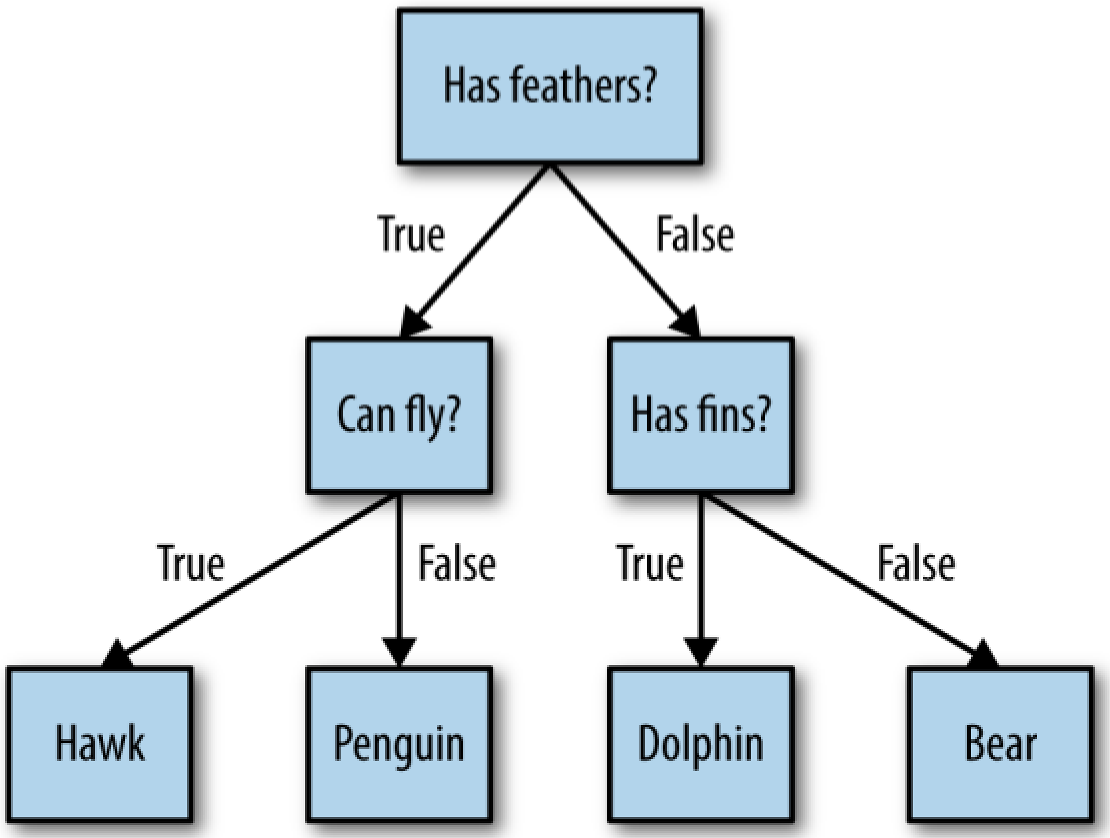
BM5702 MAKİNE ÖĞRENMESİNE GİRİŞ

Hafta 9

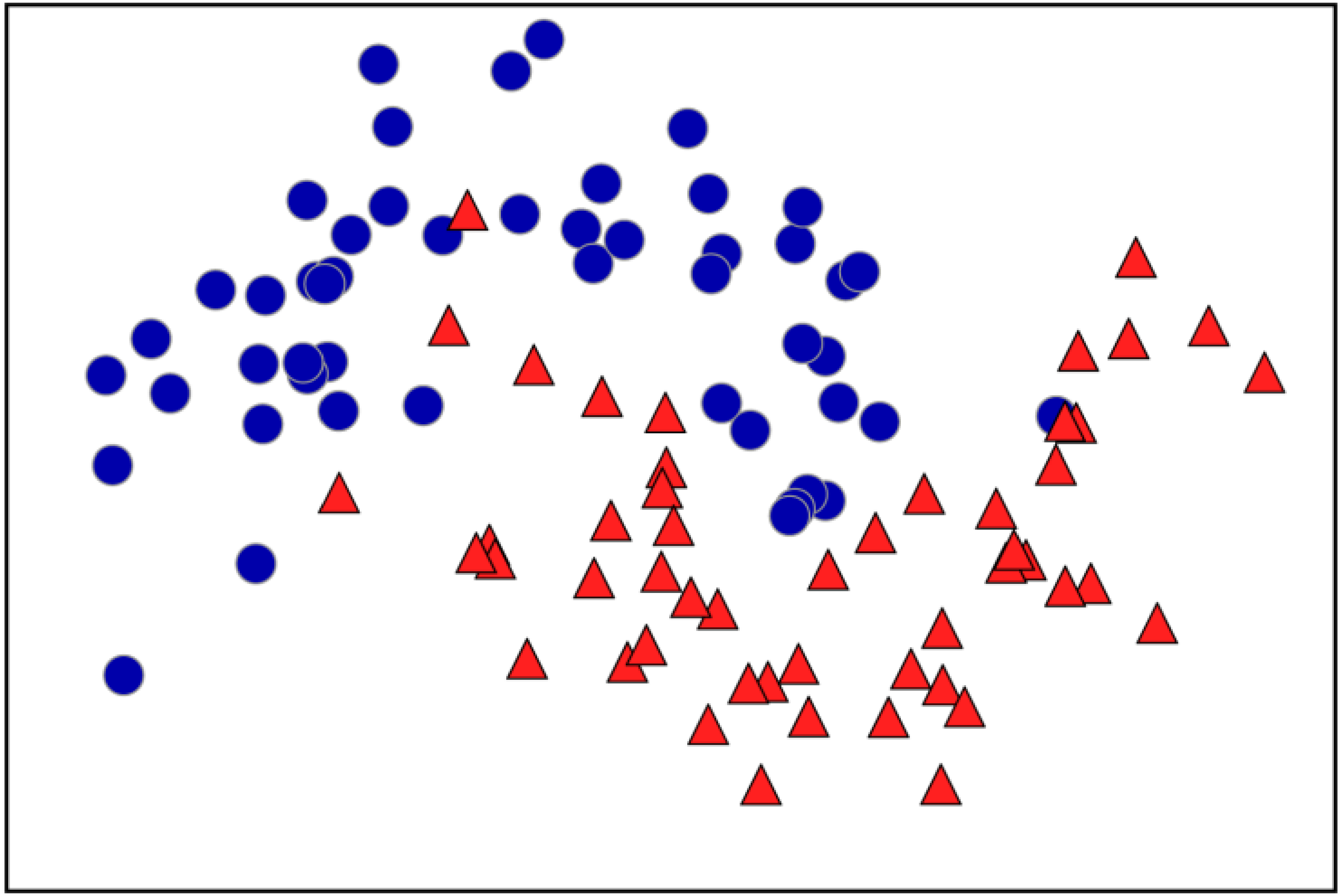
Doç. Dr. Murtaza CİCİOĞLU

Decision Trees

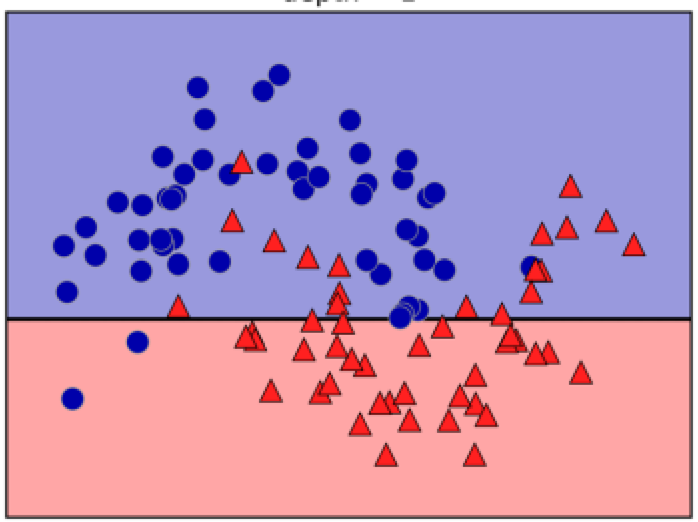
- Decision trees are widely used models for classification and regression tasks. Essentially, they learn a hierarchy of if/else questions, leading to a decision.

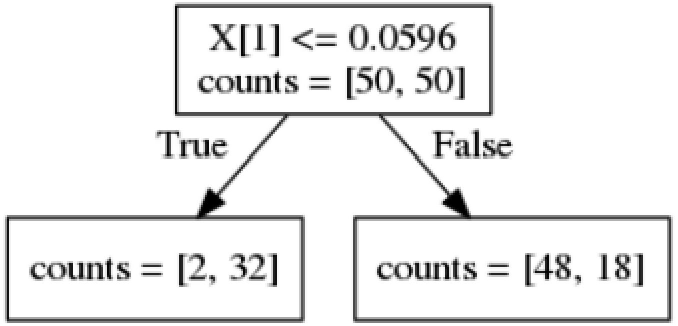


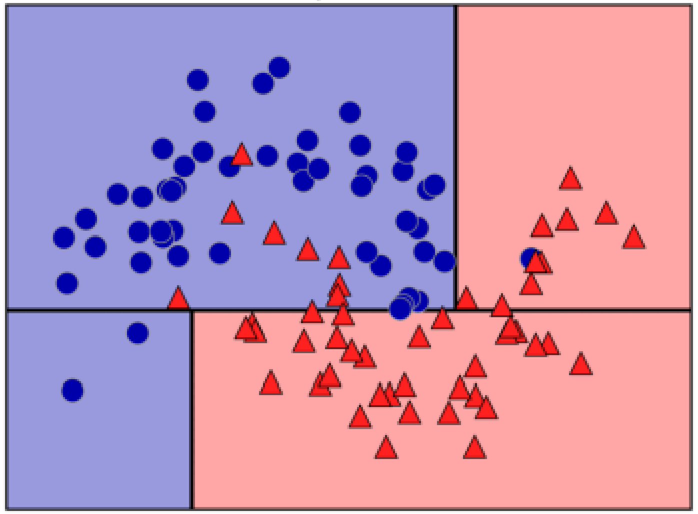
Two-moons dataset on which the decision tree

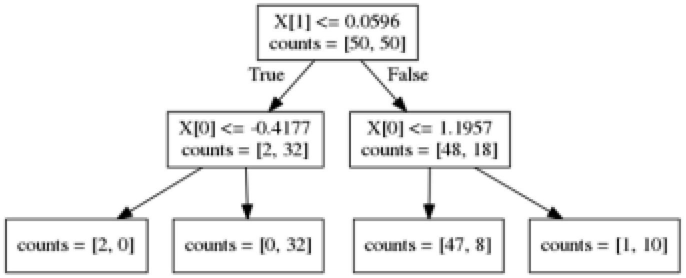


Decision boundary of tree with depth 1



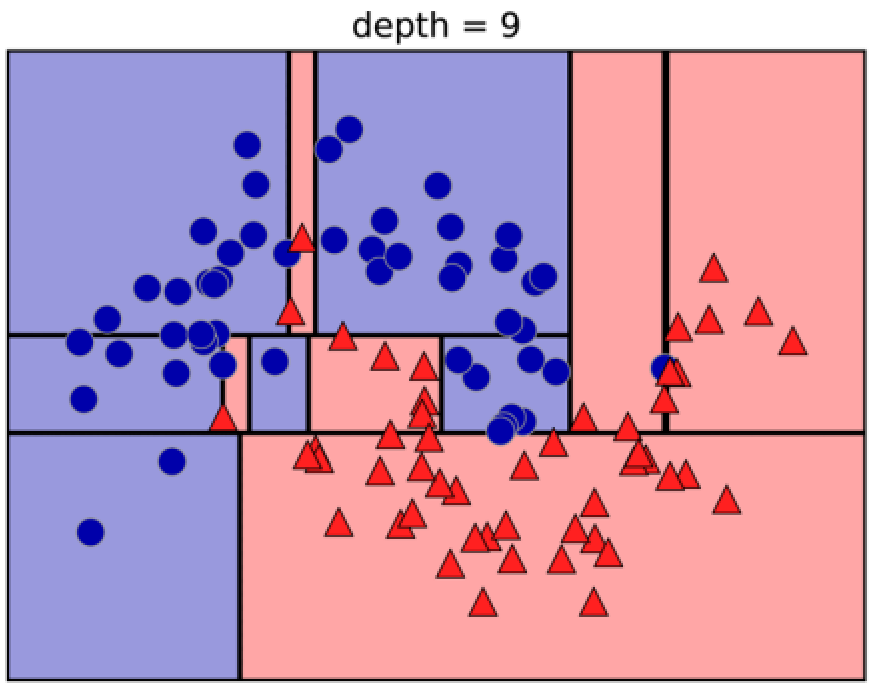


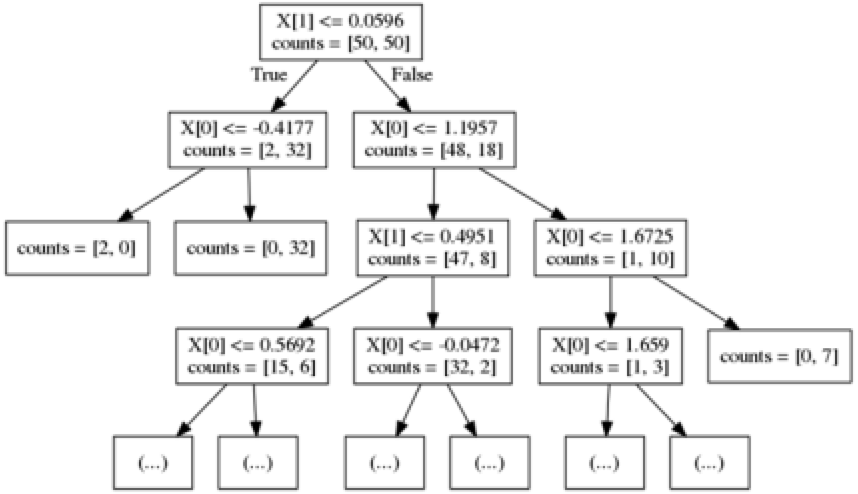




Decision boundary of tree with depth 1

- The recursive partitioning of the data is repeated until each region in the partition (each leaf in the decision tree) only contains a single target value (a single class or a single regression value). A leaf of the tree that contains data points that all share the same target value is called pure

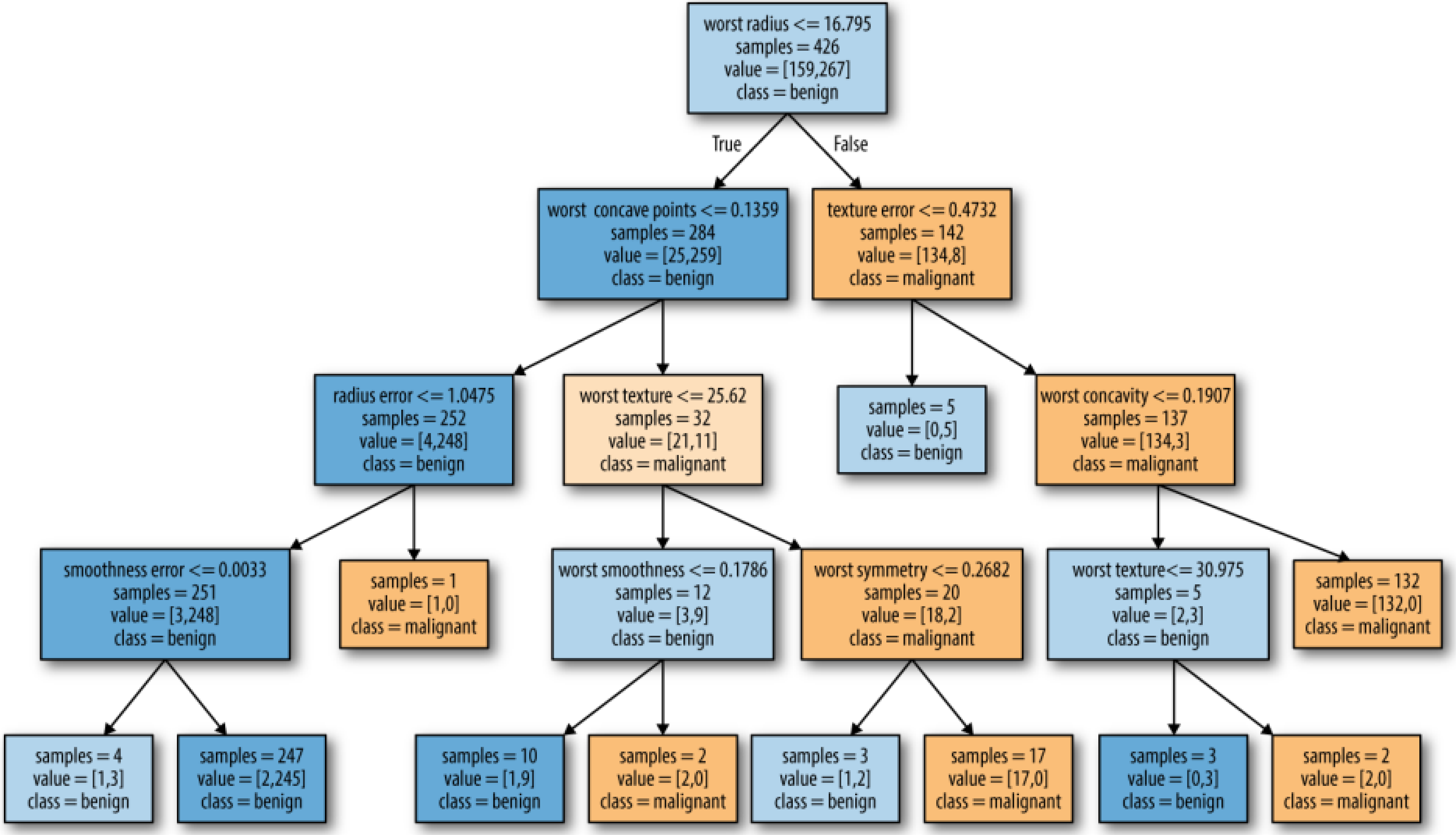




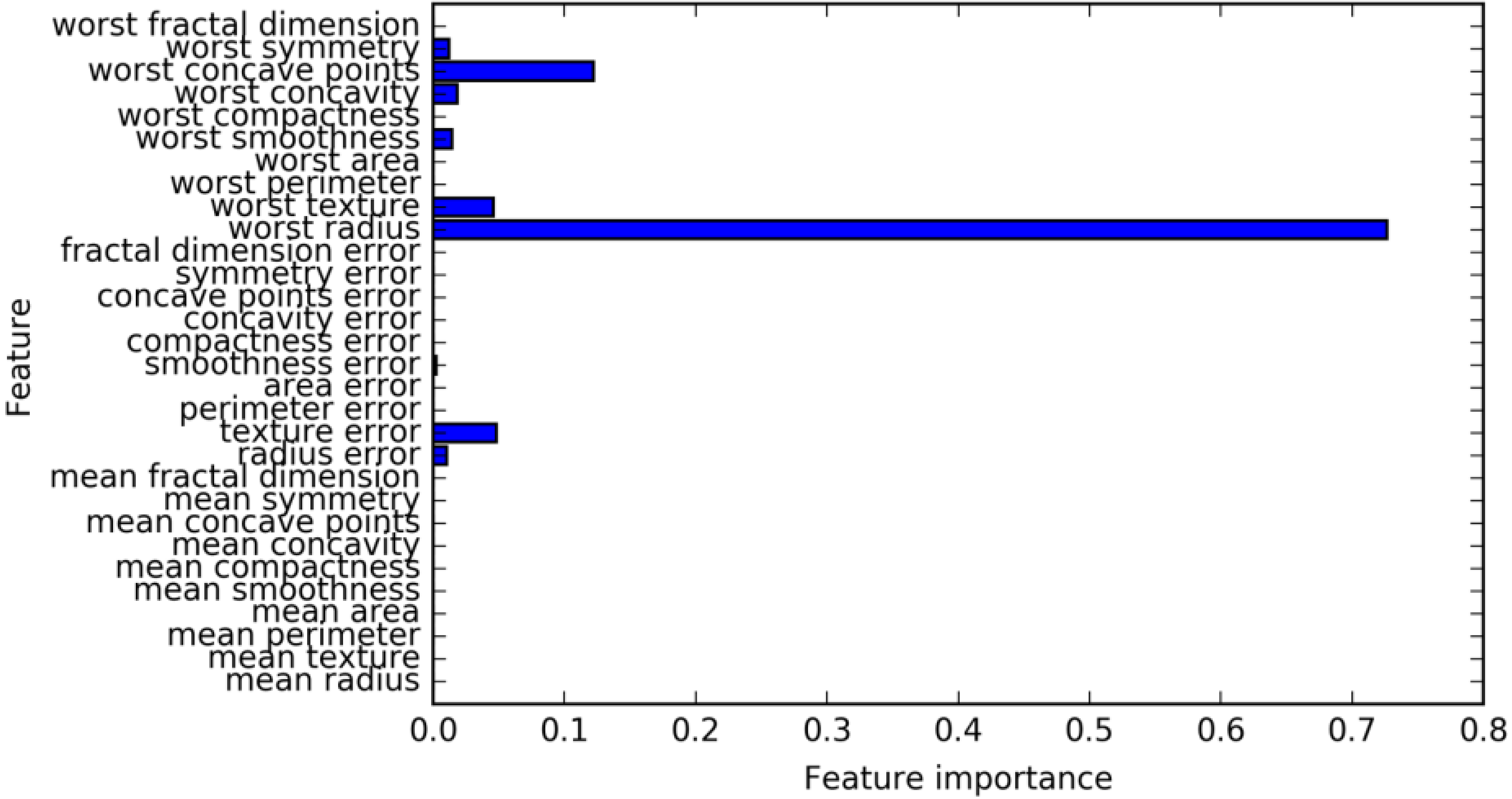
Controlling complexity of decision trees

- Typically, building a tree as described here and continuing until all leaves are pure leads to models that are very complex and highly overfit to the training data. The presence of pure leaves mean that a tree is 100% accurate on the training set; each data point in the training set is in a leaf that has the correct majority class.
- There are two common strategies to prevent overfitting: stopping the creation of the tree early (also called pre-pruning), or building the tree but then removing or collaps- ing nodes that contain little information (also called post-pruning or just pruning). Possible criteria for pre-pruning include limiting the maximum depth of the tree, limiting the maximum number of leaves, or requiring a minimum number of points in a node to keep splitting it.

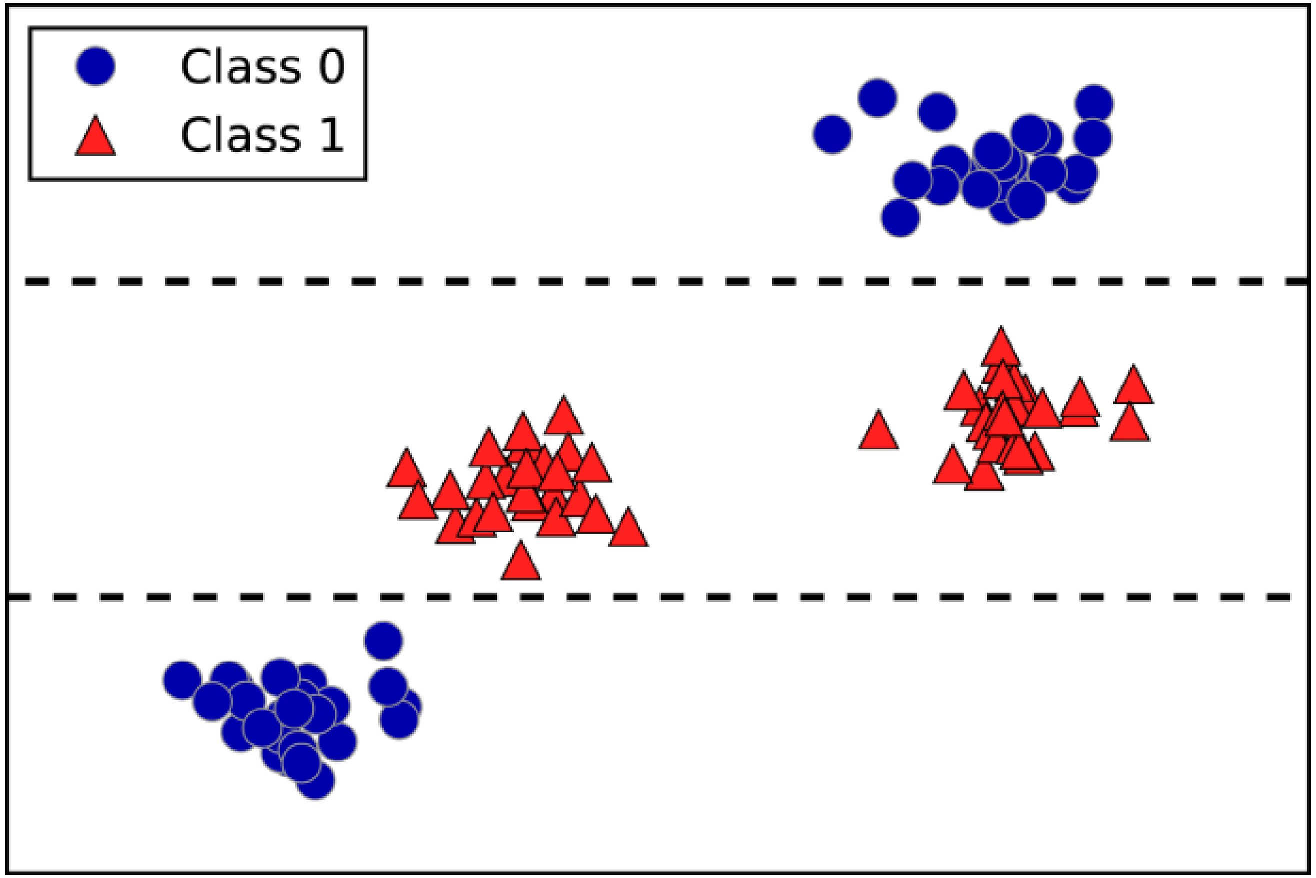
Controlling complexity of decision trees

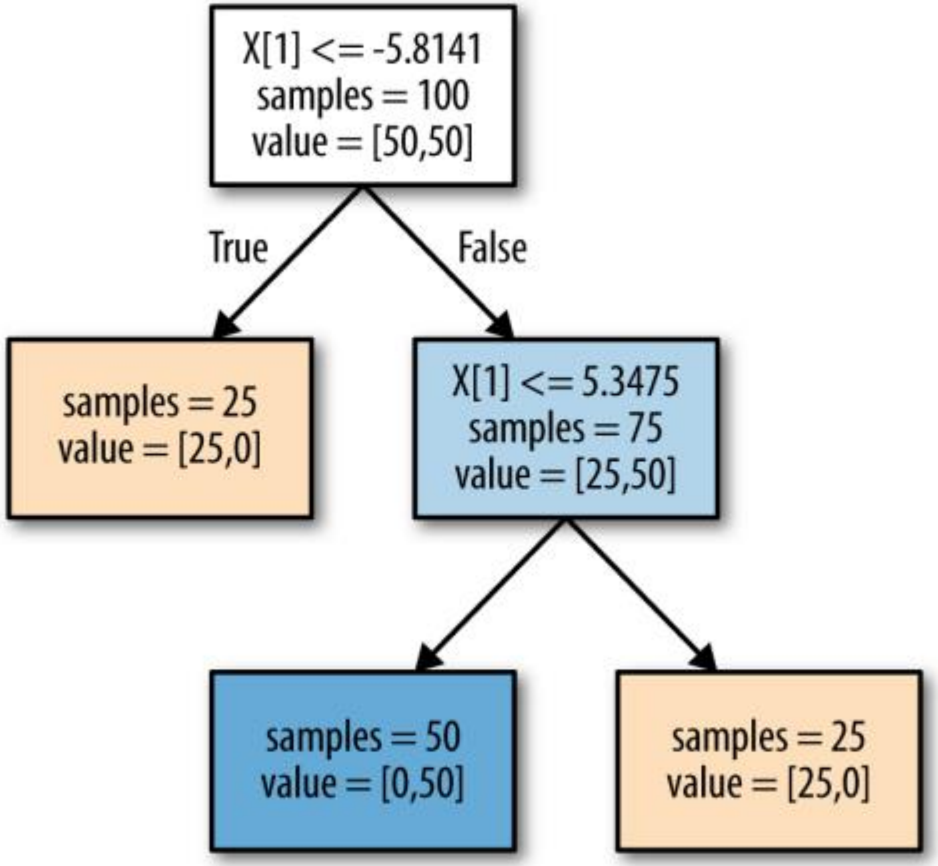


Feature importance in trees



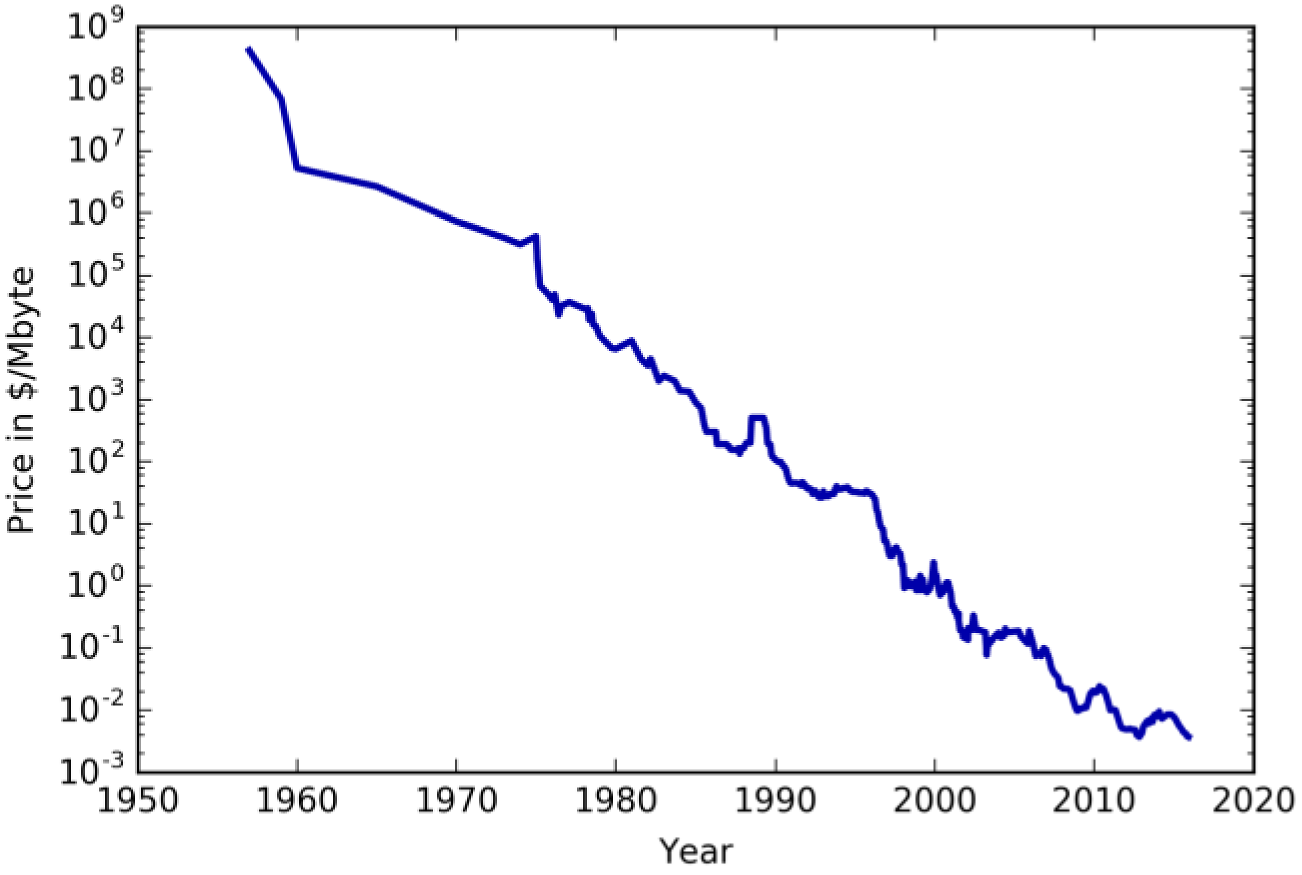
Feature importance in trees





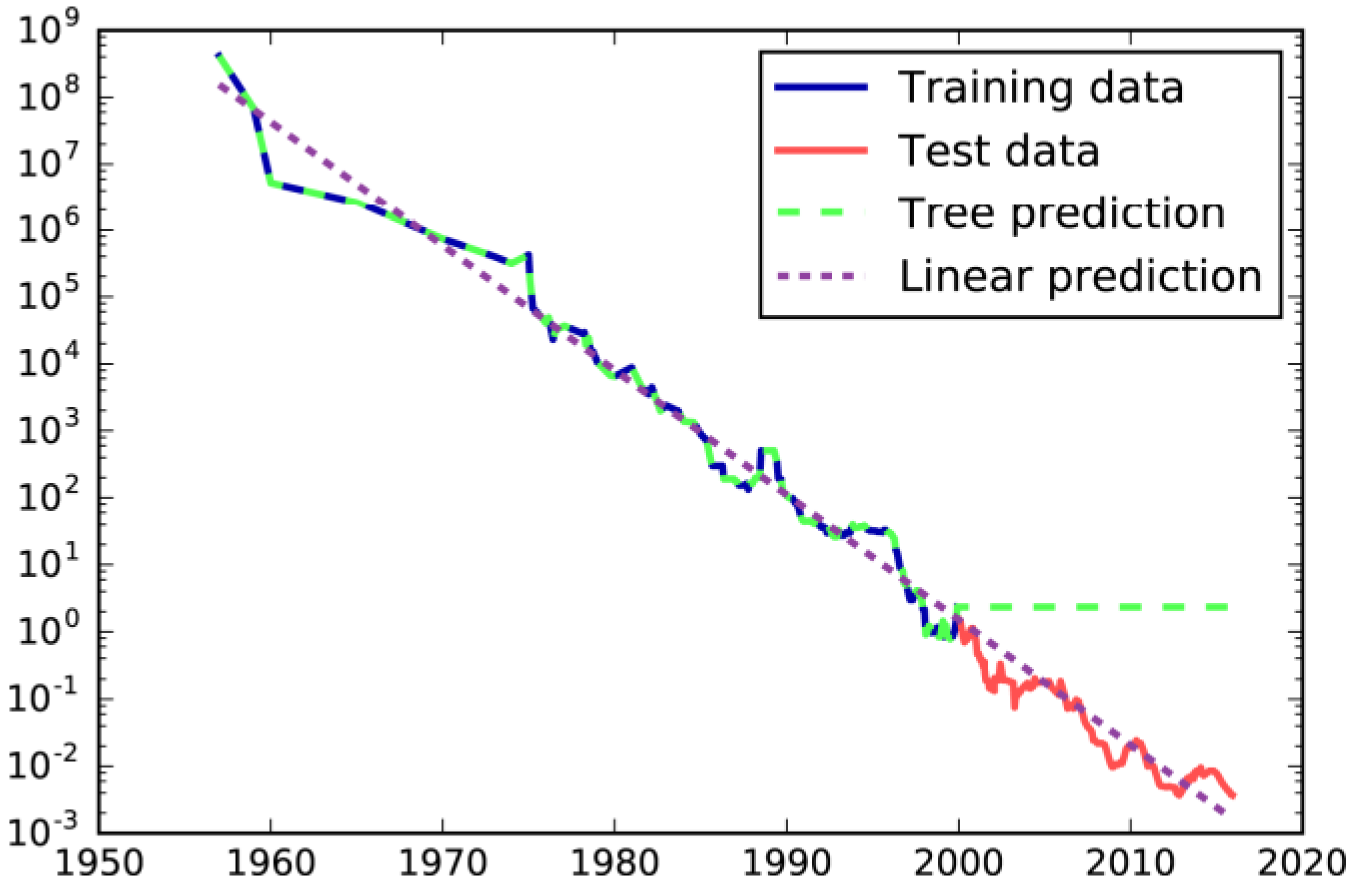
DecisionTreeRegressor

- While we focused our discussion here on decision trees for classification, all that was said is similarly true for decision trees for regression, as implemented in Decision TreeRegressor.



DecisionTreeRegressor

- We will make a forecast for the years after 2000 using the historical data up to that point, with the date as our only feature. We will compare two simple models: a DecisionTreeRegressor and LinearRegression.



Strengths, weaknesses, and parameters

- The parameters that control model complexity in decision trees are the pre-pruning parameters that stop the building of the tree before it is fully developed. Usually, picking one of the pre-pruning strategies—setting either max_depth, max_leaf_nodes, or min_samples_leaf—is sufficient to prevent overfitting.
- Decision trees have two advantages over many of the algorithms: the resulting model can easily be visualized and understood by nonexperts (at least for smaller trees), and the algorithms are completely invariant to scaling of the data.
- As each feature is processed separately, and the possible splits of the data don't depend on scaling, no preprocessing like normalization or standardization of features is needed for decision tree algorithms.

Strengths, weaknesses, and parameters

- In particular, decision trees work well when you have features that are on completely different scales, or a mix of binary and continuous features.
- The main downside of decision trees is that even with the use of pre-pruning, they tend to overfit and provide poor generalization performance.