Data Mining with Weka

Building a classifier

Ian H. Witten

Building a classifier

Use J48 to analyze the glass dataset

❖ Open file glass.arff
   (or leave it open from the last lesson)

❖ Check the available classifiers

❖ Choose the J48 decision tree learner (trees>J48)

❖ Run it

❖ Examine the output

❖ Look at the correctly classified instances
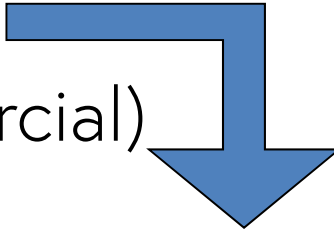   … and the confusion matrix

Building a classifier

Investigate J48

- ❖ Open the configuration panel
- ❖ Check the *More* information
- ❖ Examine the options
- ❖ Use an unpruned tree
- ❖ Look at leaf sizes
- ❖ Set minNumObj to 15 to avoid small leaves
- ❖ Visualize tree using right-click menu

Building a classifier

From C4.5 to J48

- ❖ ID3 (1979)
- ❖ C4.5 (1993)
- ❖ C4.8 (1996?)
- ❖ C5.0 (commercial)

J48

Building a classifier

- ❖ Classifiers in Weka
- ❖ Classifying the *glass* dataset
- ❖ Interpreting J48 output
- ❖ J48 configuration panel
- ❖ … option: pruned vs unpruned trees
- ❖ … option: avoid small leaves
- ❖ J48 ~ C4.5