

03-60-574

## Machine Learning

### Assignment #2

Deadline: Tuesday October 31<sup>th</sup> in classroom

Dr. Alioune Ngom

October 13, 2006

1. From the "UCI Knowledge Discovery in Databases Archive" at <http://kdd.ics.uci.edu/>, click on "Machine Learning Repository" then on "Summary page" links to access all the datasets in the repository. Download the "Adult Database" then format the data and test sets as required for WEKA (attributes description is located in the file "adult.names"). Remember to download WEKA before.
2. In WEKA Explorer, evaluate the test set using "One Rule" and "Decision Trees".
  - (a) Show rule and tree learned (at least part of it). Explain their meaning.
  - (b) Compare performance and confusion matrices of both algorithms
3. In WEKA Experimenter, compare the performance of the previous algorithms using 10-fold cross-validation. In your report, make reference to the statistical significance of the results.
4. Implement the Perceptron and Adaline learning algorithm and compare their performance and generalization ability on any simple two-class data set from the UCI repository or from your own (you may define your own separable and/or non-separable two-class functions to learned by your neurons). Experiments with different values of "learning rates" and different sizes of training sets, . . . , etc.
5. Bonus question: Whatever ideas you investigate about DT's and NN's (in particular, bacpropagation NN's) on any data set will be great for you.