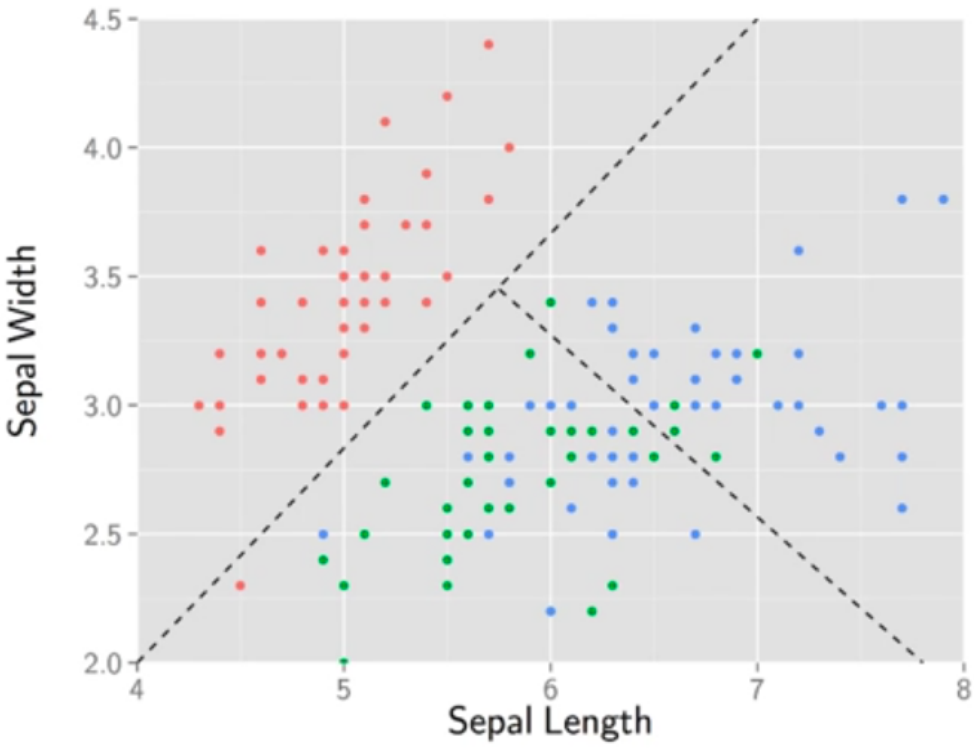
Optimal Trees

- CART's greedy training means splits are only are only locally-optimal, overall tree could be far from optimal
- MIT researchers have developed a new method for finding optimal decision trees:
  - Bertsimas and Dunn. "Optimal Classification Trees". Machine Learning, 2017
- Our method uses modern optimization techniques to train the entire tree in one step, rather than split-by-split like existing methods

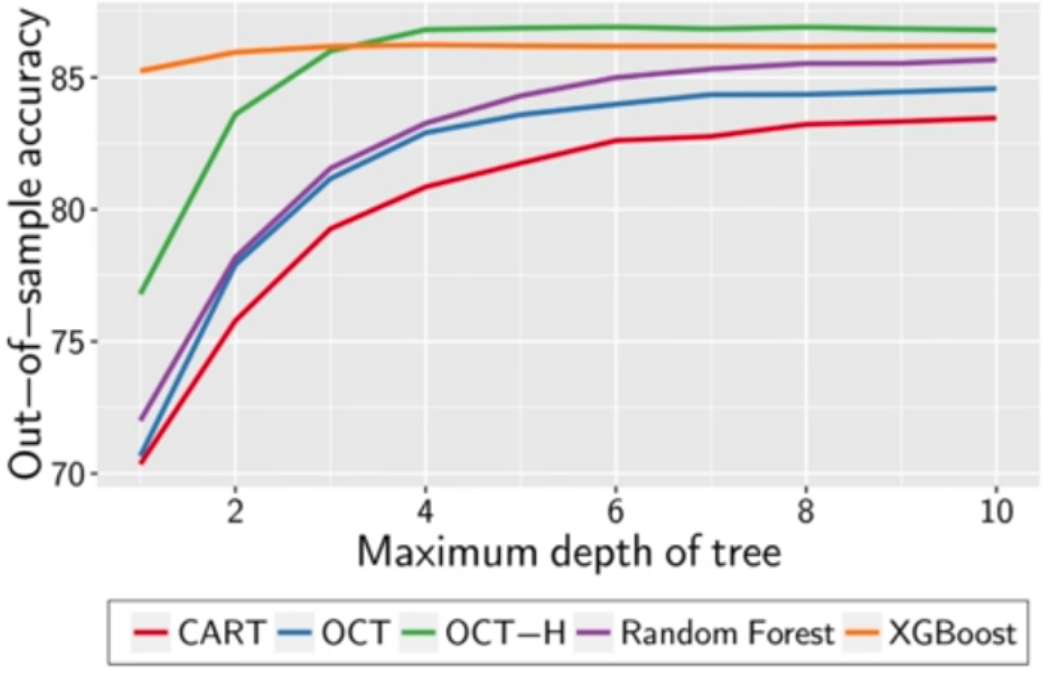
Variants of Optimal Trees

- **OCT**: trees with parallel splits (one variable per split)
- **OCT-H**: trees with hyperplane splits (can use multiple variables per split if beneficial)



Performance of OCT

- Average performance across 60 datasets:



Summary of Optimal Trees

- Practitioners often have to choose between interpretability (CART) or performance (random forest)

- Optimal Trees is a new method that maintains interpretability but delivers state-of-the-art performance