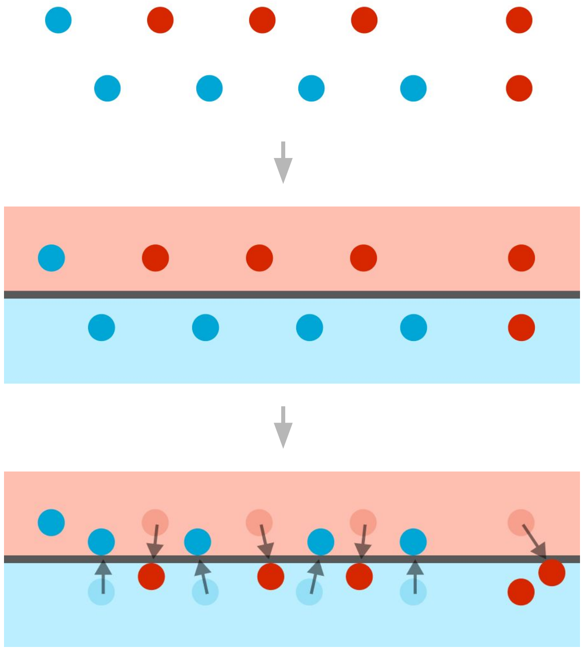
Efficient Training of Robust Decision Trees Against Adversarial Examples

Daniël Vos and Sicco Verwer

Delft University of Technology

Decision trees suffer from adversarial examples

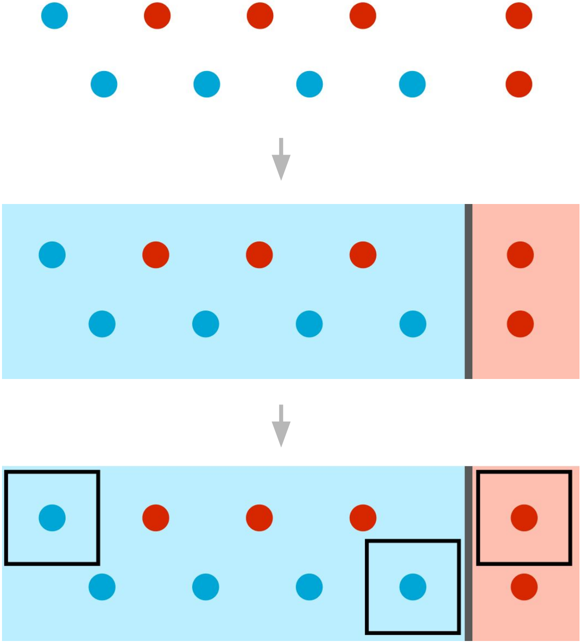


80% accurate

0% accurate against attacks

Similar example by Chen et al. (2019) ‘Robust Decision Trees Against Adversarial Examples’

We aim to fit trees using robust splits

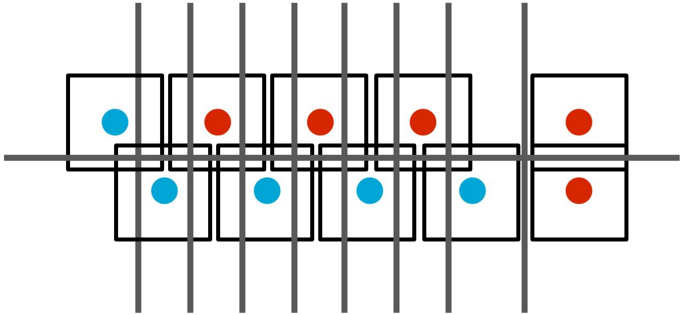


70% accurate

70% accurate against attacks

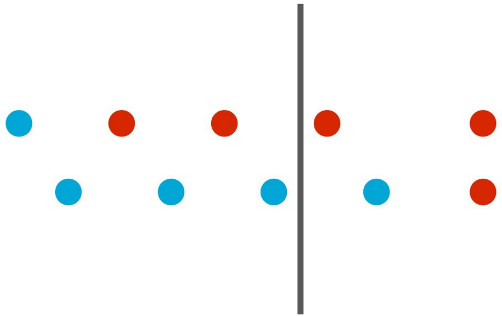
Similar example by Chen et al. (2019) ‘Robust Decision Trees Against Adversarial Examples’

Robust tree learning is slow

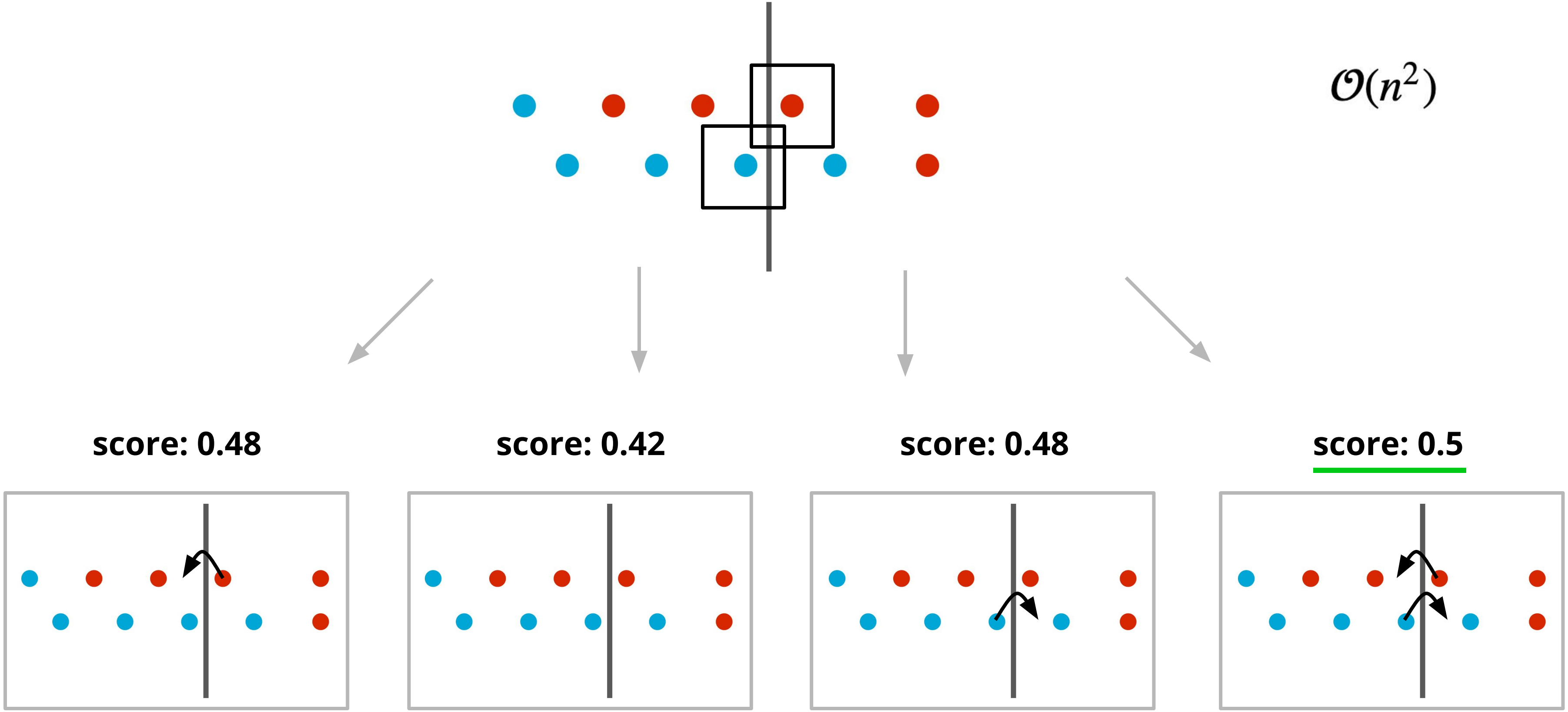


Regular trees use the Gini impurity

score: 0.42



Robust trees use the **maximum** Gini impurity



$\mathcal{O}(n^2)$

score: 0.48

score: 0.42

score: 0.48

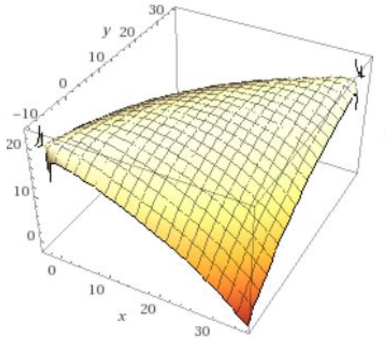
score: 0.5

GROOT, a fast algorithm for growing robust trees

$\mathcal{O}(n \log n)$

- Max. Gini impurity takes time
- Concave function

↪ Analytical solution



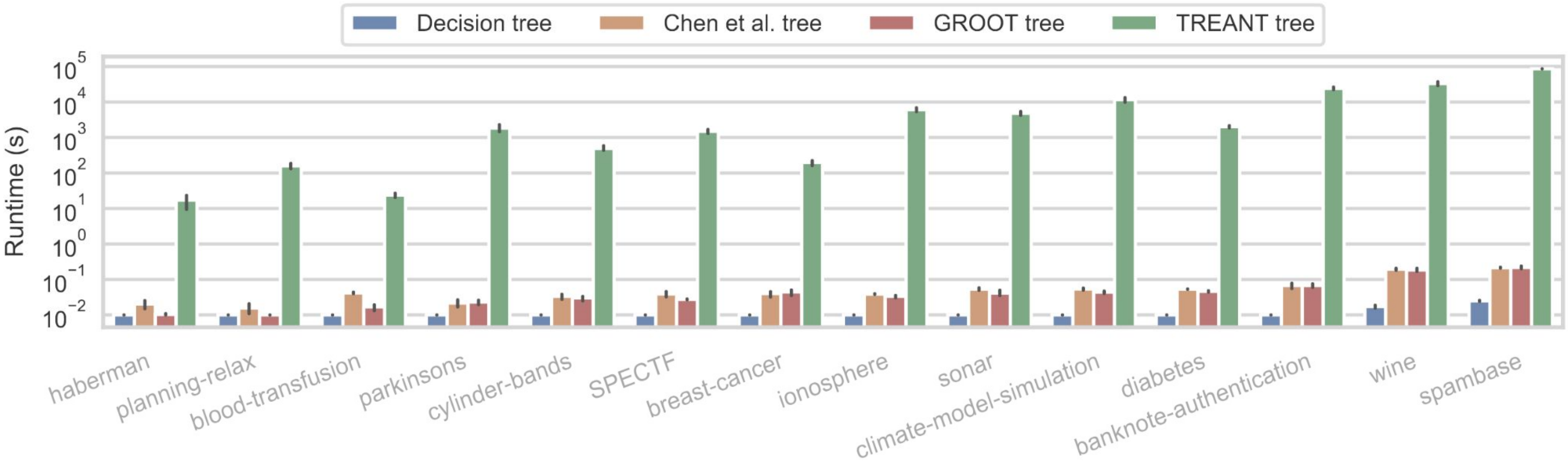
Scoring with the adversarial Gini impurity

$$S(l_0, l_1, r_0, r_1) = \frac{(l_0 + l_1) \cdot \overset{\text{Gini impurity}}{G}(l_0, l_1) + (r_0 + r_1) \cdot G(r_0, r_1)}{l_0 + l_1 + r_0 + r_1}$$

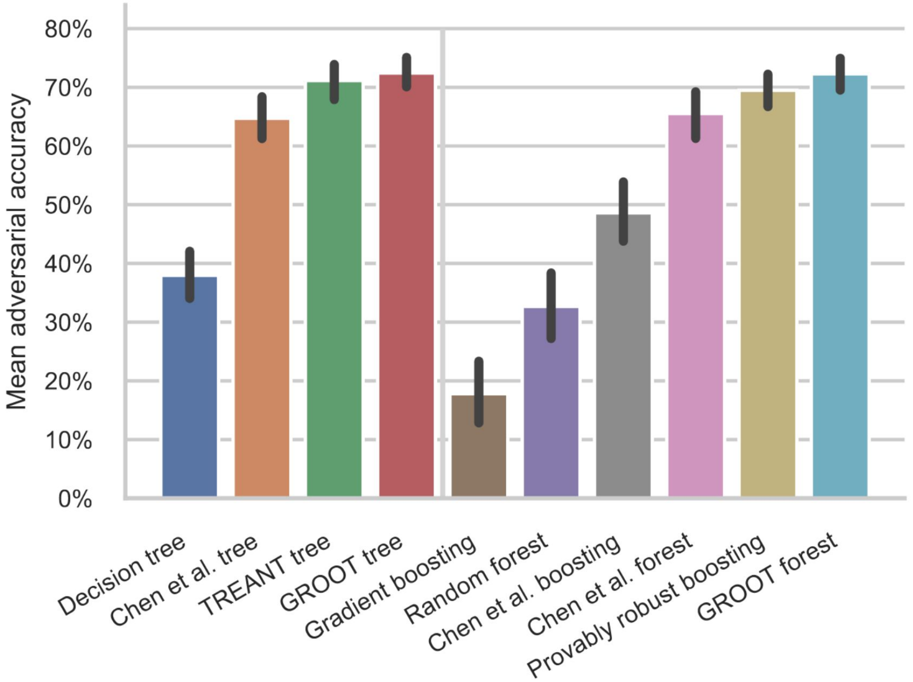
$$S_{\text{robust}}(\overbrace{l_0, l_1}^{\text{Left}}, \overbrace{r_0, r_1}^{\text{Right}}, \overbrace{i_0, i_1}^{\text{Both}}) = \max_{m_1 \in [0, i_1], m_0 \in [0, i_0]} S(l_0 + m_0, l_1 + m_1, r_0 + i_0 - m_0, r_1 + i_1 - m_1)$$

$$\boxed{m_0' = \frac{l_1(r_0 + i_0) - l_0(r_1 + i_1)}{l_1 + r_1 + i_1} + \frac{(l_0 + r_0 + i_0)m_1'}{l_1 + r_1 + i_1}}$$

GROOT is nearly as fast as regular decision trees



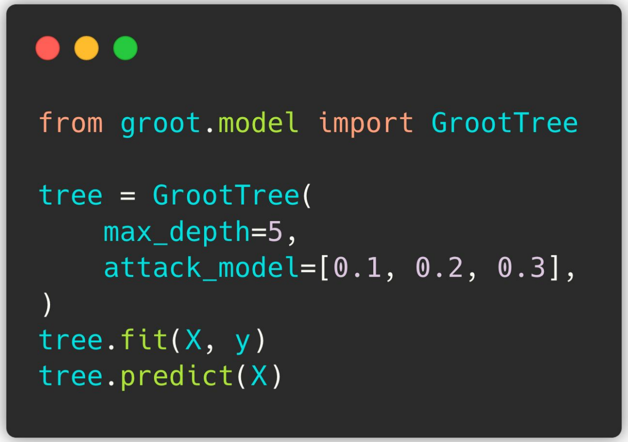
GROOT scores as well as state of the art



Summary

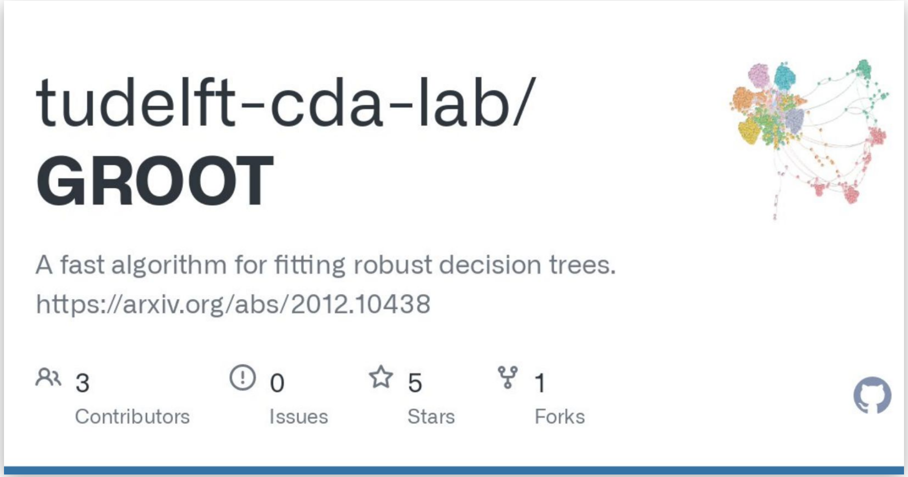
- Robust methods effective but slow
- GROOT splits efficiently
- 2-6 orders of magnitude speedup
- Competitive scores

Find **GROOT** on GitHub



```python
from groot.model import GrootTree

tree = GrootTree(
    max_depth=5,
    attack_model=[0.1, 0.2, 0.3],
)
tree.fit(X, y)
tree.predict(X)
```



tudelft-cda-lab/
GROOT

A fast algorithm for fitting robust decision trees.
https://arxiv.org/abs/2012.10438

3 Contributors
0 Issues
5 Stars
1 Forks