OPTIMIZING OPTIMAL REGRESSION TREES USING DYNAMIC PROGRAMMING

Author Mim van den Bos **Supervisor** Ir. J.G.M. van der Linden **Responsible Professor** Dr. Emir Demirović

mimvandenbos@gmail.com

1 INTRODUCTION

Finding an optimal regression tree is an NP-hard problem but recent algorithmic techniques and hardware can handle larger datasets.

Several algorithmic techniques exist to increase scalability for classification trees [1].

Target	Area > 100	Bedroom > 2
55000	0	0
58000	0	0
83000	1	0
99100	1	1
70000	0	1

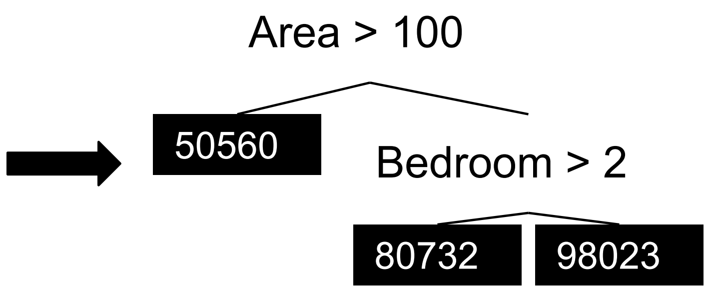

Figure 1: Example of finding a regression tree from training data

2 MAIN QUESTION

Can the scalability of optimal regression trees be improved by adapting methods for classification trees?

3 RESEARCH METHOD

- Show that Mean Squared Error with a penalty for the number of nodes is "seperable". (Can be solved independently for subtrees)
- Implement regression in the STreeD [2] framework with a regularization term.
- Adapt the specialized depth two algorithm to work for regression.
- Implement lower bounds from a previous paper [3].
  - Equivalent Points
  - k-Means Equivalent Points
- Find a novel upper bound on the contribution of a single instance for use with the similarity lower bound.
- Compare the runtime to OSRT [3], a state of the art method.

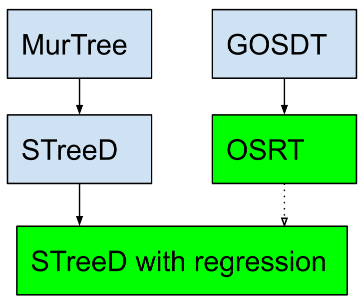

Figure 2: Evolution of optimal methods. Green is regression

4 RESULTS & CONCLUSION

- 12 Datasets are tested on various maximum tree depths and regularization weights for a reliable comparison.
- Experiments were run on the DelftBlue supercomputer.
- Experiments show an order of magnitude speed improvement over OSRT.
- Experiments show the lower bounds are effective in pruning suboptimal solutions.
- Two improvements to STreeD during development.
- Future work can inspect the performance difference for datasets with more than 1 million instances.

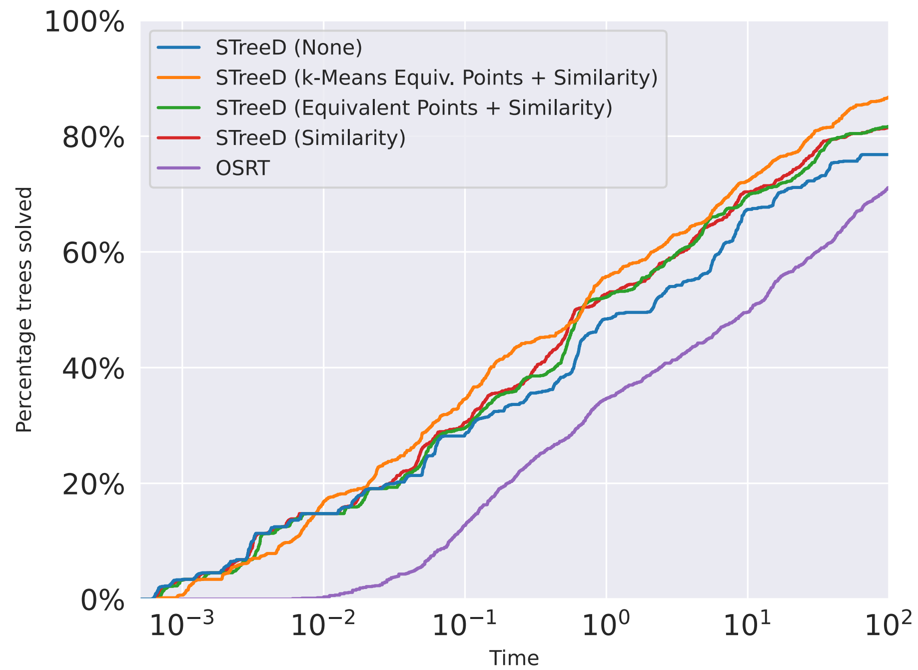

Figure 3: Number of different trees solved within an amount of time. Note the exponential axis.

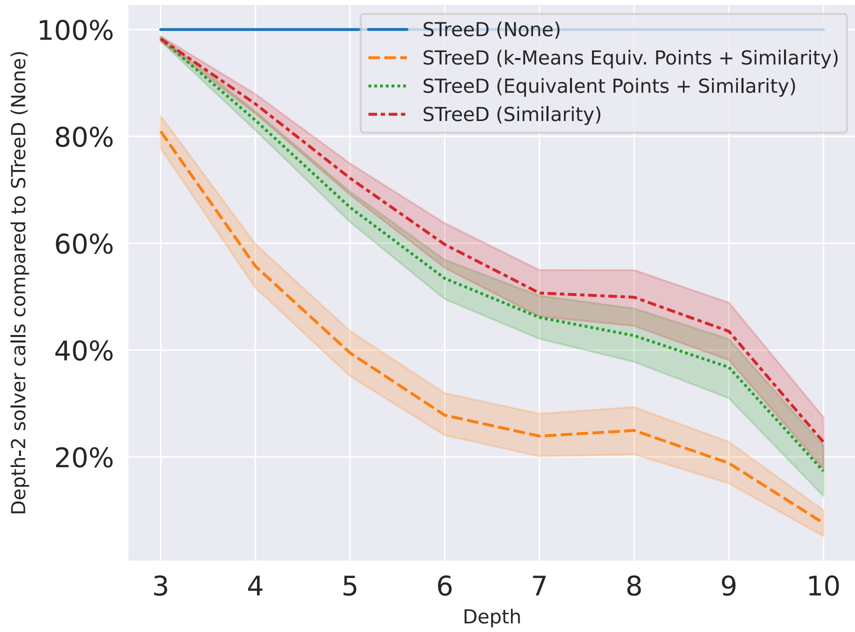

Figure 4: Pruning potential of the implemented lower bounds, lower is better

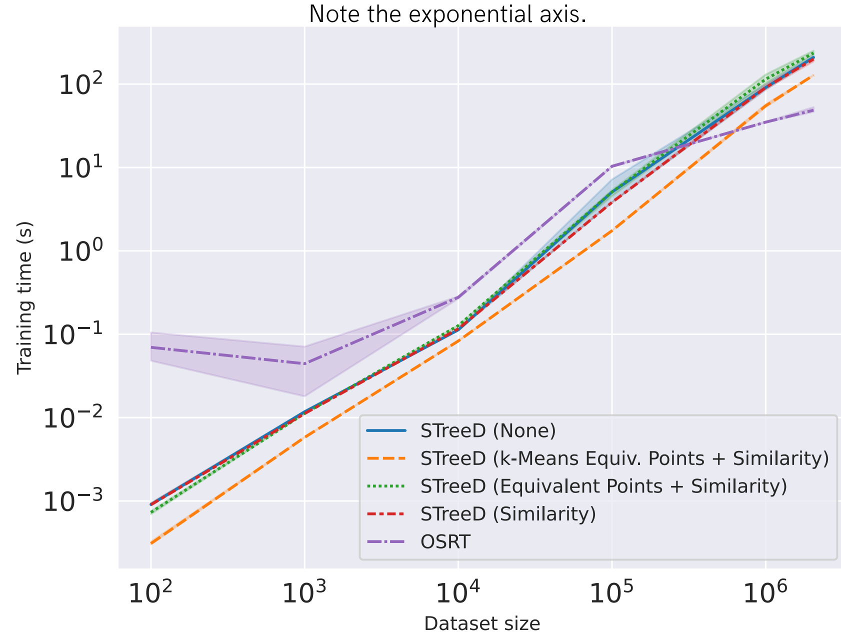

Figure 5: Time to find a tree as a function of dataset size

REFERENCES

[1] Demirović, E., Lukina, A., Hebrard, E., Chan, J., Bailey, J., Leckie, C., Ramamohanarao, K., & Stuckey, P. J. (2022). MurTree: Optimal Decision Trees via Dynamic Programming and Search. Journal of Machine Learning Research, 23(26).
[2] Van der Linden, J.G.M., de Weerdt, M.M., Demirović, E. (2023). Optimal Decision Trees for Separable Objectives: Pushing the Limits of Dynamic Programming. ArXiv [Cs.LG]. Retrieved from http://arxiv.org/abs/2305.19706
[3] Zhang, R., Xin, R., Seltzer, M., & Rudin, C. (2023). Optimal Sparse Regression Trees. ArXiv [Cs.LG]. Retrieved from http://arxiv.org/abs/2211.14980