

Optimal Sparse Decision Trees



Xiyang Hu

Carnegie Mellon
University



Cynthia Rudin

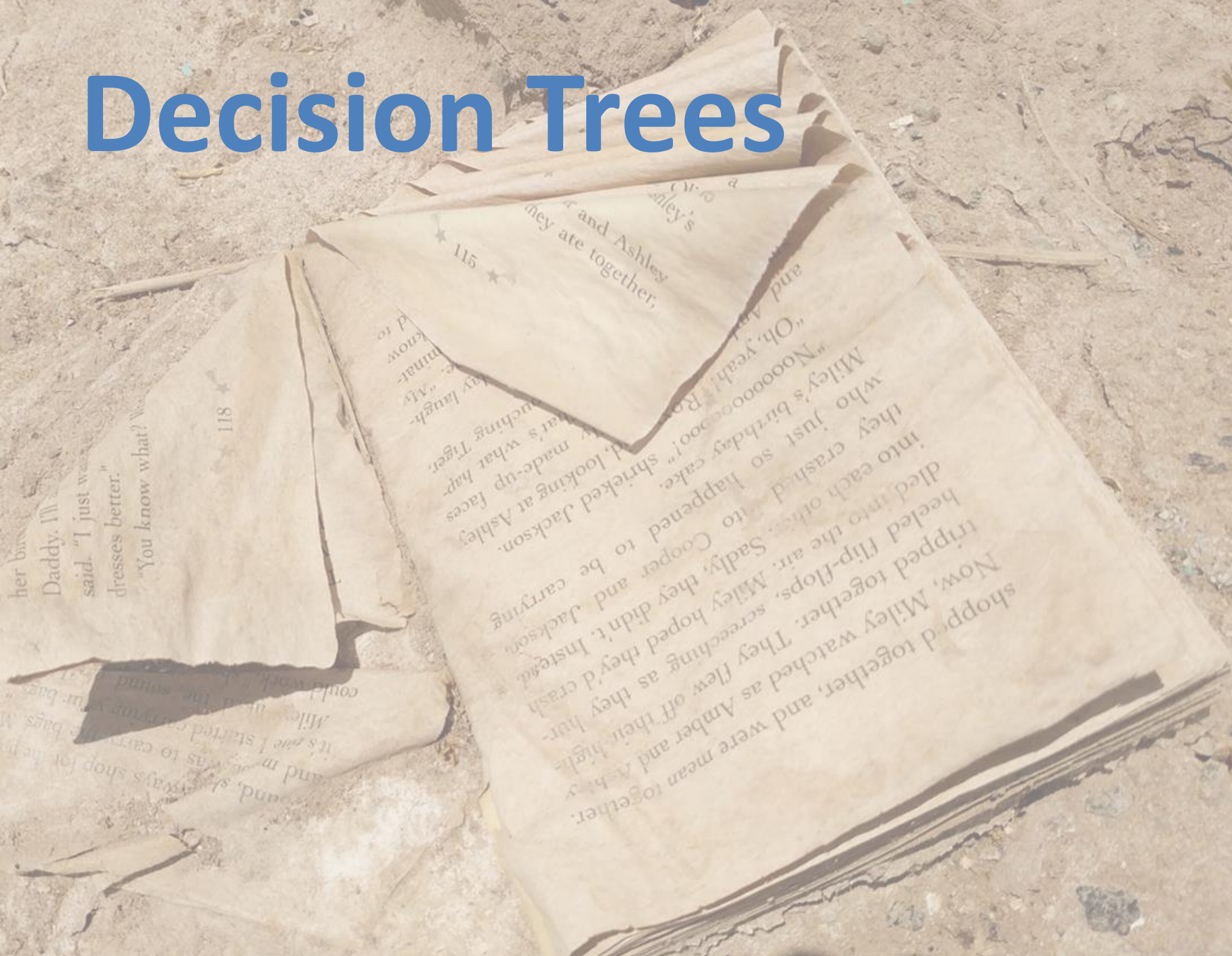
Duke University



Margo Seltzer

University of British
Columbia

Decision Trees



her b...
Daddy, I'll
said. "I just want
dresses better."
"You know what?"

118

115

and Ashley
they ate together.

"Oh, yeah!"
"Nooooooo!"
"My birthday cake!"
shrieked Jackson.
who just crashed so happened to be carrying
they crashed into the air. Sadly, they didn't. Instead
healed flip-flops, Milley hoped as they hurtled
tripped together. They flew off their high-heeled flip-flops, and were mean together.
tripped together, and were mean together.
shopped together, and were mean together.
Now, Milley watched as Amber and Ashley
died into the air. Sadly, they didn't. Instead
into each other. It happened to be carrying
shrieked Jackson.
made-up faces
what Tiger
My laugh-
know
d to

Decision Trees



C4.5

her birthday
Daddy. I'll
said. "I just
dresses better."
"You know what?"

115

and Ashley
they ate together.

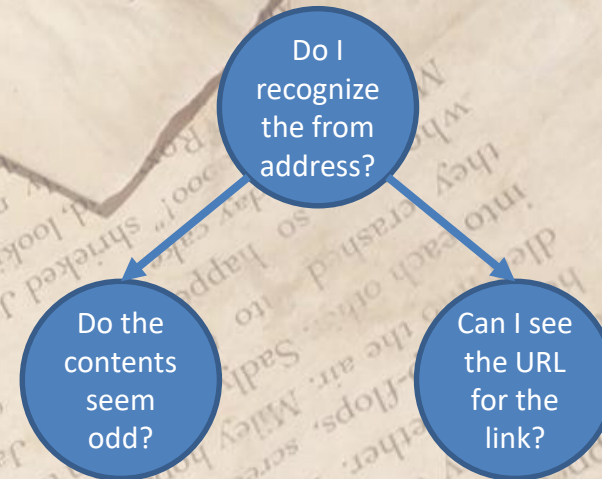
know
what
"My
they laugh-

what hap-
shrieked Jackson.

Now, Milley watched as Amber and Ashley
shopped together, and were mean together.
tripped together. They flew off their high-
heeled flip-flops. Milley hoped as they'd crash
into the air. Sadly, they didn't. Instead,
died into the air. Sad, they didn't. Jackson
they just so happened to be carrying
"Oh, yeah! Ro!" shrieked Jackson.
"Nooooooo!"
"Oh, yeah! Ro!"
"Oh, yeah! Ro!"
"Oh, yeah! Ro!"

Decision Trees

Should I click on the link in this email?

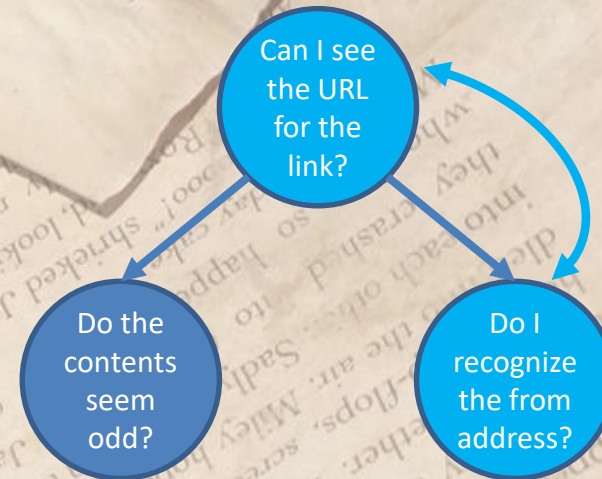


C4.5



Decision Trees

Should I click on the link in this email?



C4.5



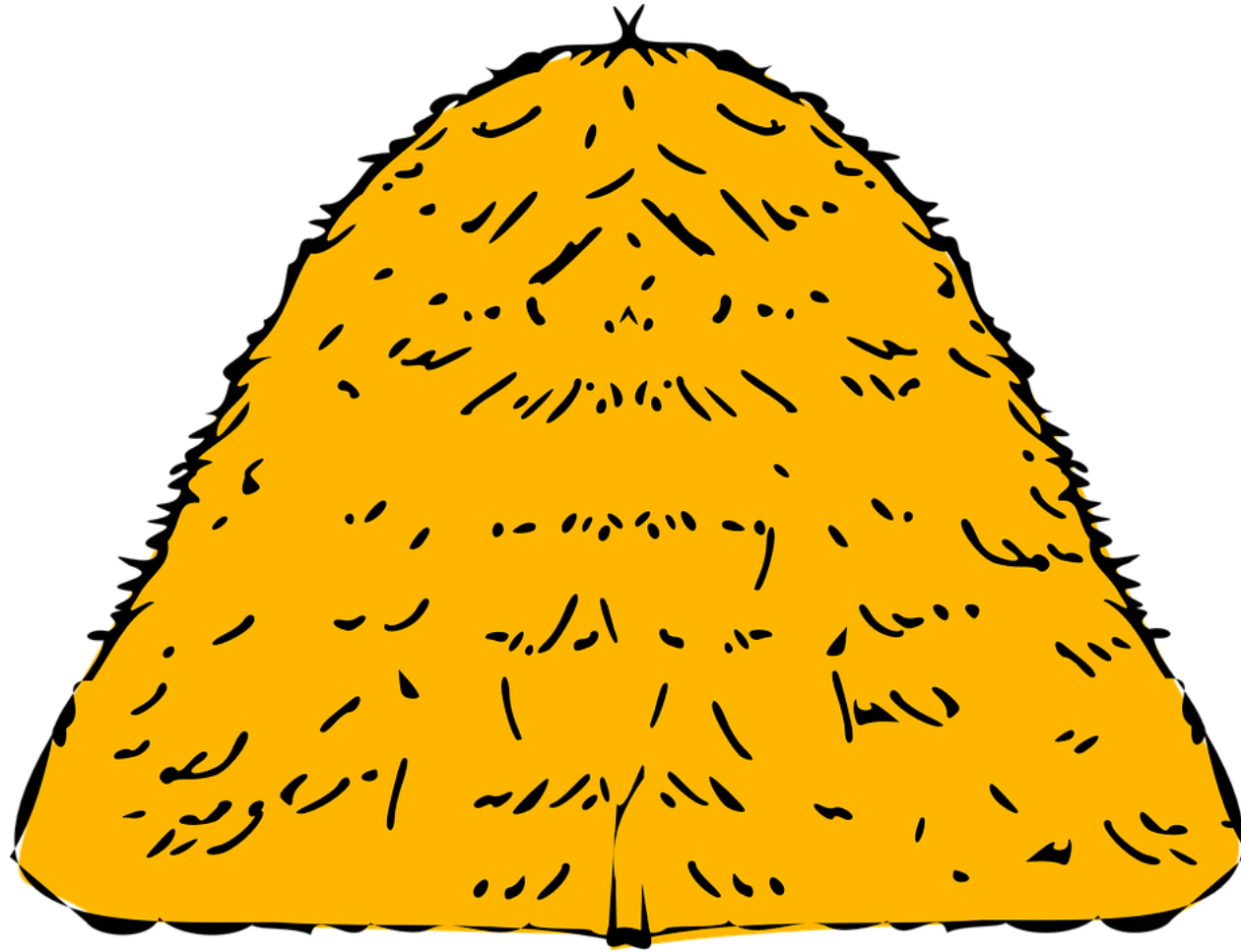
Why not just find the Best Tree?



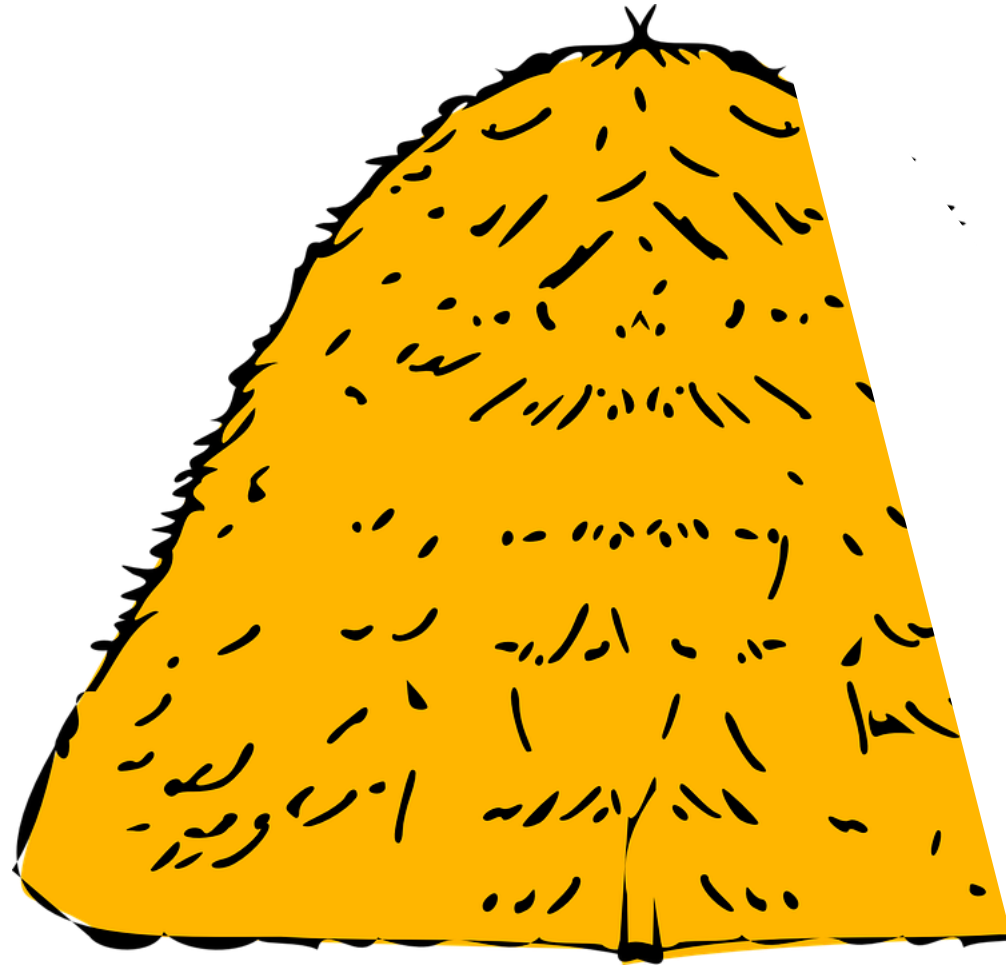
Why not just find the Best Tree?

$O(10^{28})$

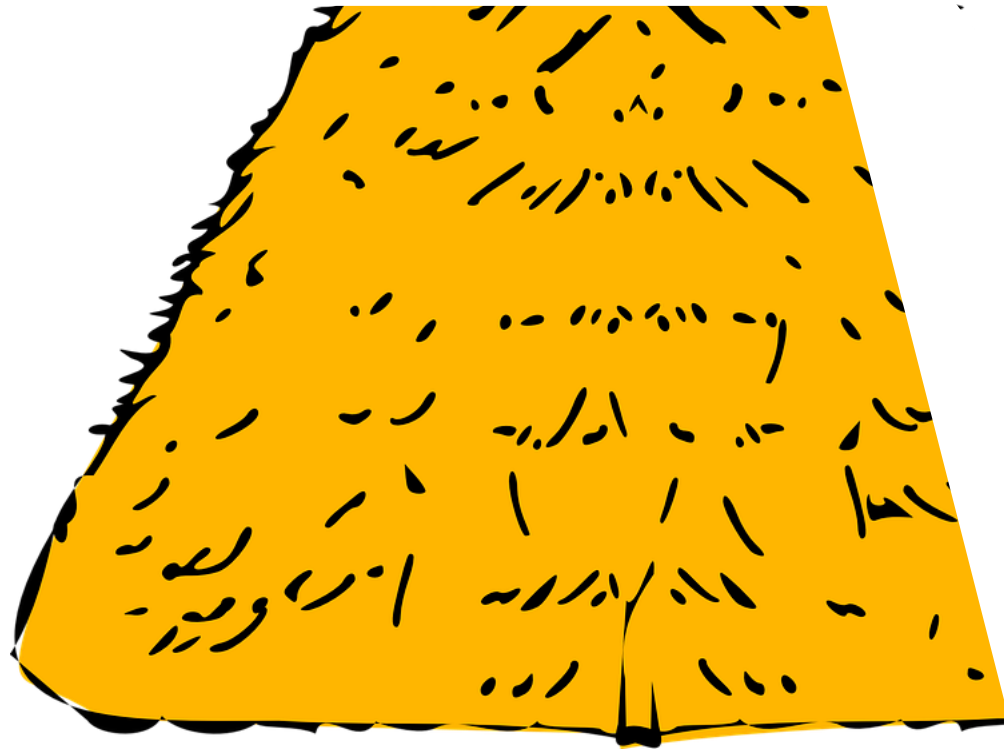
Could we Effectively Search that Space?



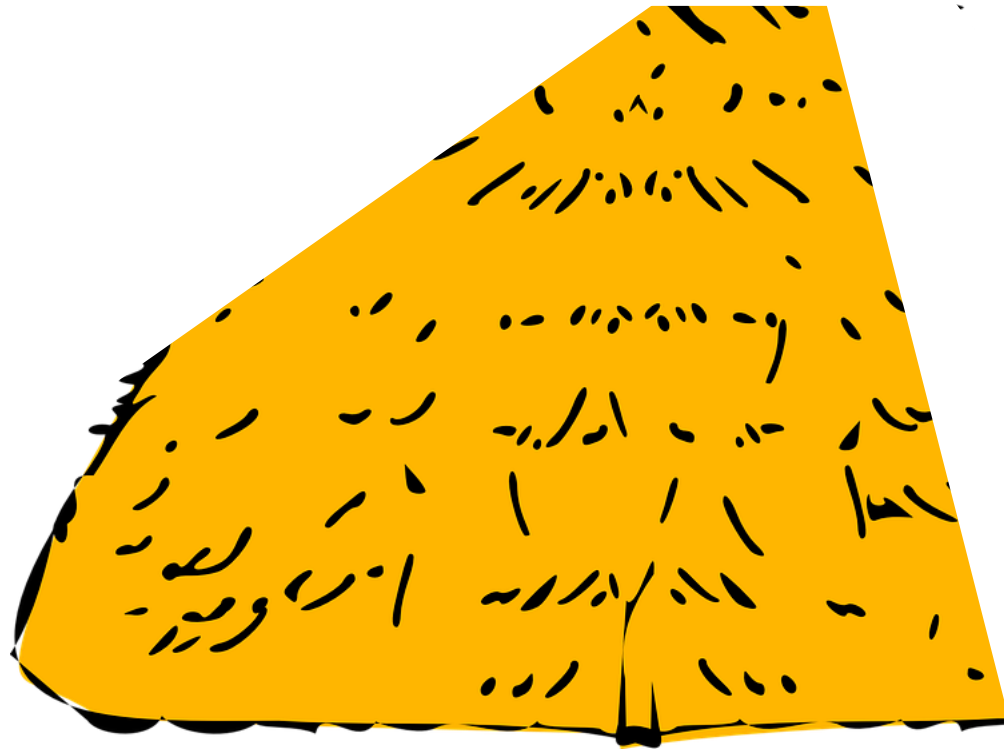
Could we Effectively Search that Space?



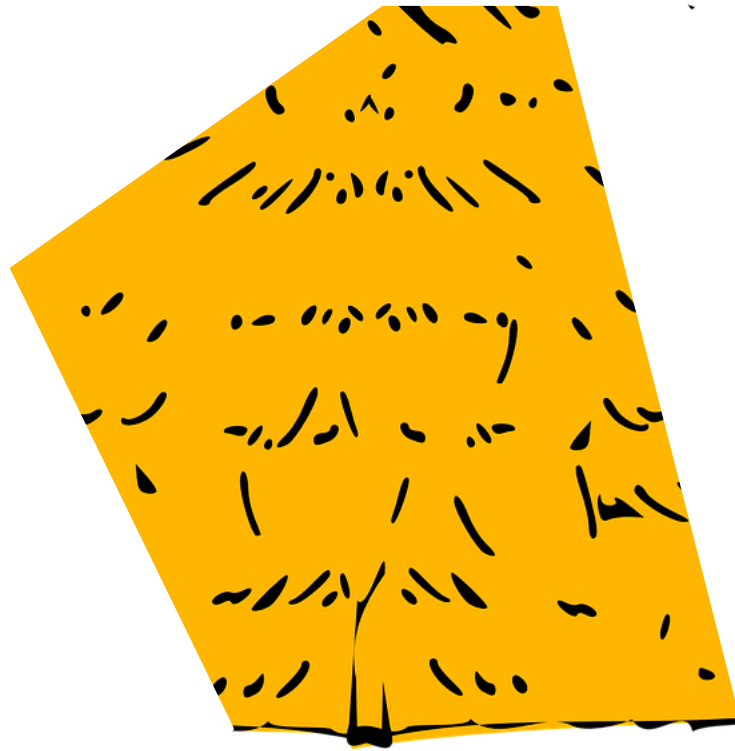
Could we Effectively Search that Space?



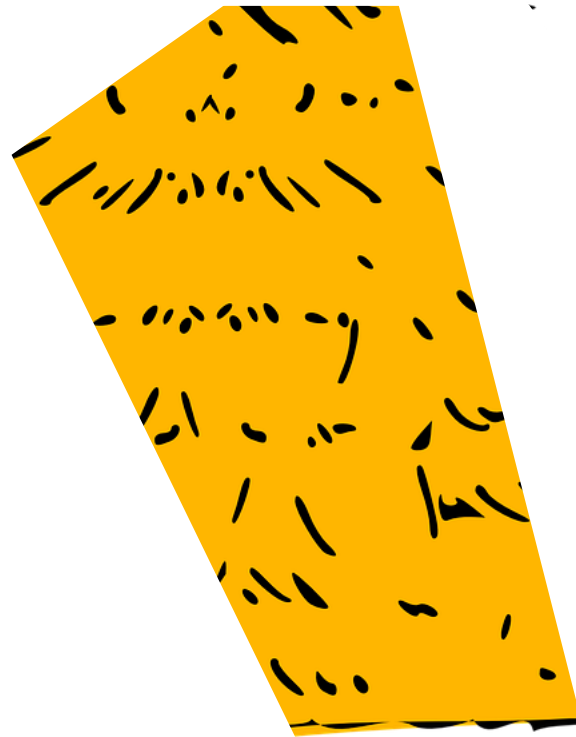
Could we Effectively Search that Space?



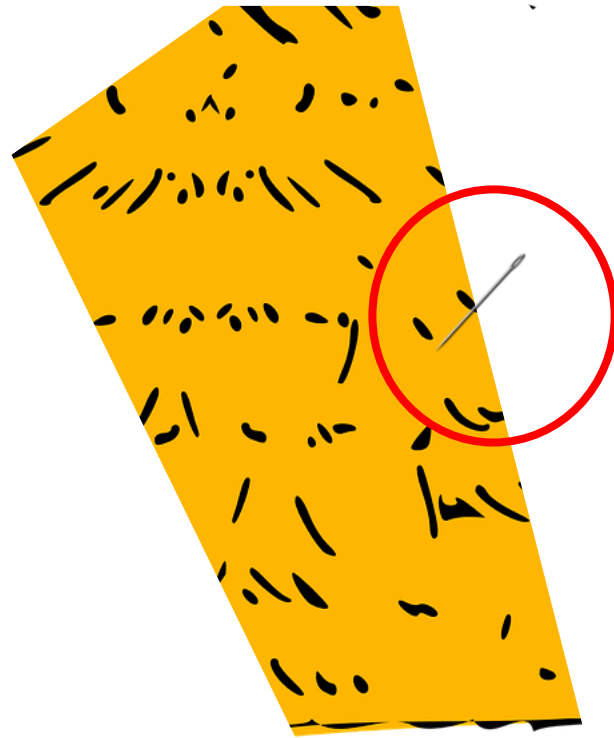
Could we Effectively Search that Space?



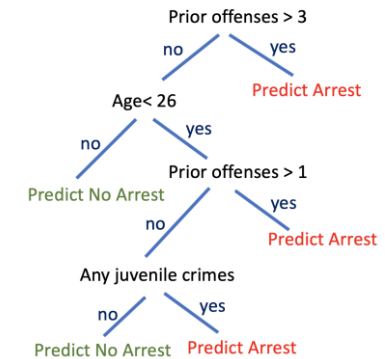
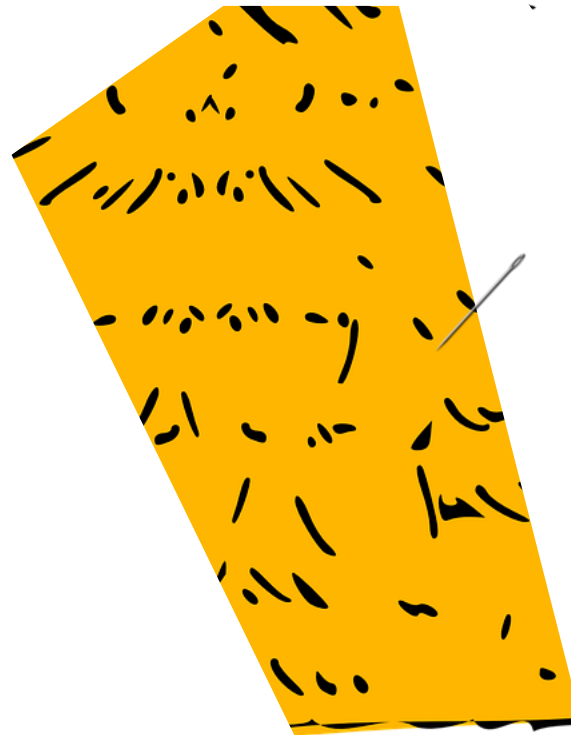
Could we Effectively Search that Space?



Could we Effectively Search that Space?



Could we Effectively Search that Space?



The Optimization Problem

$$\hat{L}(\text{tree}, \{(x_i, y_i)\}_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[\text{tree}(x_i) \neq y_i]} + C(\# \text{ leaves in tree})$$

The Optimization Problem

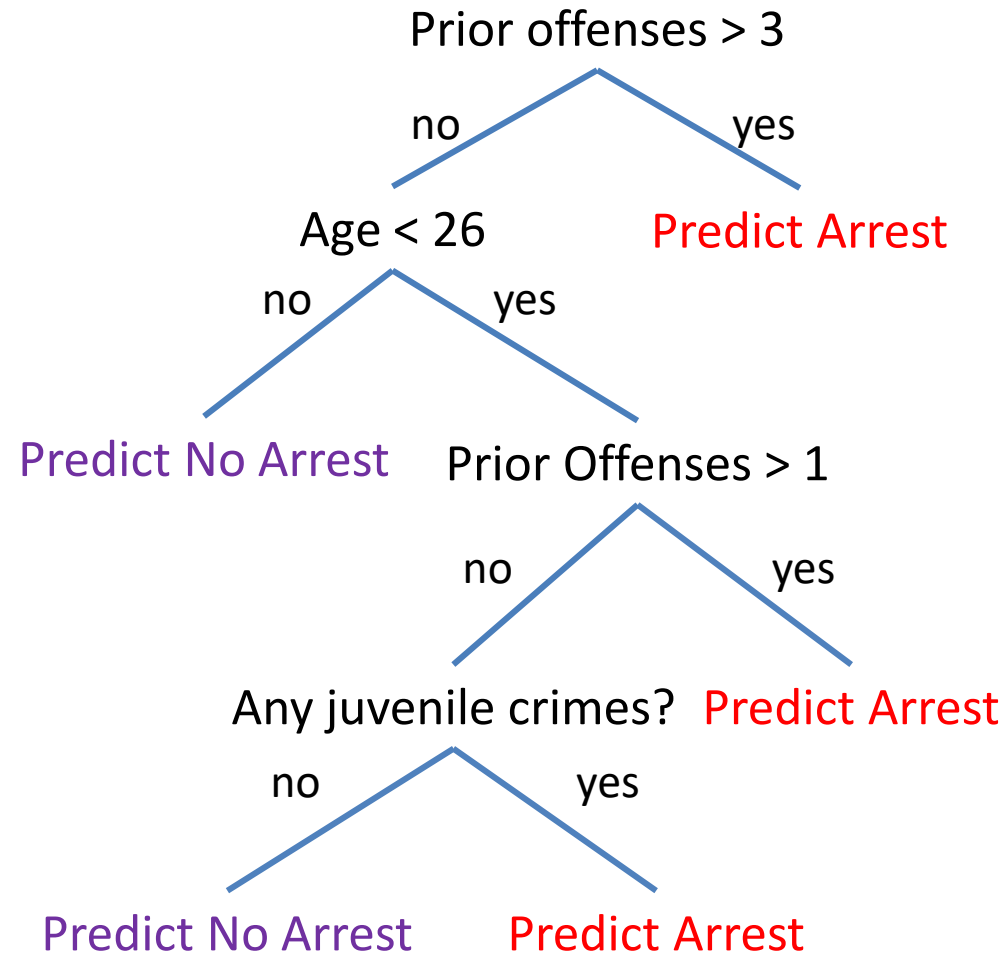
$$\hat{L}(\text{tree}, \{(x_i, y_i)\}_i) = \underbrace{\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[\text{tree}(x_i) \neq y_i]}}_{\text{Misclassification error}} + C(\# \text{ leaves in tree})$$

The Optimization Problem

$$\hat{L}(\text{tree}, \{(x_i, y_i)\}_i) = \underbrace{\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{[\text{tree}(x_i) \neq y_i]}}_{\text{Misclassification error}} + \underbrace{C(\# \text{ leaves in tree})}_{\text{Sparsity}}$$

Optimal Sparse Decision Tree

(Broward County Recidivism Data)



Optimal Sparse Decision Trees

Branch and Bound



Good Scheduling Order



Strong Bounds



Incremental Computation

Optimal Sparse Decision Trees

Branch and Bound



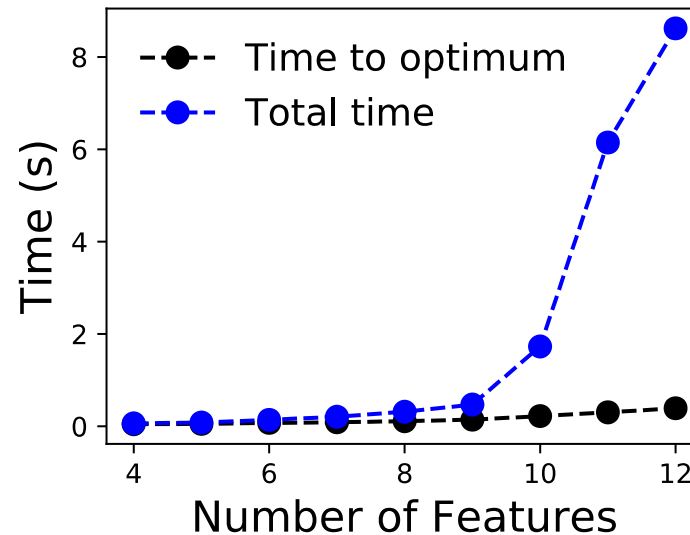
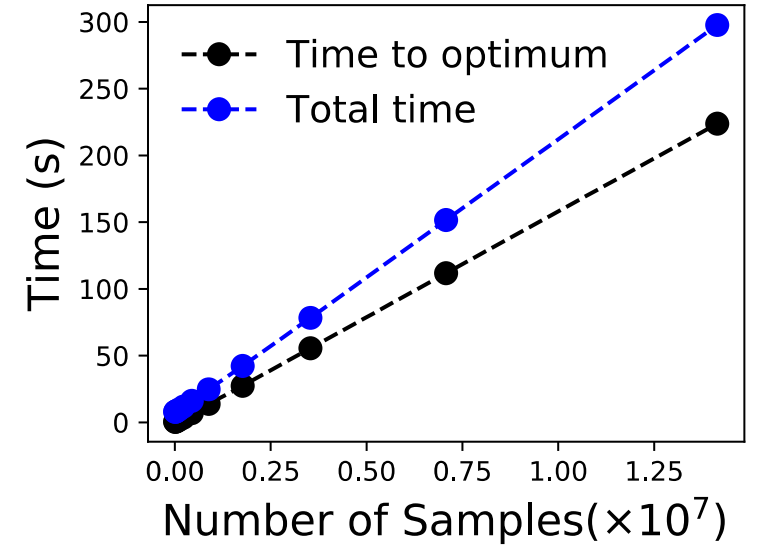
Good Scheduling Order



Strong Bounds



Incremental Computation



FAST

Optimal Sparse Decision Trees

Branch and Bound



Good Scheduling Order

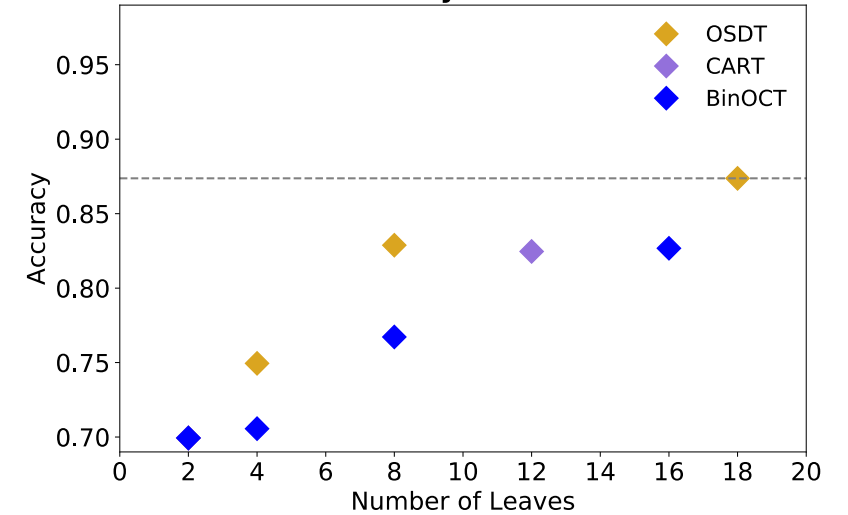


Strong Bounds

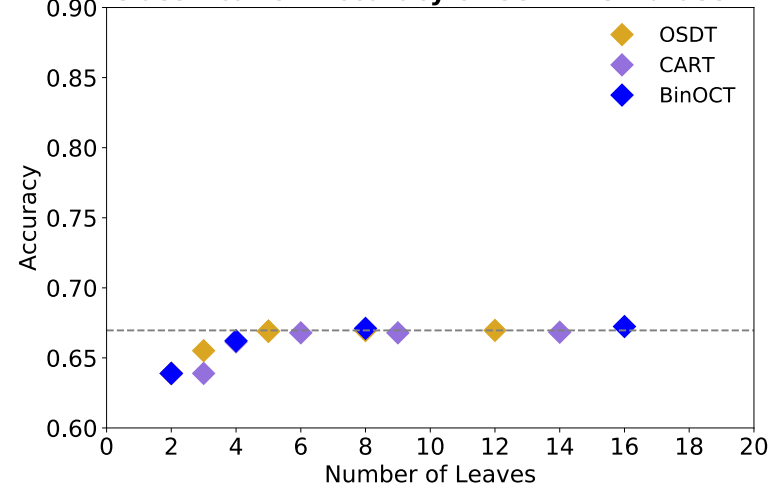


Incremental Computation

Classification Accuracy of Tic-Tac-Toe Dataset



Classification Accuracy of COMPAS Dataset

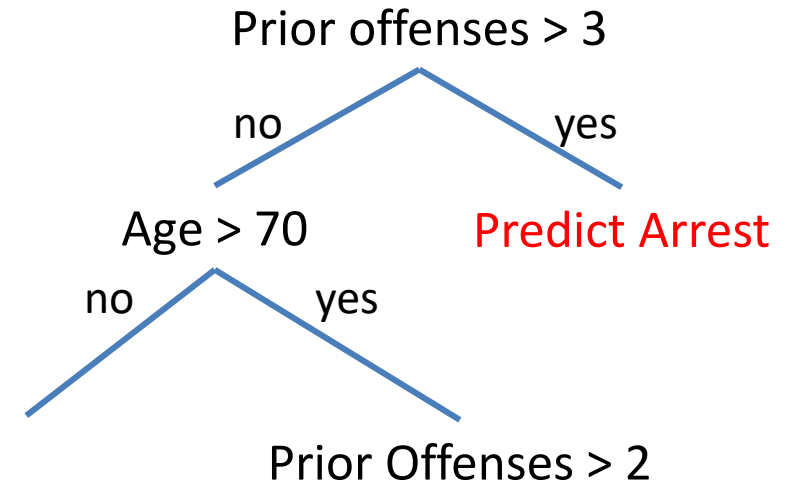


Accurate

Bounding the Search Space

Lower Bound on Node Support

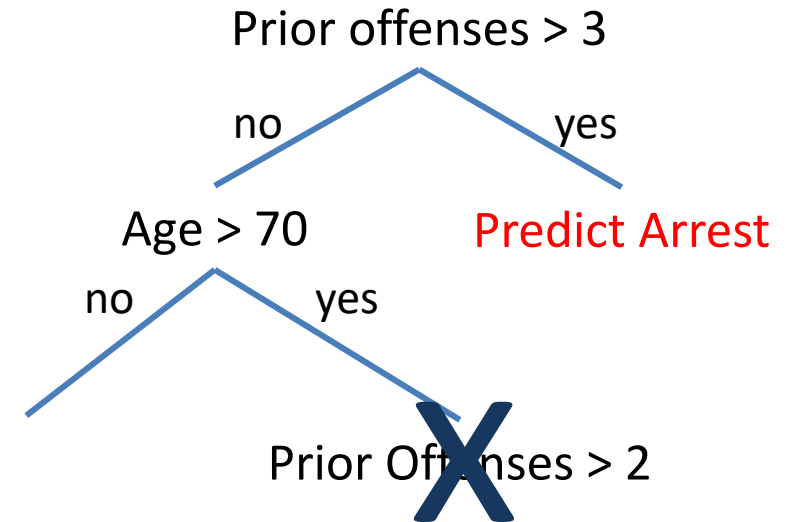
Theorem: For an optimal tree, the support of each node must be above $2C$.



Bounding the Search Space

Lower Bound on Node Support

Theorem: For an optimal tree, the support of each node must be above $2C$.

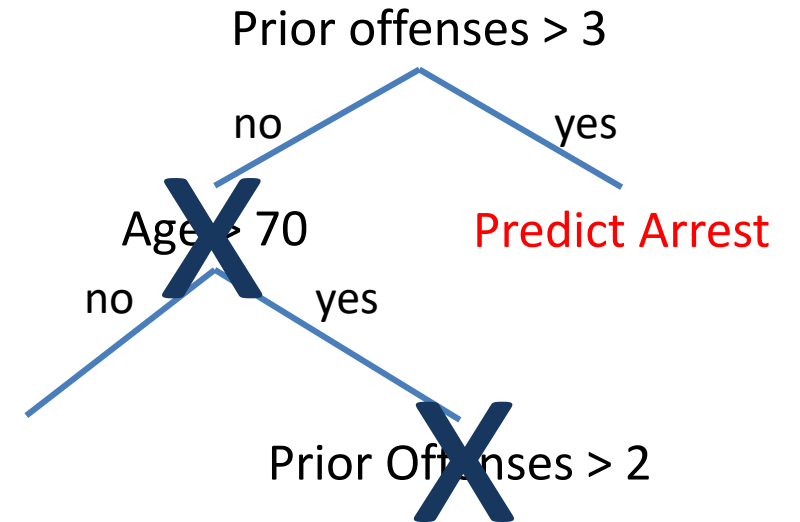


Node support
insufficient to produce
optimal solution

Bounding the Search Space

Lower Bound on Node Support

Theorem: For an optimal tree, the support of each node must be above $2C$.

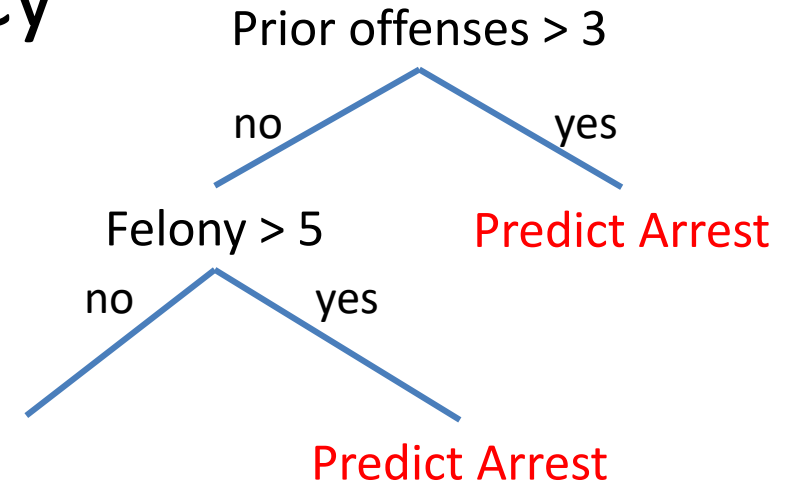


Node support
insufficient to produce
optimal solution

Bounding the Search Space

Lower Bound on Classification Accuracy

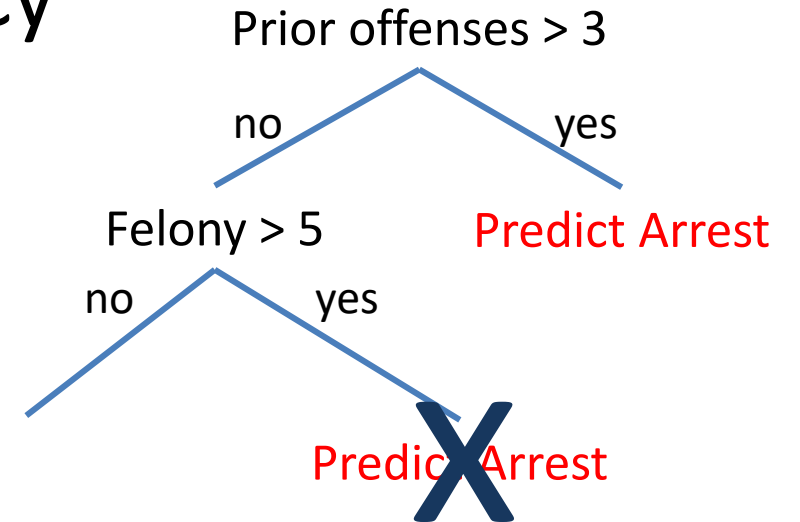
Theorem: Each leaf of an optimal tree correctly classifies at least fraction C of the data



Bounding the Search Space

Lower Bound on Classification Accuracy

Theorem: Each leaf of an optimal tree correctly classifies at least fraction C of the data

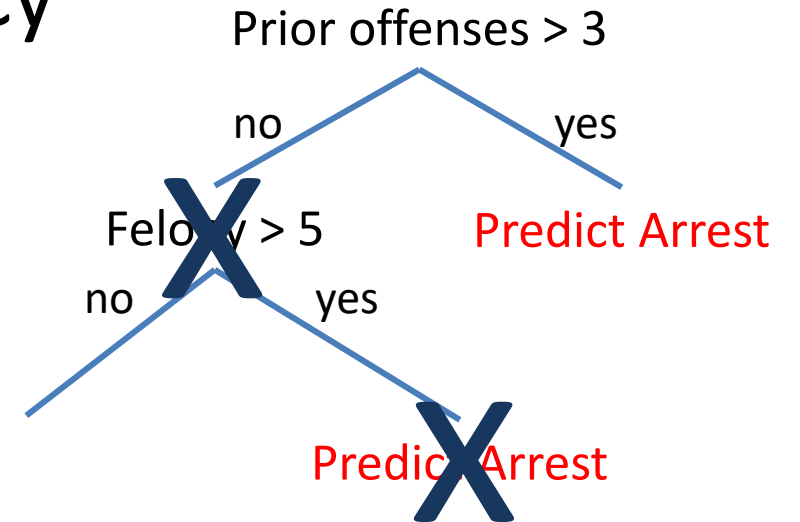


Doesn't classify at least Cn points correctly.

Bounding the Search Space

Lower Bound on Classification Accuracy

Theorem: Each leaf of an optimal tree correctly classifies at least fraction C of the data

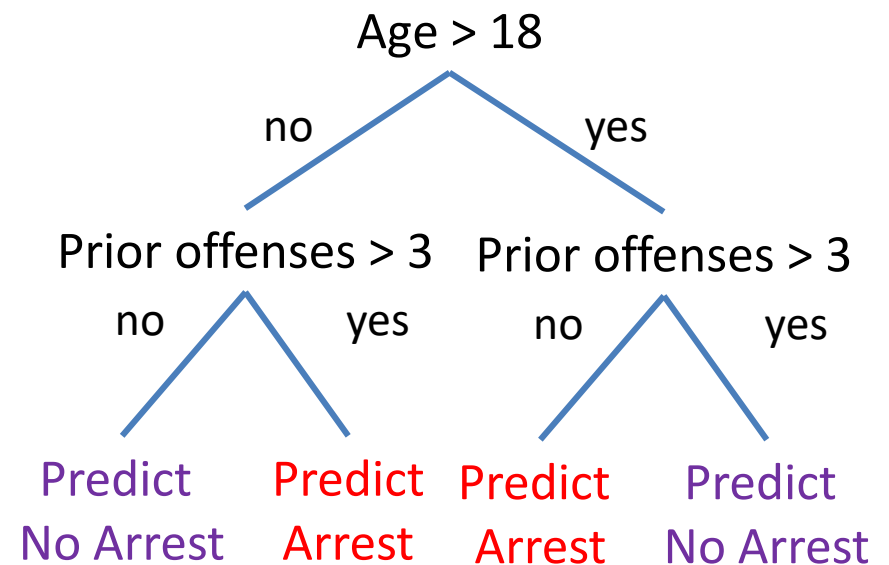
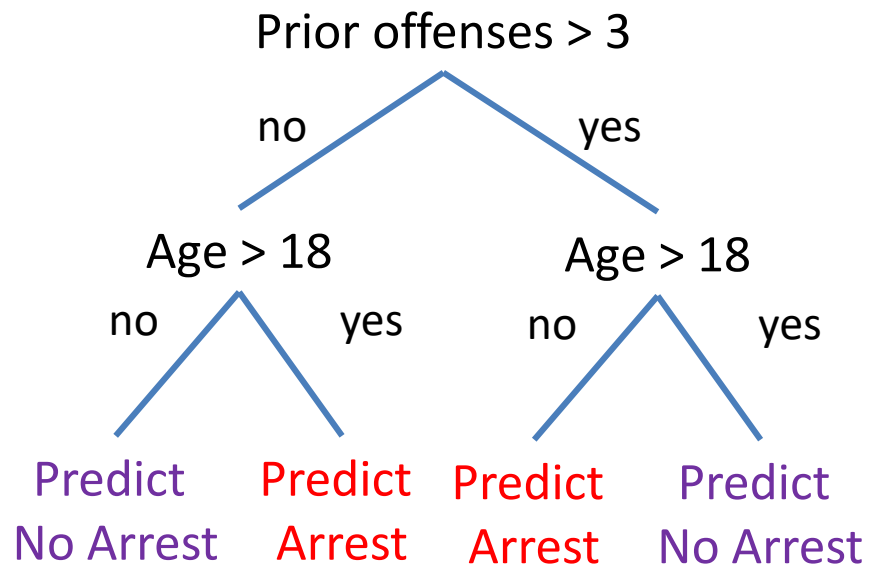


Doesn't classify at least Cn points correctly.

Bounding the Search Space

Permutation Bound

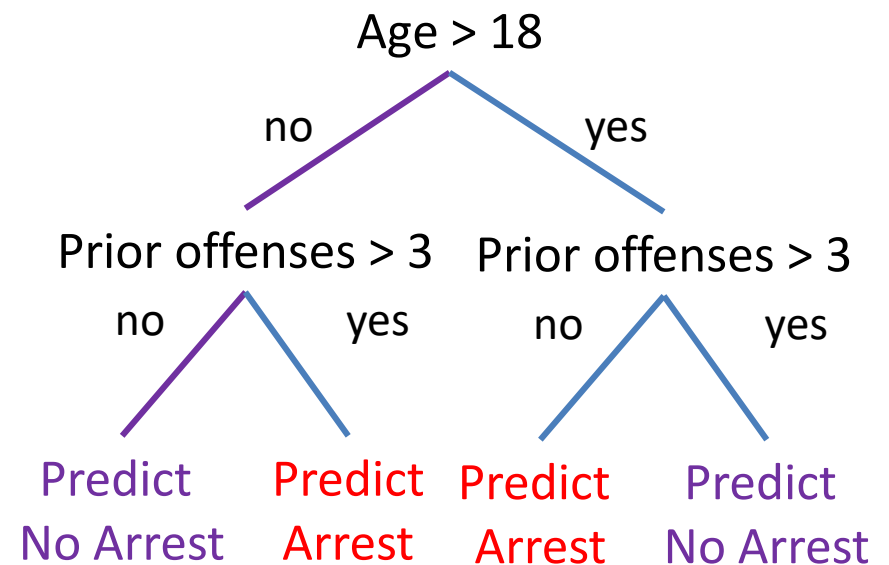
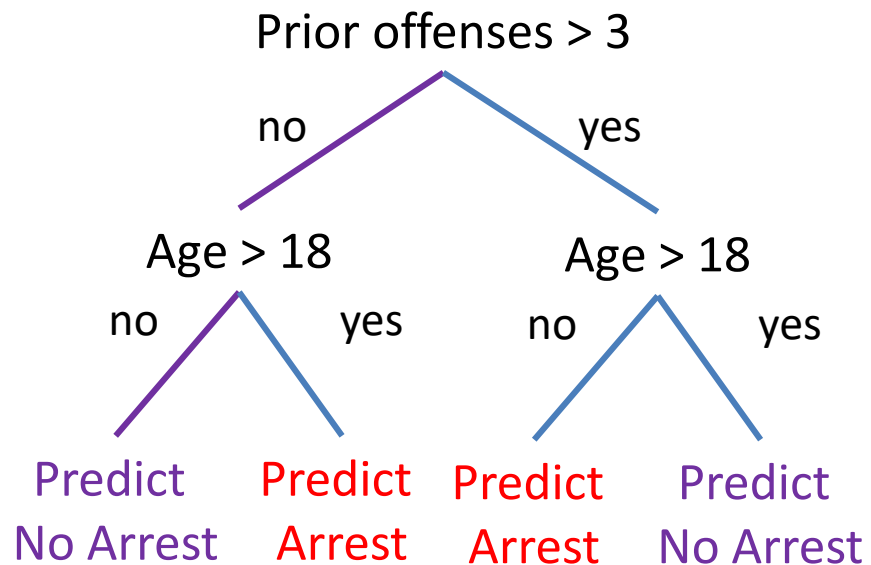
Theorem: If two trees have the same leaves, up to a permutation, all their child trees will be the same, so one of them can be pruned.



Bounding the Search Space

Permutation Bound

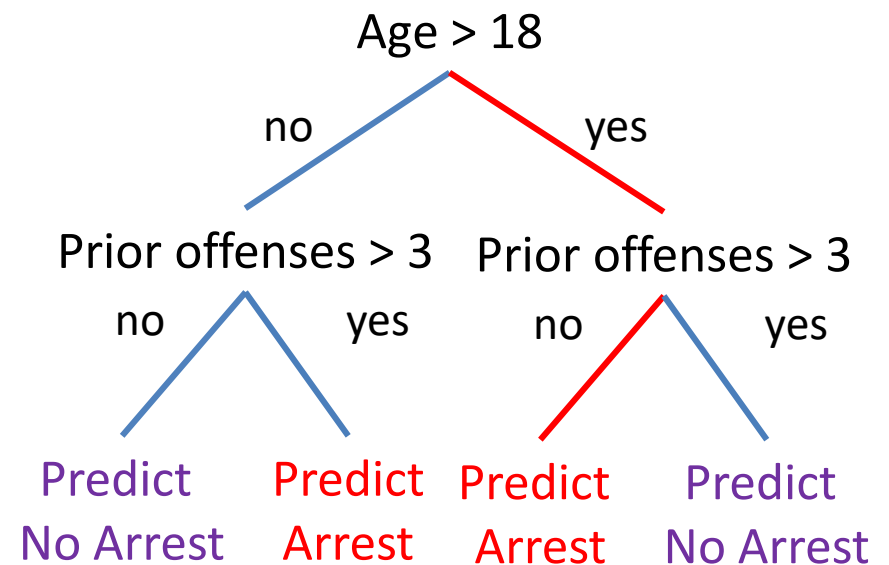
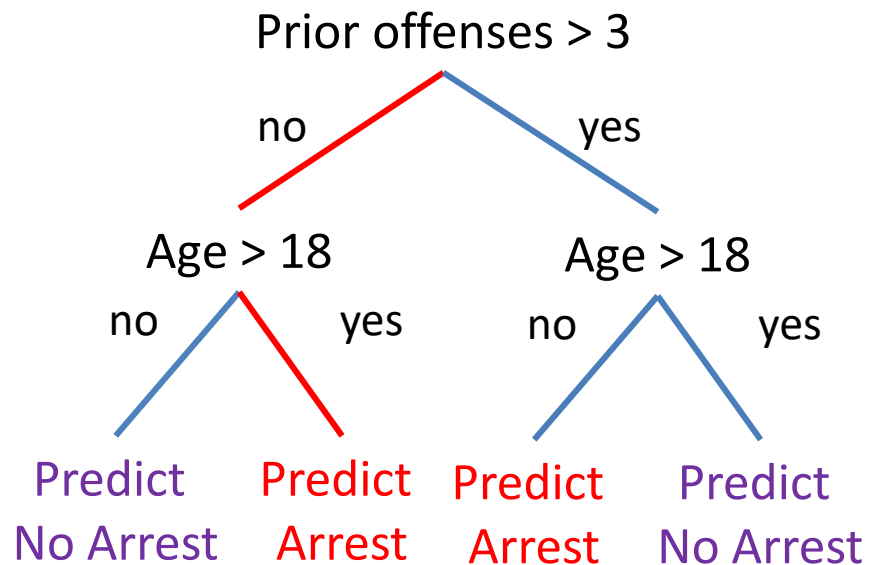
Theorem: If two trees have the same leaves, up to a permutation, all their child trees will be the same, so one of them can be pruned.



Bounding the Search Space

Permutation Bound

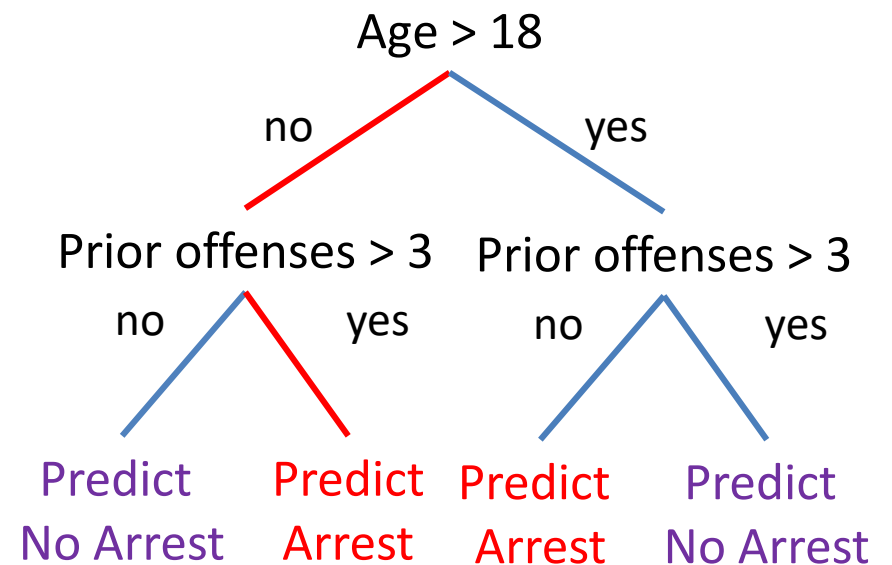
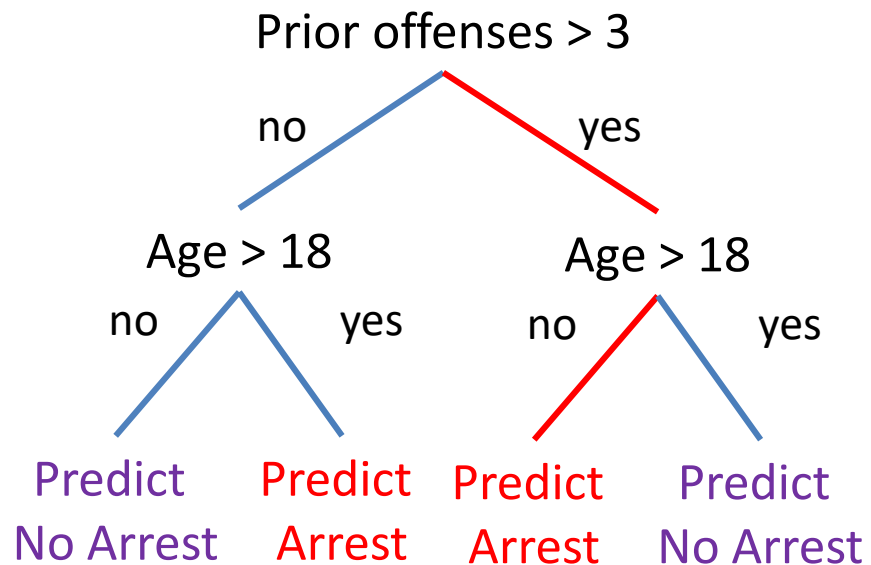
Theorem: If two trees have the same leaves, up to a permutation, all their child trees will be the same, so one of them can be pruned.



Bounding the Search Space

Permutation Bound

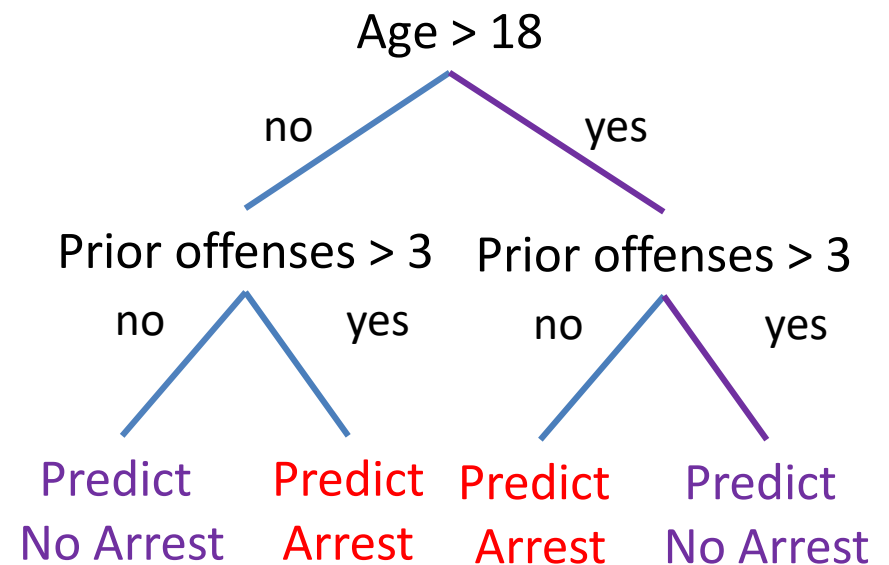
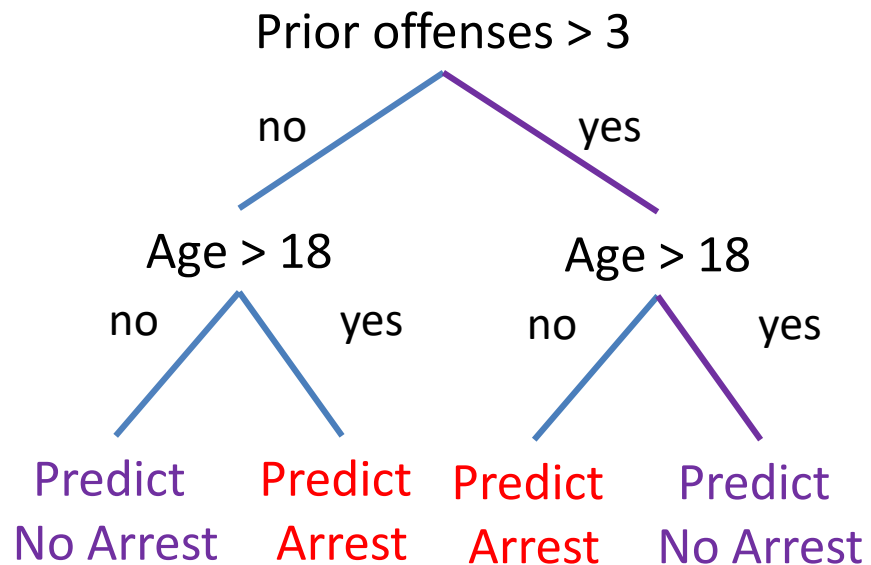
Theorem: If two trees have the same leaves, up to a permutation, all their child trees will be the same, so one of them can be pruned.



Bounding the Search Space

Permutation Bound

Theorem: If two trees have the same leaves, up to a permutation, all their child trees will be the same, so one of them can be pruned.



Bounding the Search Space

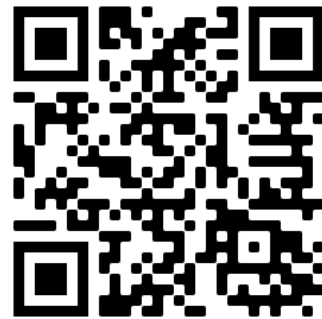
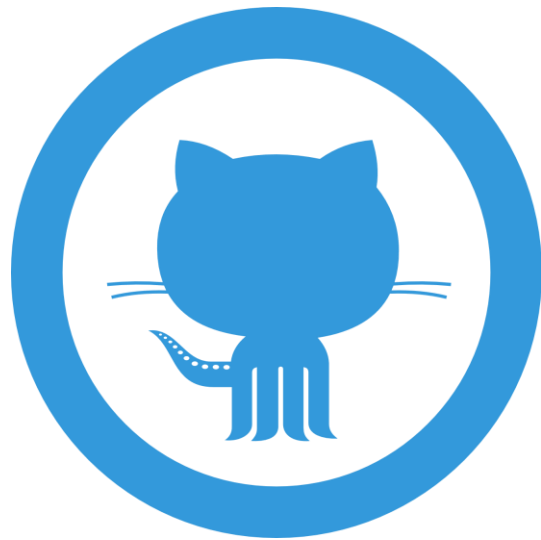
- Other bounds enable even more pruning
 - Equivalent points bound: Samples with the same features, but different predictions will produce misclassifications regardless of model.
 - Bound on the number of leaves: Regularization value bounds the number of leaves.

Optimal Sparse Decision Trees

Open Source

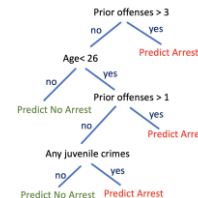


Accurate



<https://github.com/xiyanghu/OSDT>

FAST



Interpretable