heuristic vs Optimal DTs

Constructing optimal decision trees is NP-Complete problem

→ Heuristic DTs often used in practice:

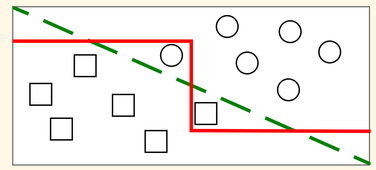
✓ Fast construction
✕ Maximum accuracy not guaranteed
✕ Difficult to add extra constraints

Optimal DTs address shortcomings, now feasible thanks to improved algorithms

3. Multi- vs Uni- Variate DTs

Univariate DT - internal nodes split on single feature (most common) → cheaper to construct

Multivariate DT - internal nodes split on combination of features → potentially closer to ground truth



e.g.: Distinguish between □ and ○

- - Multivariate Boundary

– Univariate Boundary

4. Is it possible to build multivariate optimal DTs?

- Are the increases in accuracy noticeable over a wide array of sample datasets?
- What is the further cost penalty of turning already costlier optimal trees multivariate?

5. Methodology

Search Space Increases Dramatically with Multivariate Splits

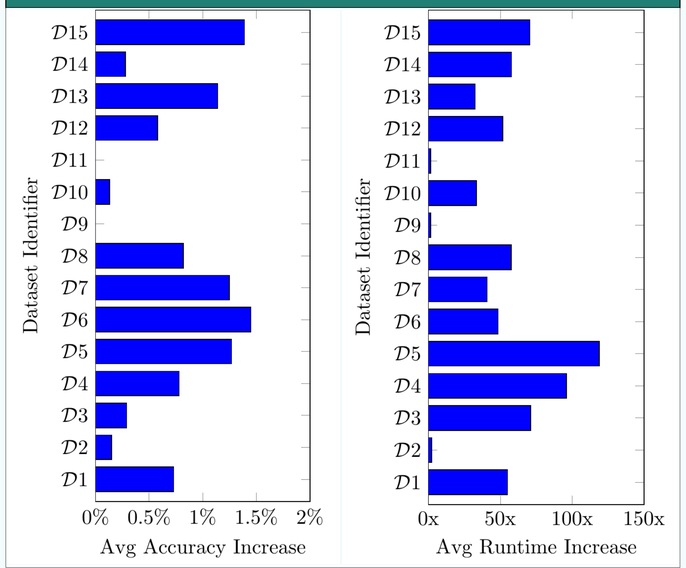
→ Apply extra constraints:

- Utilise strictly bivariate predicates
- Binary features → use Boolean ops (*AND*, *OR*, *XOR*)
- Limit multivariate splits to first α tree levels

Search space increase remains large nonetheless:

	α = 1	α = 2
f = 10	*14.5x*	*3048.62x*
f = 50	*74.5x*	*413493.62x*

6. Result Review (depth 4, α=1)



7. Conclusions & Further Work

Takeaways:

- Multivariate Optimal DTs feasible, but only with severe limitations
- Accuracy Improvements present, but not groundbreaking
- Orders of mangitude longer runtime prohibitive for real-world use

Worth Exploring:

- General implementation optimisations
- Multivariate-specific pruning techniques
- Different kinds of multivariate predicates

TU Delft

Abele Mălan
A.M.Malan@student.tudelft.nl

Supervisor:
Emir Demirović

July 2021