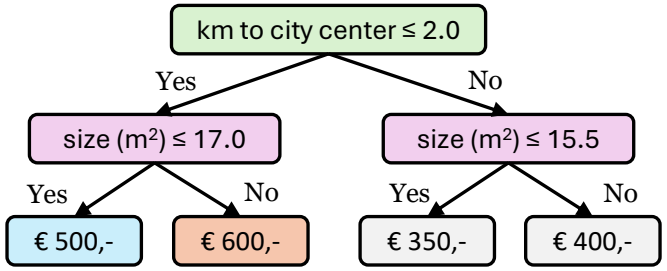
Robust Optimal Regression Trees:
Predicting optimal values even when under attack

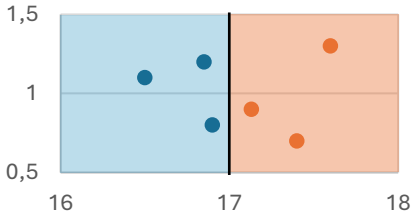
TU Delft

1. Introduction

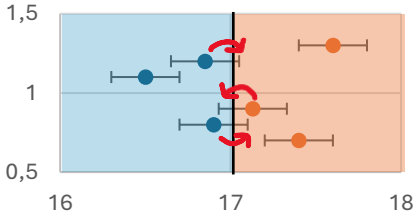
Regression trees are interpretable models that take in features and predict a value, such as rent prices:



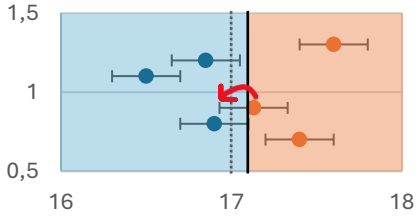
Optimal regression trees minimize the prediction error:



While the adversary maximizes prediction error by changing data:



This is why we need **robust** optimal regression trees:

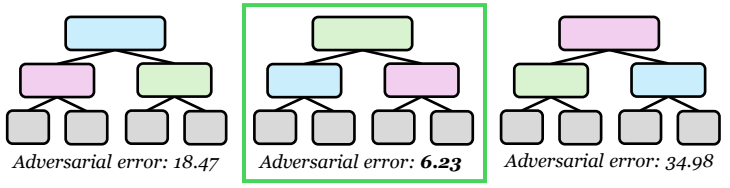


2. Main question:

How can we compute robust optimal regression trees using exhaustive search with pruning?

3. ForTree Approach

Exhaustively search all trees for a set of prediction values and pick the lowest adversarial prediction error:



Prune if the regular prediction error or the partial adversarial error is higher than the minimum found so far:

4. Results

Adversarial R^2 accuracy scores on multiple datasets. Bold is best score, * is best tree at timeout.

Dataset	Chen et al.	TREANT	ForTree
Airfoil	-0.16	**0.01**	0.00*
Auction	**0.45**	0.08	0.36
AutoMPG	-1.59	0.36	**0.48**
Household	timeout	timeout	**-0.01***
OpticalNet.	**0.04**	**0.04**	-0.01*
SeoulBike	-0.08	timeout	**0.02***
Servo	0.10	**0.14**	0.13
Synch.	**0.25**	**0.25**	**0.25***
Yacht	-0.88	**0.29**	0.27

5. Conclusion

ForTree finds trees with higher average adversarial accuracy than previous methods.

Searching more prediction values can improve accuracy, and scalability can be improved.

6. Contact:

Email: j.c.vleeschdraager@student.tudelft.nl

Author: Jesse Vleeschdraager
Supervisor: Ir. J.G.M. van der Linden
Responsible Professor: Dr. Emir Demirović