

Assignment 5

Decision Trees and WEKA

CS482/682: Artificial Intelligence

Due: December 9, 2013

Tools:

- 1) WEKA: <http://www.cs.waikato.ac.nz/ml/weka/downloading.html>
- 2) UCI ML Repository: <http://archive.ics.uci.edu/ml/datasets.html>

Objectives:

- 1) Get the “Abalone” dataset from the UCI ML Repo
- 2) Use a decision tree(C4.5/J48) learner in WEKA to predict the number of rings Abalone will have
- 3) (Graduate Students Only): Repeat the assignment using another dataset. Sushil recommends the “Breast Cancer” dataset, but you’re welcome to use any dataset you want

Write up the following to turn in:

- 1) Describe your process and methods. What parameters (testing/training size, classification target, subtree raising etc.) did you try using with the decision tree? What, if anything, did you do to preprocess the dataset? Why did you select those parameters/preprocessing?
- 2) What were your results? (Overall accuracy, accuracy of each leaf/branch, false positives, false negatives, etc. on the training and testing sets). Show what decision trees you found.
- 3) What do the results tell us? Why are the results (in)accurate? What class were some samples misclassified as, and why? Why did changing parameter(s) improve/degrade accuracy?