In this assignment, you are to implement the ID3 decision tree learning algorithm. We have prepared a basic [ToolKit](http://dml.cs.byu.edu/~cgc/docs/mldm_tools/Assignments/ToolKit.html) for you that you may find useful. Of course, you are welcome to implement everything from scratch, but remember that some of the functionality you will be implementing here will also be useful when implementing Backpropagation (see below). In either case, you should look at the ToolKit description so you may also become familiar with the [ARFF file format](http://dml.cs.byu.edu/~cgc/docs/mldm_tools/Assignments/arff.html), as well as [cross-validation](http://dml.cs.byu.edu/~cgc/docs/mldm_tools/Assignments/accuracy.html). We will ignore reduced error pruning here.

In order to get full credit, you must hand-in a detailed report of your work, carefully covering all of the following:

* Correctly implement the ID3 decision learning tree algorithm. (Note: you may wish to use a simple data set, like [Lenses](http://dml.cs.byu.edu/~cgc/docs/mldm_tools/Assignments/Datasets/lenses.arff), that you can check by hand, to test your algorithm and make sure that it is working correctly; you should be able to get about 68% predictive accuracy on lenses with cross-validation). Your implementation must support the following.
  + An option for choosing the splitting criterion: information gain or accuracy.
  + Some mechanism to handle continuous-valued attributes.
  + Some mechanism to handle unknown (or missing) attribute values.
* Use your ID3 algorithm on the [Iris](http://dml.cs.byu.edu/~cgc/docs/mldm_tools/Assignments/Datasets/iris.arff) problem.
  + Induce a decision tree using the entire dataset with information gain as the splitting criterion. Give a visual representation of the tree.
  + Induce a decision tree using the entire dataset with accuracy as the splitting criterion. Give a visual representation of the tree. Compare this tree with the one obtained with information gain as the splitting criterion.
  + Evaluate predictive accuracy using 10-fold cross-validation for information gain and accuracy. Compare the results.
* Repeat the experiment with the [Voting](http://dml.cs.byu.edu/~cgc/docs/mldm_tools/Assignments/Datasets/voting.arff) problem.
  + Describe and justify the method you used to handle missing values.
* Extend your algorithm so that, when accuracy is the splitting criterion, it may use up to 2 conditions in the tests at each node (e.g., attrX = Vx and attrY = Vy). You may choose to make that an user-specified option.
  + Induce a decision tree using the entire dataset with this extended algorithm for both the [Iris](http://dml.cs.byu.edu/~cgc/docs/mldm_tools/Assignments/Datasets/iris.arff) problem and the [Voting](http://dml.cs.byu.edu/~cgc/docs/mldm_tools/Assignments/Datasets/voting.arff) problem. Give a visual representation of the trees and compare them with those obtained above.
  + Explain why it may be necessary to thus extend the decision tree learning algorithm when using accuracy as the splitting criterion (and why the extension is of little value when information gain is the splitting criterion).