OVERFITTING AND PRUNING

SYRACUSE UNIVERSITY
School of Information Studies

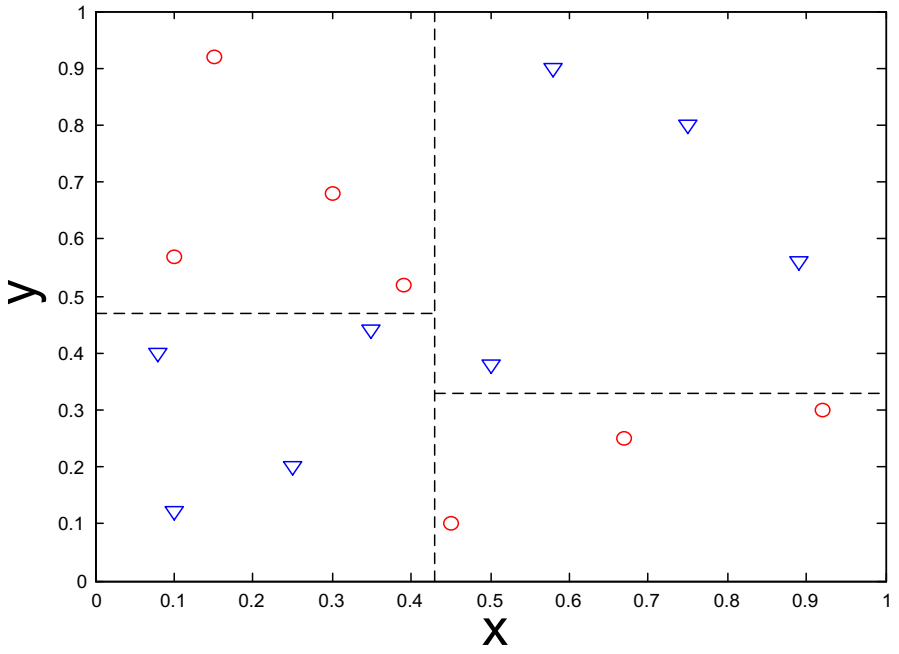
CHARACTERISTICS OF DECISION TREE INDUCTION

Decision tree (DT) is a nonparametric algorithm, meaning, it does not require any prior assumptions regarding the type of probability distributions satisfied by the class and other attributes.

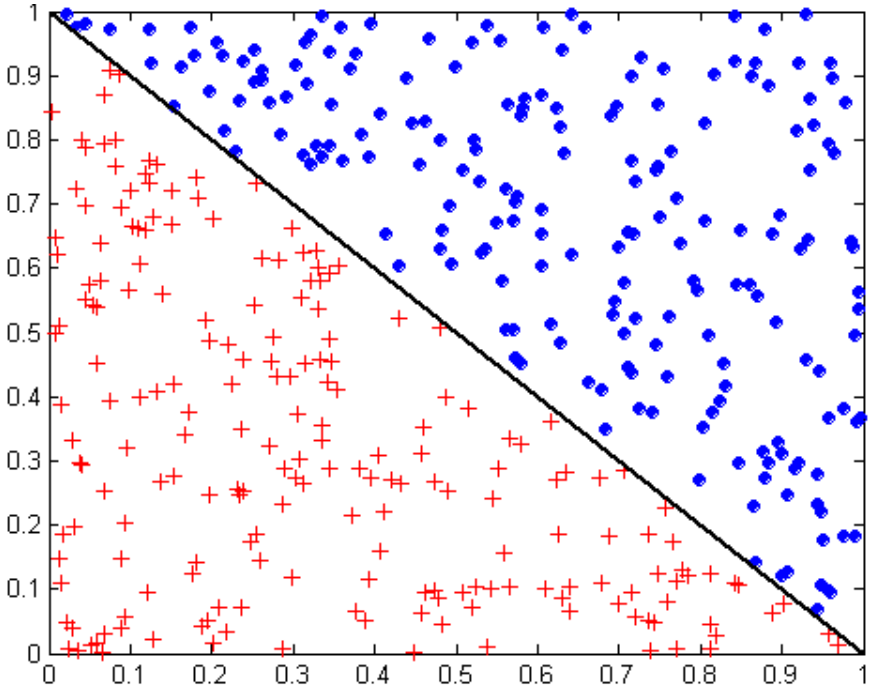
Linear classification algorithms are parametric algorithms because they assume the decision boundary is linear, such as a line in two-dimensional space.

“Decision boundary” means the border between two neighboring regions of different classes.

THERE IS NO SILVER BULLET



Nonlinear



Linear

MODEL OVERFITTING

Decision trees have the particular problem of overfitting.

There may not be enough examples to fully represent all possible cases that may arise in the future.

If decision tree is fully developed, it may be too detailed a fit to the training data and lead to more errors on the test data.

E.g., assume we are looking for patterns of buyers for a certain product. In the training data set, no women purchased a product; the DT algorithm may learn a pattern that "if women, no purchase." But this training data set included very few women, and actually, there were women who bought this product. In such cases, the DT model overfit the training data and lost precision in future prediction.

Occam's razor (preference of small trees)

MODEL OVERFITTING

Generally speaking, complex models are more likely to overfit than simple models.

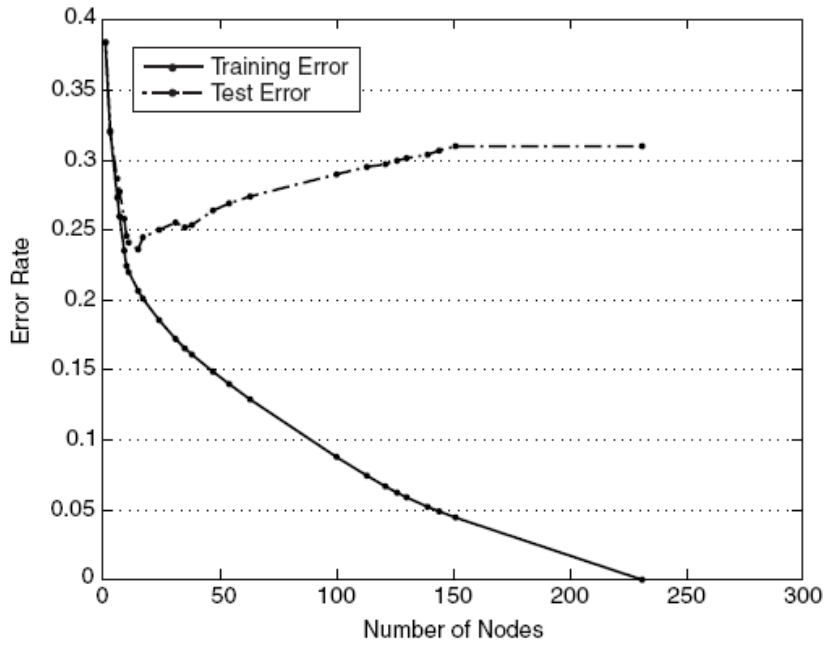


Figure 4.23. Training and test error rates.

For decision tree, number of nodes indicates model complexity.

In this figure, the higher the number of nodes, the lower the training error and the higher the test error, meaning, increasingly complex models are increasingly overfitting.

OVERFITTING AND TREE PRUNING

Two approaches to avoid overfitting:

Prepruning: Halt tree construction early—do not split a node if information gain falls below a threshold.

- Difficult to choose an appropriate threshold

Postpruning: Remove branches from a "fully grown" tree—get a sequence of progressively pruned trees.

- Use a set of data different from the training data to decide which is the "best pruned tree"

SUMMARY OF DECISION TREES

Strengths of decision trees are that they are:

- Fast in prediction
- Interpretable patterns
- Robust to noise

Weaknesses of decision trees are that they:

- Tend to overfit (pruning helps)
- Are error prone with too many classes
- Are computationally expensive in training (compared to the low cost in prediction)