Optimal Sparse Decision Trees

Xiyang Hu

Carnegie Mellon University

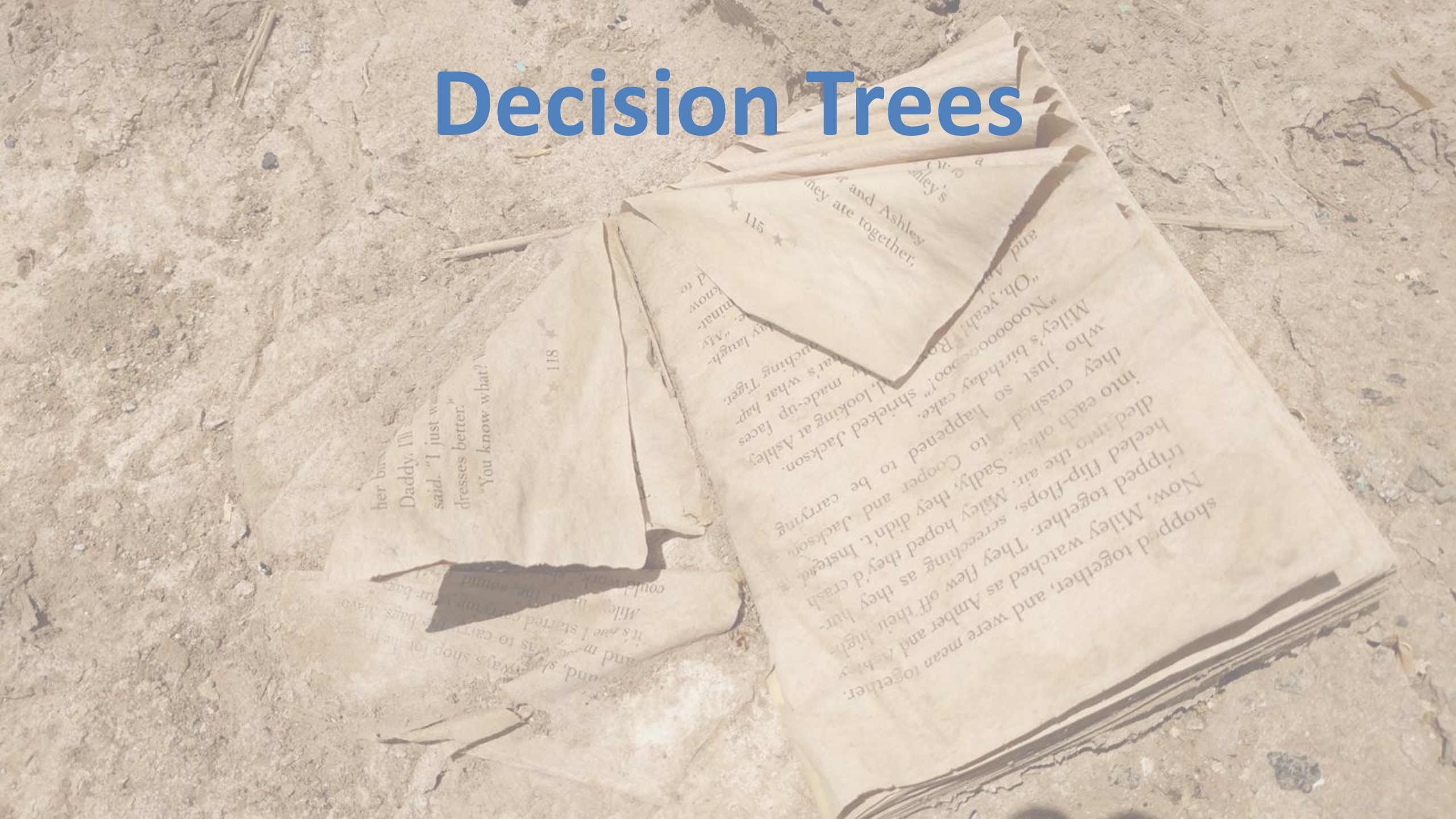
Cynthia Rudin

Duke University

Margo Seltzer

University of British Columbia

Decision Trees



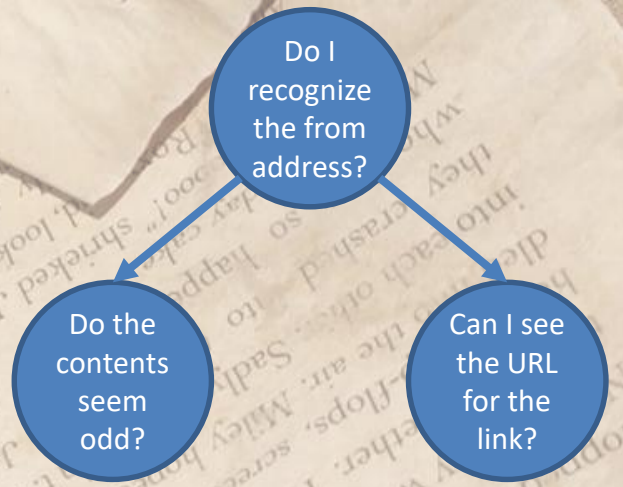
Decision Trees

C4.5

Decision Trees

C4.5

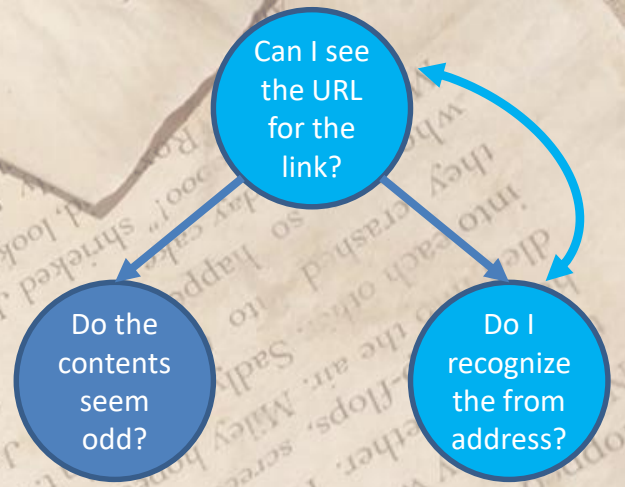
Should I click on the link in this email?



Decision Trees

C4.5

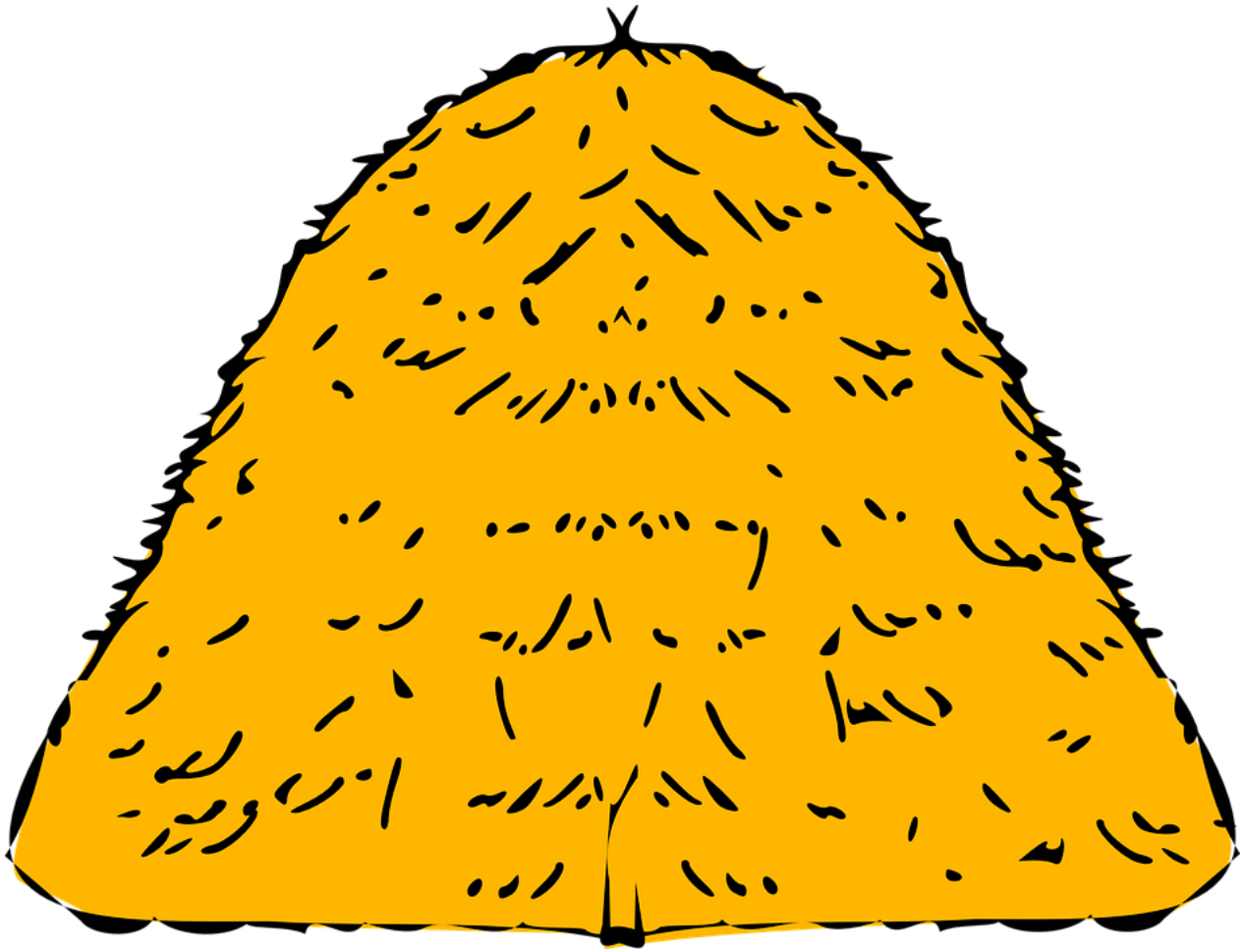
Should I click on the link in this email?



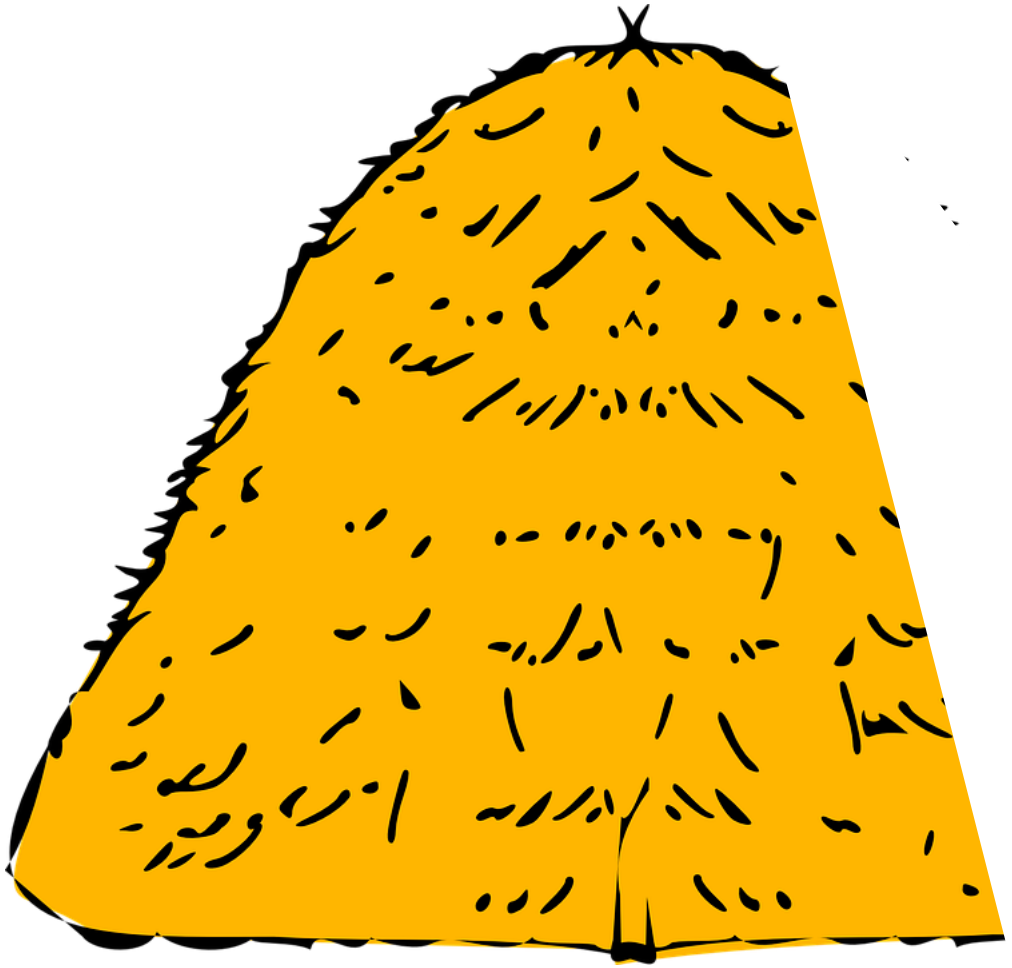
Why not just find the Best Tree?


Why not just find the Best Tree?
O(10^28)

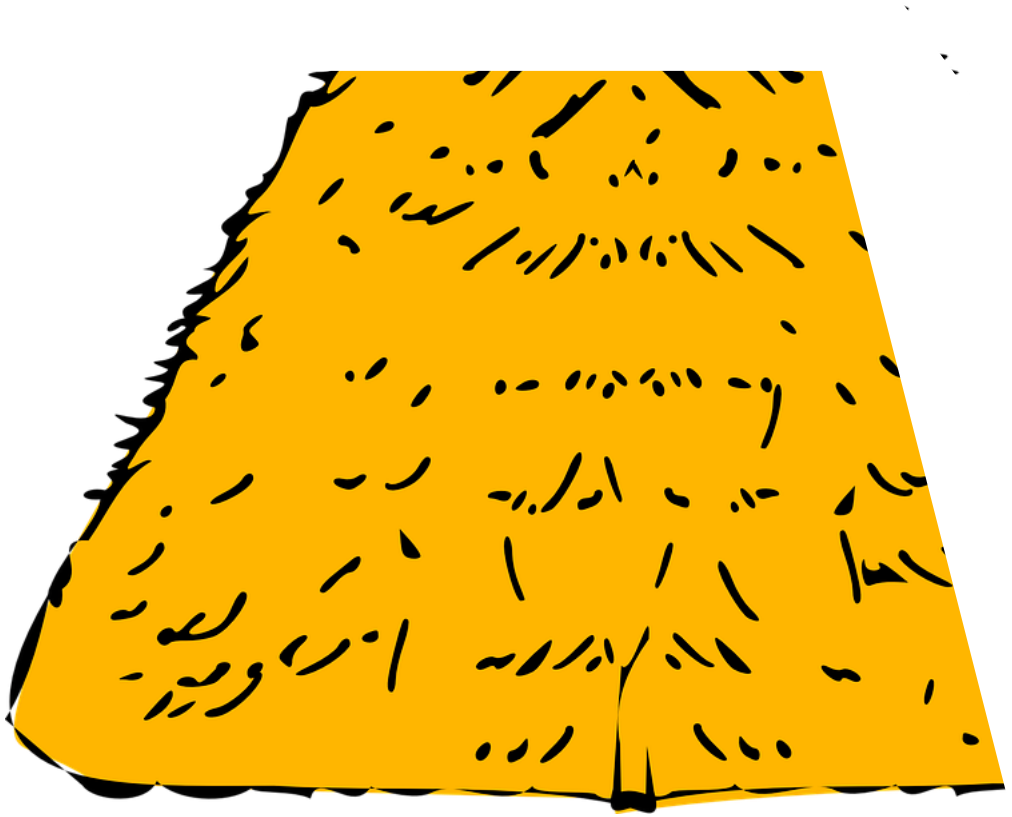
Could we Effectively Search that Space?



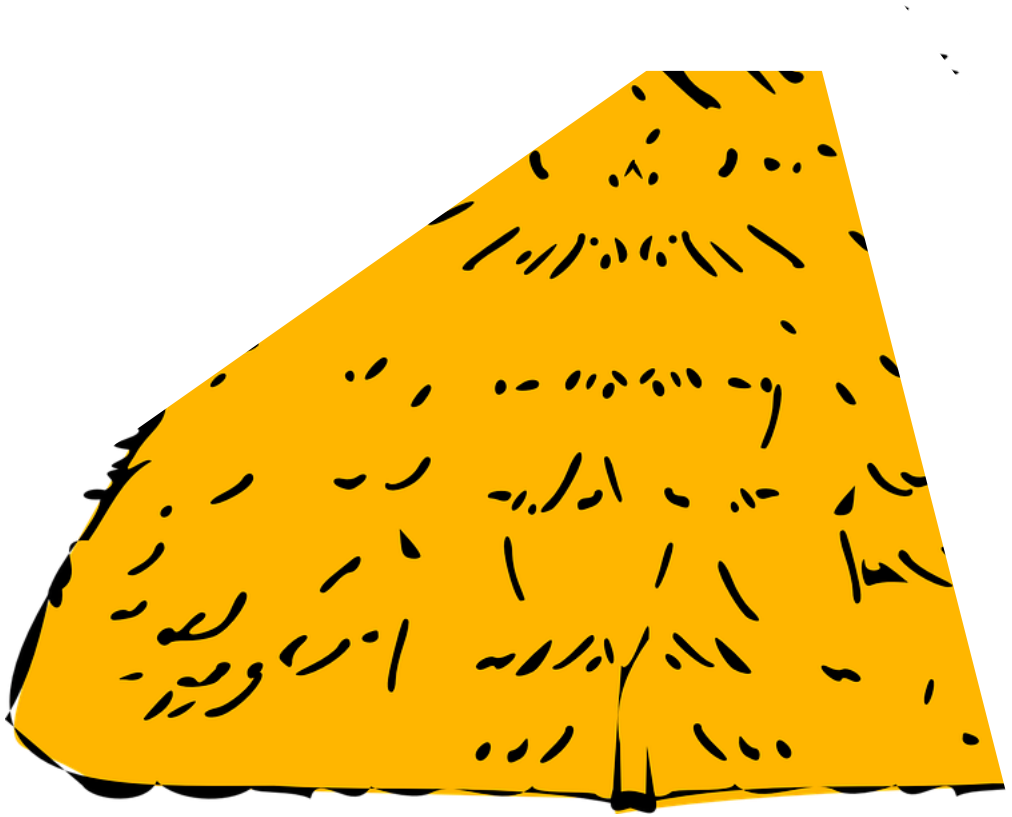
Could we Effectively Search that Space?



Could we Effectively Search that Space?



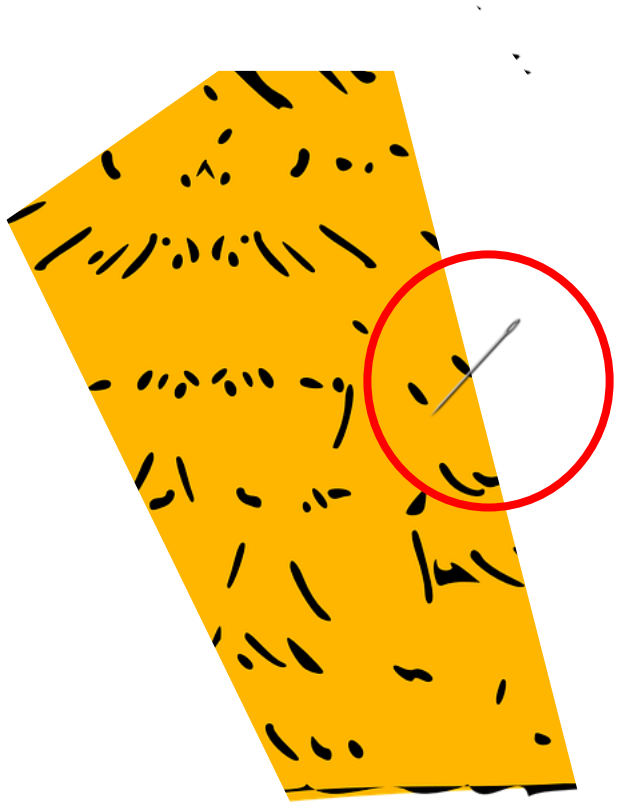
Could we Effectively Search that Space?



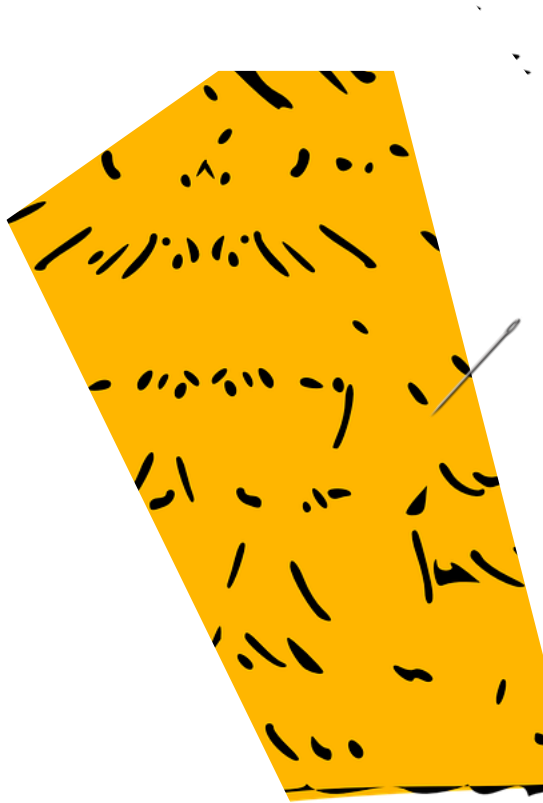
Could we Effectively Search that Space?

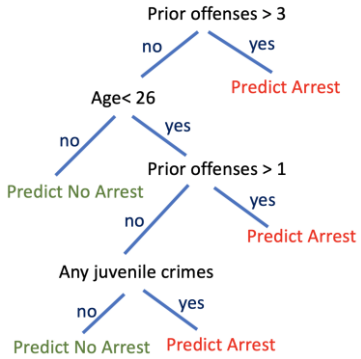
Could we Effectively Search that Space?

Could we Effectively Search that Space?



Could we Effectively Search that Space?





The Optimization Problem

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

The Optimization Problem

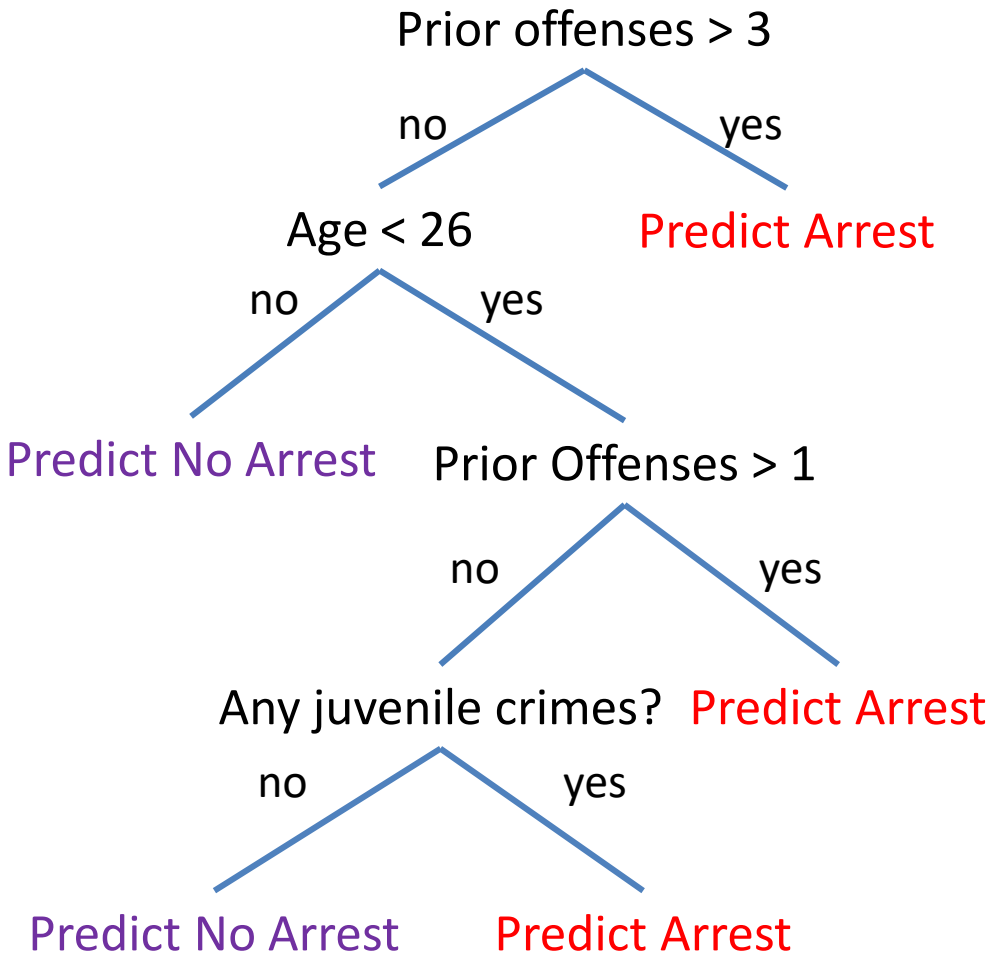
$$\hat{L}(\text{tree},\{(x_i,y_i)\}_i) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} 1_{[\text{tree}(x_i)^1\ y_i]}}_{\text{Misclassification error}} + C(\#\,\text{leaves in tree})$$

The Optimization Problem

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

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

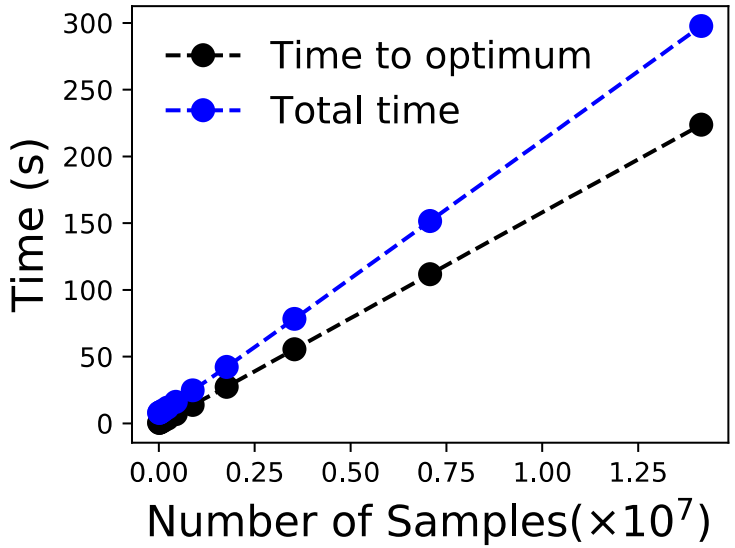
+

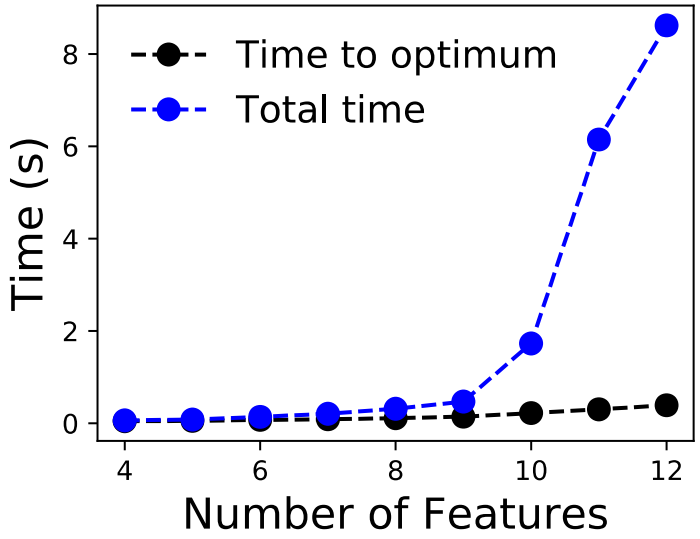
Strong Bounds

+

Incremental Computation

=





FAST

Optimal Sparse Decision Trees

Branch and Bound

+

Good Scheduling Order

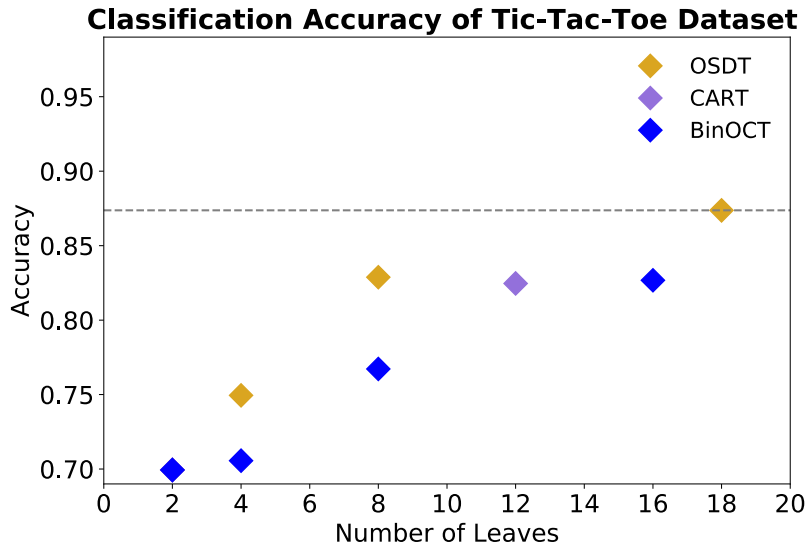
+

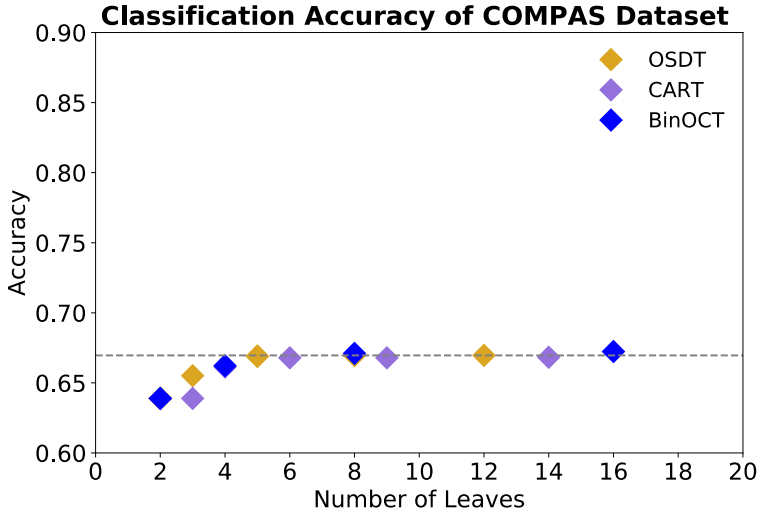
Strong Bounds

+

Incremental Computation

=



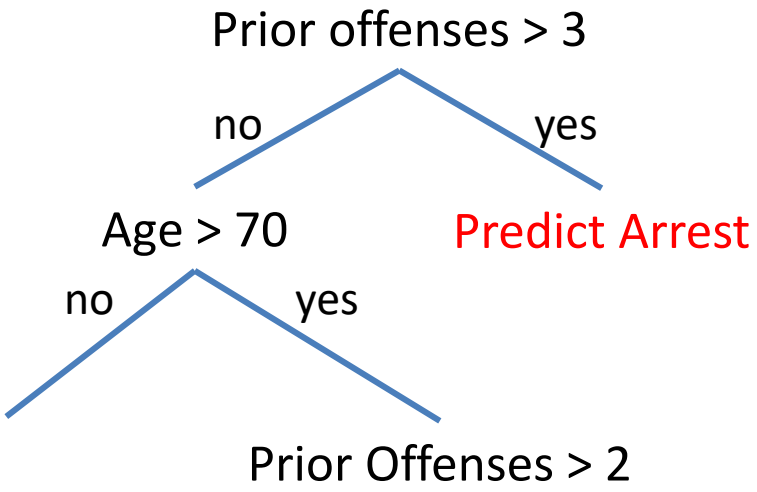


Accurate

Bounding the Search Space

Lower Bound on Node Support

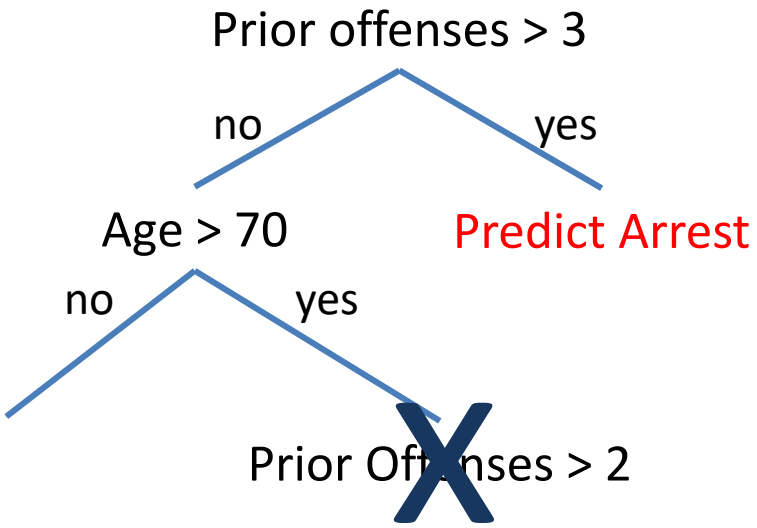
Theorem: For an optimal tree, the support of each node must be above $2C$.



Bounding the Search Space

Lower Bound on Node Support

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

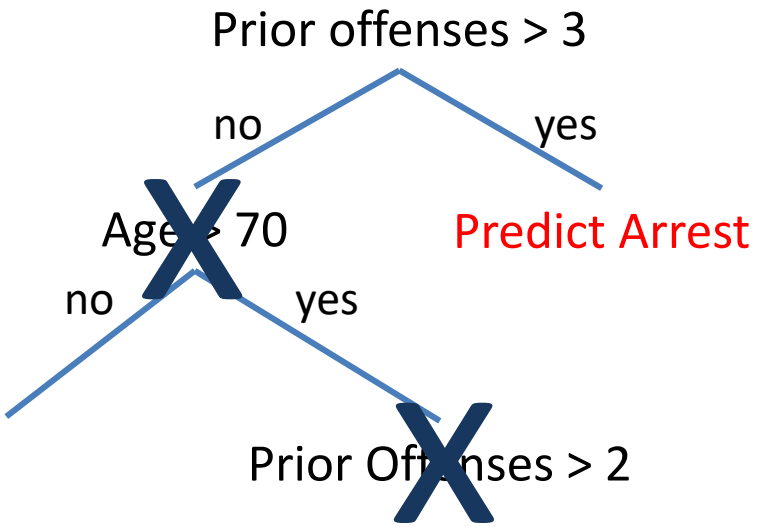


Node support insufficient to produce optimal solution

Bounding the Search Space

Lower Bound on Node Support

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

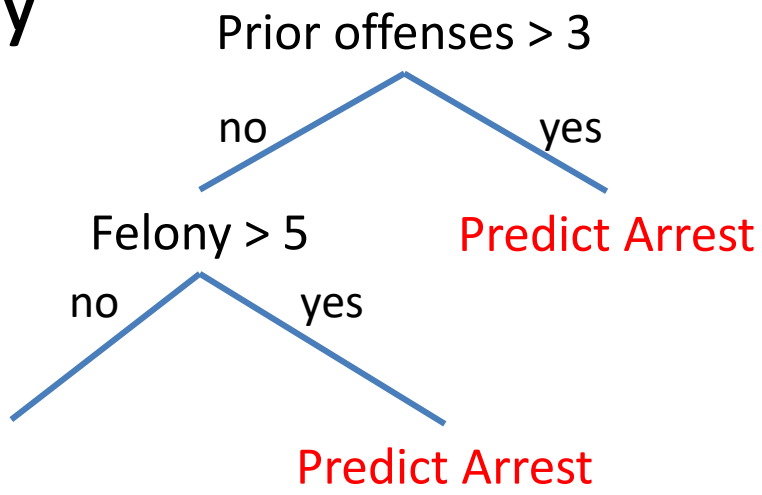


Node support insufficient to produce optimal solution

Bounding the Search Space

Lower Bound on Classification Accuracy

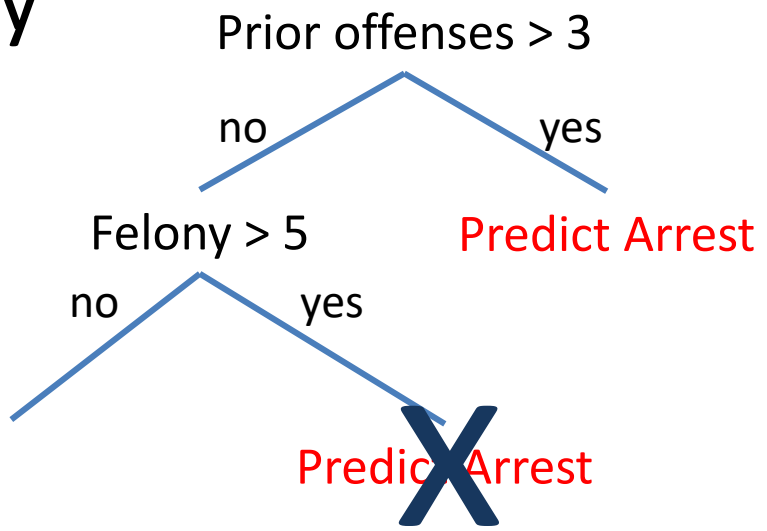
Theorem: Each leaf of an optimal tree correctly classifies at least fraction *C* of the data



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

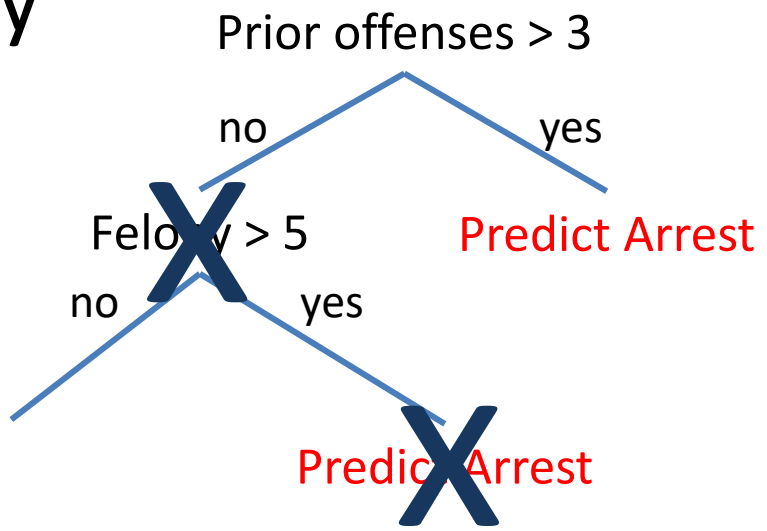


Doesn't classify at least Cn points correctly.

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

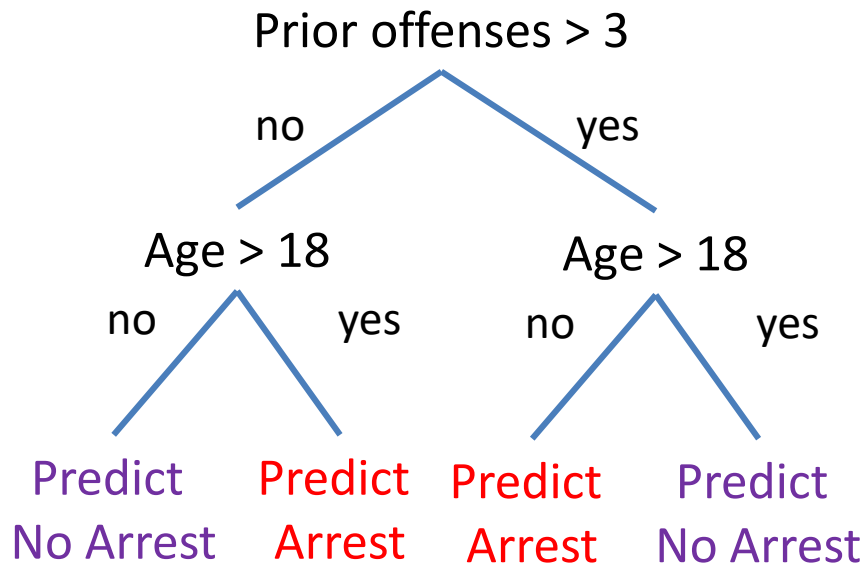


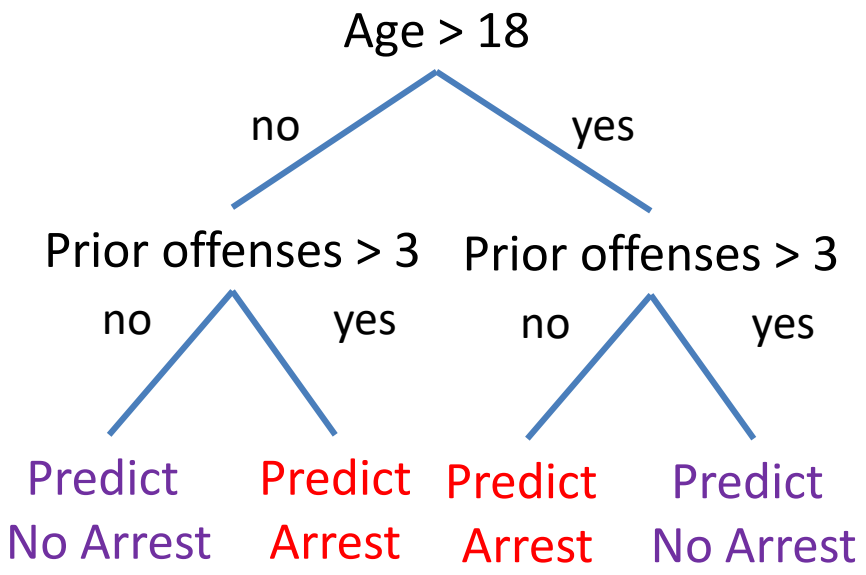
Doesn't classify at least Cn points correctly.

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.

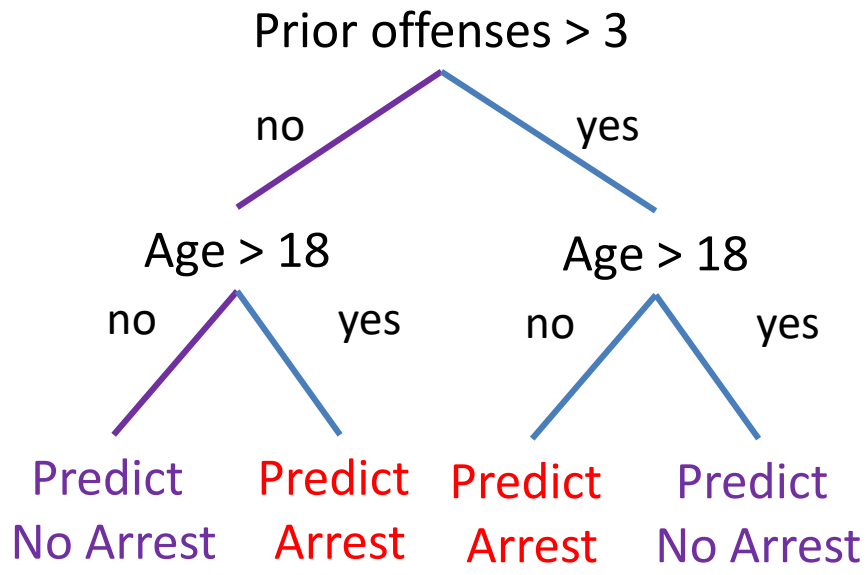


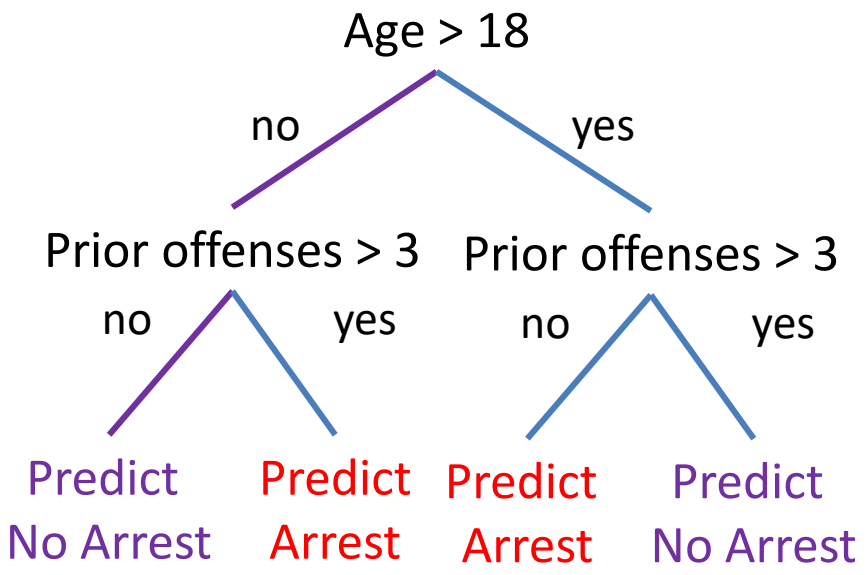


Bounding the Search Space

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.

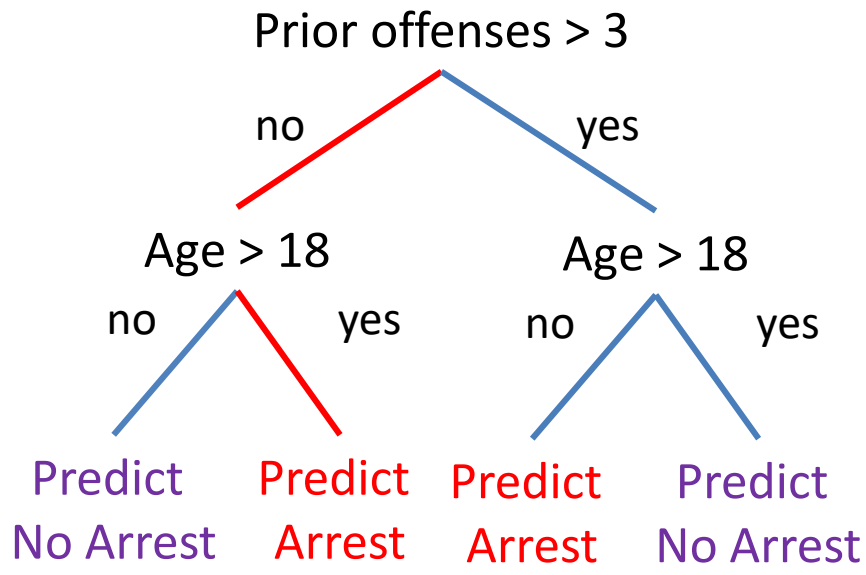


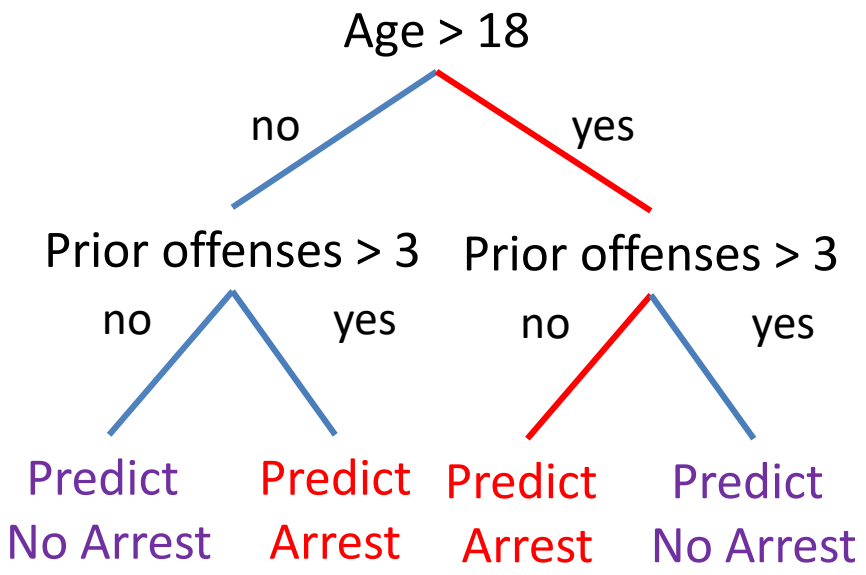


Bounding the Search Space

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.

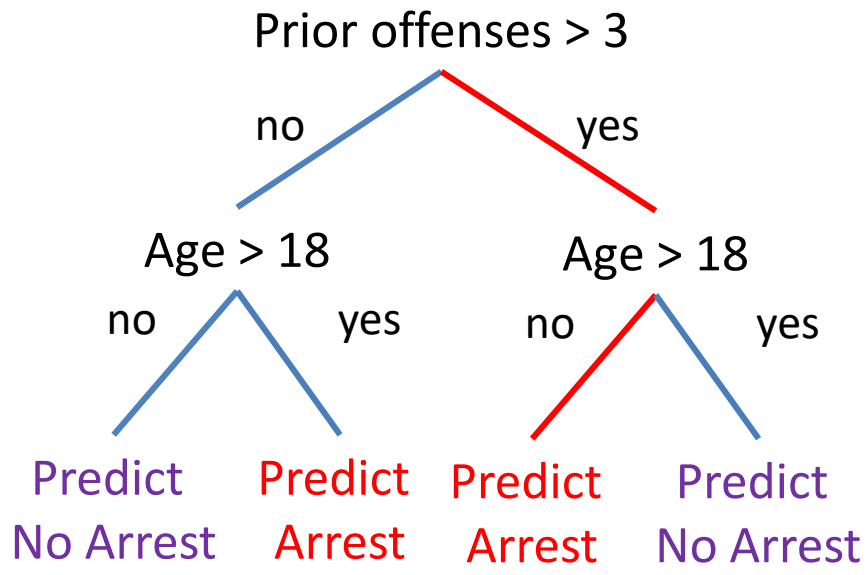


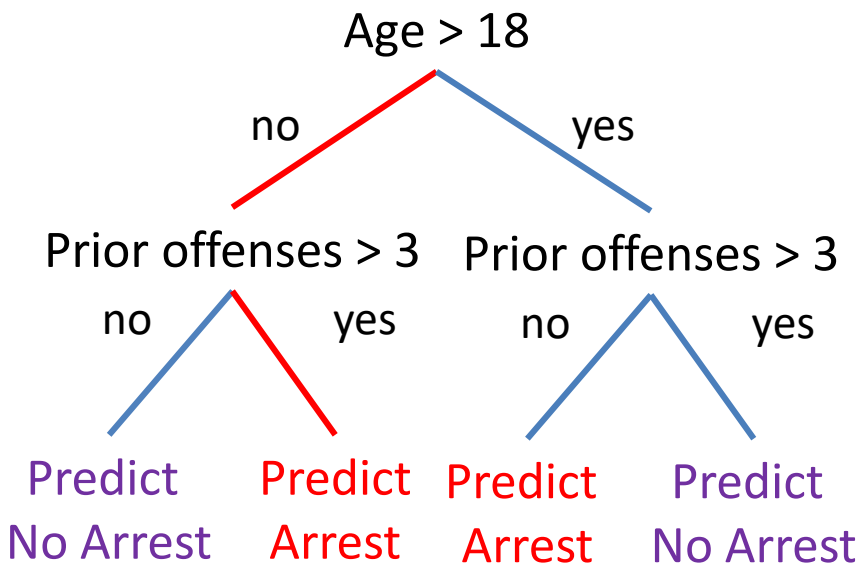


Bounding the Search Space

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.

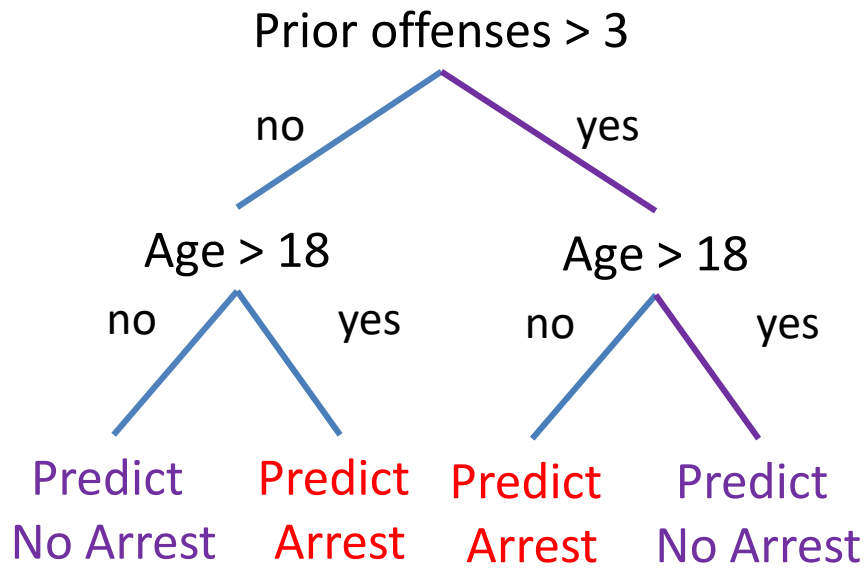


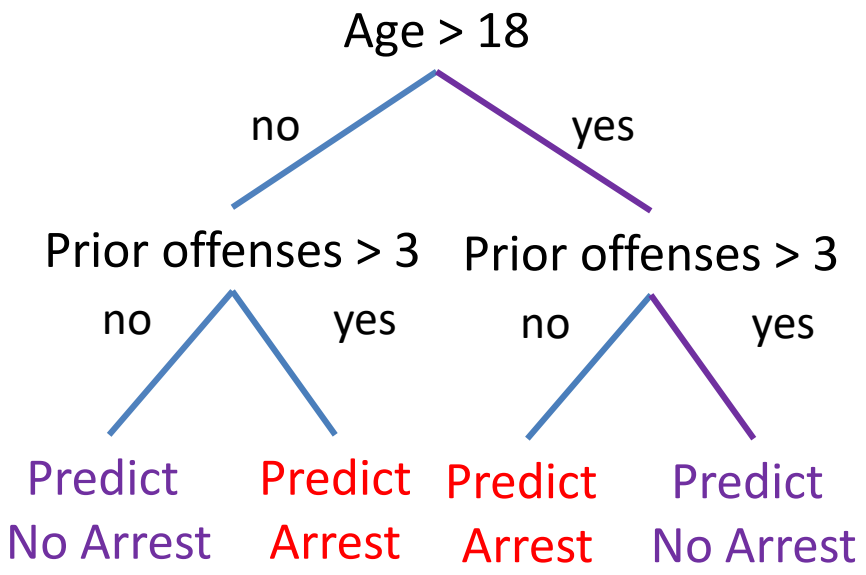


Bounding the Search Space

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

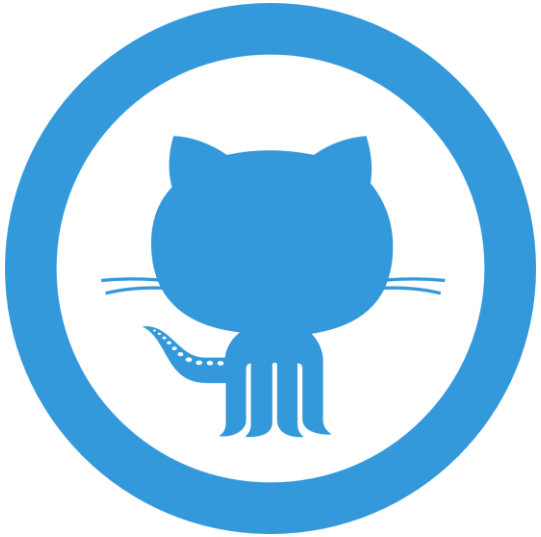
- Other bounds enable even more pruning
  - Equivalent points bound: Samples with the same features, but different predictions will produce misclassifications regardless of model.
  - Bound on the number of leaves: Regularization value bounds the number of leaves.

Optimal Sparse Decision Trees

Accurate

Open Source

FAST



Interpretable

https://github.com/xiyanghu/OSDT