ID3 greedy algorithm, grow tree top down.

Hypothesis space : complete space

No backtracking

Properties: less sensitive to errors( scan through all data)

Inductive bias: shorter trees are preferred. High info gain attributes are preferred ---- preference bias/ search bias

Occam’s razor

Overfitting:

1.def

2.avoid overfitting

a. stop growing the tree earlier

b. post-prune

1. Use a separate set of examples, distinct from the training examples, to evaluate the utility of post-pruning nodes from the tree. ----- training and validation set approach

-----reduced-error pruning.

Pruning a decision node consists of removing the subtree rooted at that node, making it a leaf node, and assigning it the most common classification of the training examples affiliated with that node.

Rule post pruning

**1.** Infer the decision tree from the training set, growing the tree until the training data is fit as well as possible and allowing overfitting to occur.

2. Convert the learned tree into an equivalent set of rules by creating one rule for each path from the root node to a leaf node.

**3.** Prune (generalize) each rule by removing any preconditions that result in improving its estimated accuracy.

4. Sort the pruned rules by their estimated accuracy, and consider them in this sequence when classifying subsequent instances.

Why convert to rules:

distinguishing among the different contexts in which a decision node is used.

distinguishing among the different contexts in which a decision node is used.

Improve readability

Alternative measures for selecting attributes(instead of info gain):

Gain ratio: penalize attributes by incorporating a term—split info





Missing values handling:

1. Assign most common value
2. Assign probability to each of possible values ( weight average), fractionally down to branch

Handling attributes with differing costs: