WEKA J48 TUTORIAL

SYRACUSE UNIVERSITY
School of Information Studies

J48 ALGORITHM

J48 is an implementation of the famous C4.5 algorithm to construct decision tree.

J48 provides many parameters to tune.

Every time you tune a parameter, a new decision model will be created.

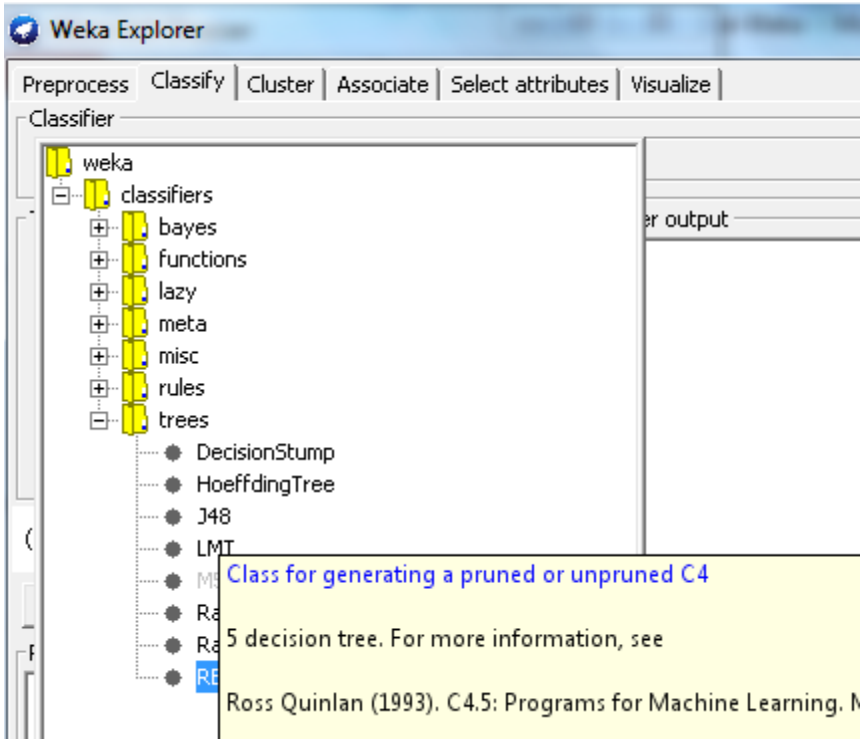
USE J48 IN WEKA

Two key questions:

How to tune parameters to get better models?

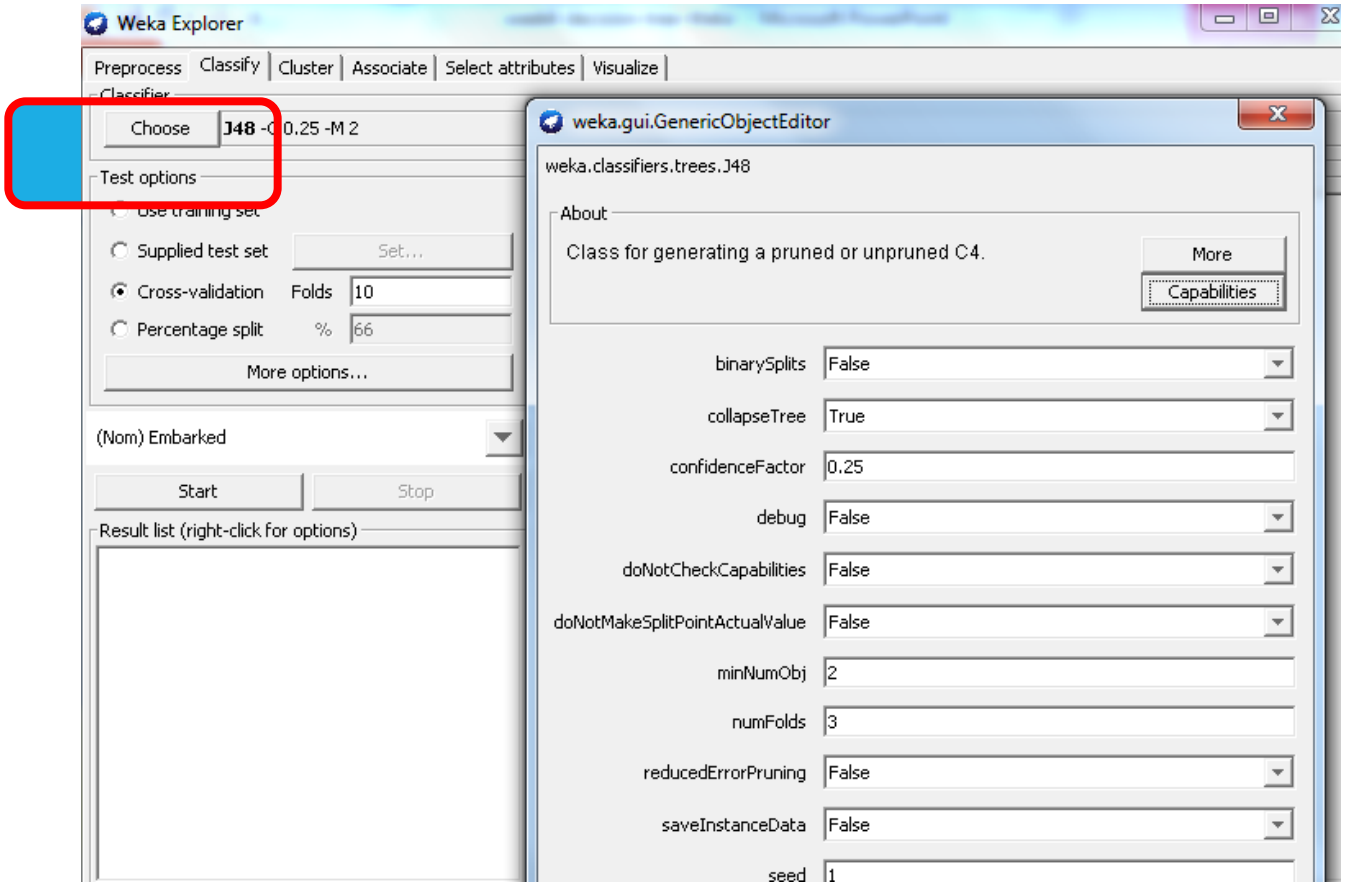
How to evaluate which model is the best?

HOW TO TUNE PARAMETERS TO GET BETTER DECISION TREE MODELS



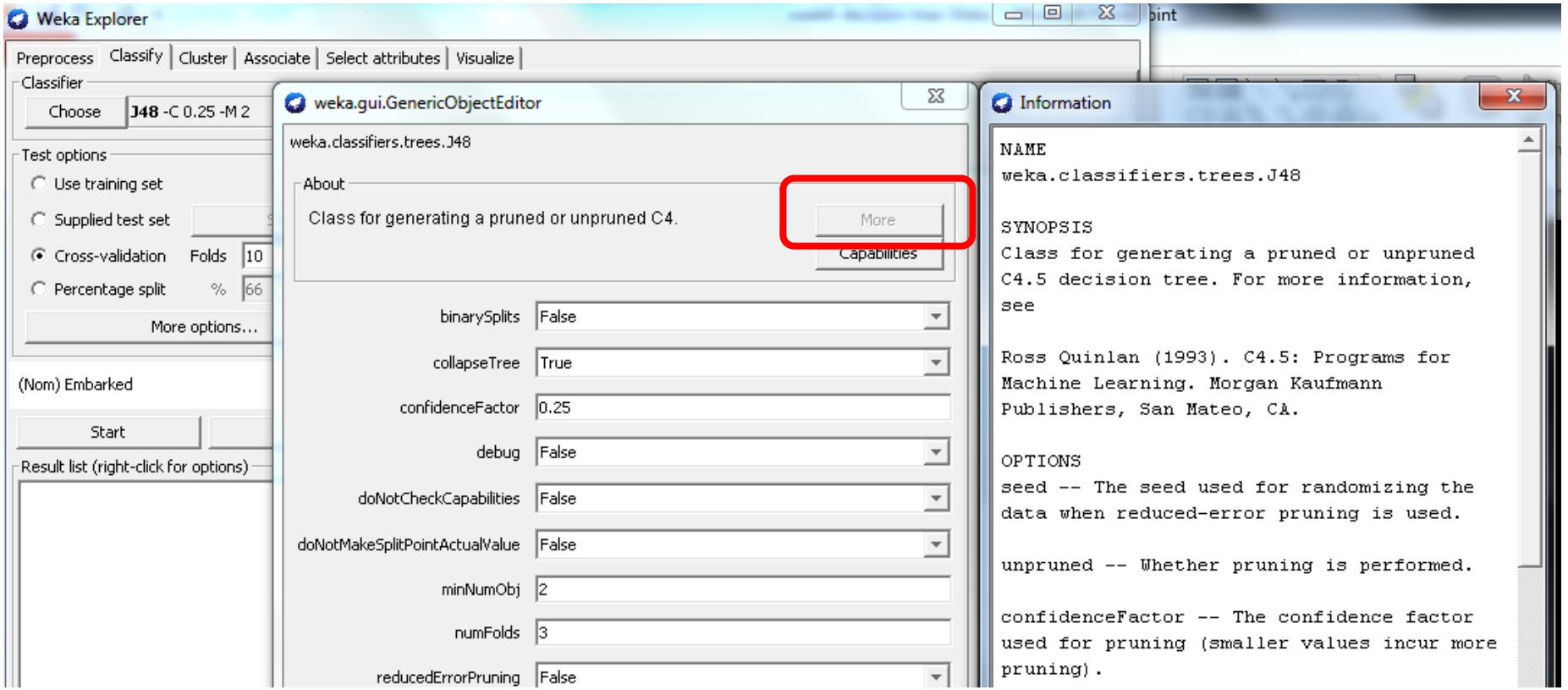
Step 1: Find J48 algorithm, located in the “Classify” tab, under “Trees.”

J48 PARAMETER PANEL



Click the algorithm (in red box) to pop up the parameter panel.

J48 PARAMETER MEANING



Click the “More” button (in red box) to pop up the explanation for the parameters.

J48 PARAMETERS

Tuning these parameters requires the theoretical knowledge of their purpose and empirical knowledge of their performance on different kinds of data.

Several important parameters to tune

- "BinarySplit": True or False
  - True: A deep tree with two branches at each level
  - False (default): A wide tree with many branches at each level
  - Which one works better? Depends on data

J48 PARAMETERS

"**unpruned**": True or False

True: Grow a tree completely without pruning.

False (default): Prune the tree.

"**ConfidenceFactor**": numeric (0 to 1)

Decide how aggressively to prune the tree.

Smaller values incur more pruning.

Too aggressive pruning results in a too small tree that does not capture all patterns.

Too conservative pruning results in a large tree that overfits the training data.

How to find the balance point?

Use a good evaluation method, such as cross-validation (covered in later slides).

J48 PARAMETERS

"**minNumObj**": Integer (1 to infinity)

The minimum number of examples in the leaf node

Default value: 2

It means all leaves with only one data example are pruned.

Increasing this value would result in more aggressive pruning.

J48 PARAMETERS

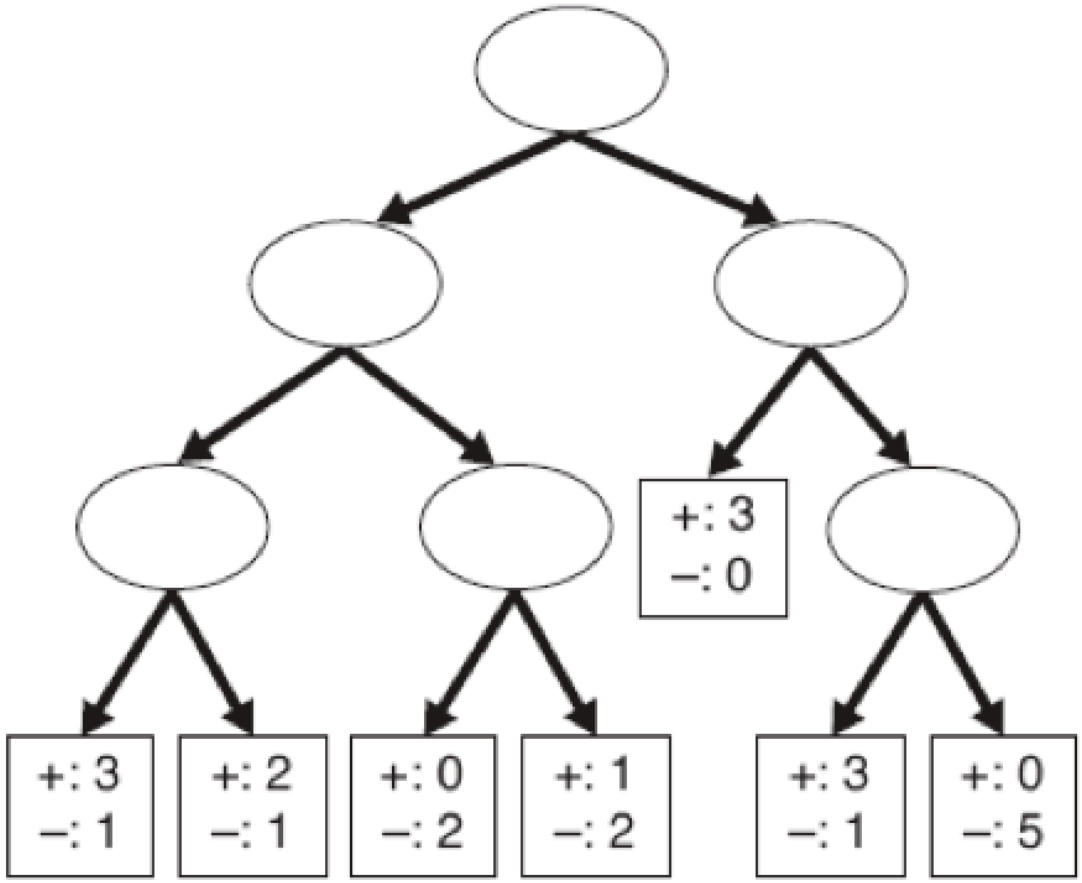
Further techniques to prune a tree

"reducedErrorPruning": Replace a subtree with a leaf node with the most popular category label

"subtreeRaising": A subtree replaces its parent

REDUCED ERROR PRUNING

+:5
-:2



Decision Tree, T_L