Load the PIMA diabetes data set (diabetes on the hints web page). It contains 768 examples. Use rules.OneR with the default settings. Test on the training set. Now do a 10-fold cross-validation. What do you notice about accuracy. Why do you think this is true?

Now build a J48 decision tree. Use unpruned on the training set. Now use pruned but change C (the certainty factor) to 0.05. Record the accuracy and the size of the tree. Next try a 10-fold cross validation using the heavily pruned and default pruning C = 0.25. Provide an analysis of accuracy and anything else you think is interesting.

Compare the decision tree results to rules.OneR.

Note: Weka runs on netcluster.cse.usf.edu or sunray.cse.usf.edu.