Robert Herriott
Undergraduate
Machine Learning
Spring 2014
Decision Tree Classifier

Running this one is the same as the others, just double click the jar and follow the instructions. There are no special packages or anything you need.

I decided to just follow the algorithm to the tee, I even commented in the pseudo-code above each line of actual code. The methods I wrote were:

ID3: This follows the outlined algorithm exactly.
AttributeIndex: This finds the index of an attribute in a Tumor's list of attributes based on an supplied name, I needed this since I decided to make my attributes objects.
Classify: This takes a tumor as an argument and just traverses the tree to classify it.
Homogenous: This tells you if a set of tumors is all malignant or all benign.
MostCommon: This gets a majority vote for benign or malignant for a set of tumors.
Entropy: Obviously calculates entropy for a set of tumors.
Information Gain: Uses Entropy to find the amount of information gained by splitting based on a particular attribute.
GreatestGain: This takes a set of tumors and a set of attributes and tells you which attribute causes the greatest information gain.
Subset: Takes a set of tumors, an attribute name, and an attribute value and returns the subset of tumors that have that attribute with that value.

Results: My classifier was able to correctly classify 66 out of the 70 tumors, or about 94%.

Root and next level questions:
My root asked the question of Size, on the next level, splitting by size produces homogenous groups for all size values except 1 and 2, with tumors of size 1 then asking the question of Nuclei, and tumors of size 2 asking the question of Epithelial.

I was able to classify all the tumors, I just followed the algorithm exactly, creating nodes for all possible values even if I didn't actually encounter them, and if the sets of tumors for those answers were empty, I just labeled them with a majority vote of the previous level.

I didn't have any unresolvable issues.

Read CSV

fil   out

Numerical to ...

exa   exa
       ori

Validation

tra   mod
       tra
%   ave
       ave