

CS341 Artificial Intelligence – Spring 2015

Program 3 – Bagging

50 points

Due: Apr. 17

For this project you are to write code to build a collection of decision trees using the bagging technique. The problem to be learned is the same as in the decision tree lab we worked on in class: whether or not Tommy will play on the Wii or not. I will provide the following code for you:

1. `p3Gen.cpp` – code which will generate sets of examples. You will use this to generate both training sets and test sets.
2. `example.cpp` – code which defines an `example` class as well as predefined constants appropriate for this project: the number of attributes (6), attribute names (“AdultAtHome”, “GameInterest”, etc), and the number of possible values for each attribute (2 for “AdultAtHome”, 3 for “GameInterest”, etc).

Your code should do the following:

1. prompt the user for a training set file name and the number of hypotheses (i.e., decision trees) to generate.
2. read in the training set and generate the specified number of decision trees using bagging and the information-theory based methods described in class.
3. prompt the user for a test set file name
4. use a simple majority vote system on each of the test cases and report on the percentage of correct classifications.

You should hand in the following

1. Your code, properly commented
2. A report on how effective this method is. You should run your code for a variety of training set sizes (no larger than 40 though) and number of hypotheses (use only odd numbers of hypotheses so that you’ll always have a winner when voting). Your report should comment on your results.