

CS4320/5314 – Artificial Intelligence
Spring 2009
Lab # 4
Machine Learning – An Introduction to WEKA
Due electronically before 11:59 p.m., Monday April 6, 2009

The purpose of this project is to become familiar with the WEKA machine learning tool. Following are the instructions.

1. Download the WEKA data mining software in java and install it on your computer. You will need to read the tutorial that comes with the software explaining how to use the explorer interface.
2. Download 3 datasets from the UCI Machine Learning Repository (use Google to find it). Try to choose datasets that look interesting, don't just pick simplest ones. For extra credit, add your own dataset related to your research.
3. For each of the datasets evaluate the performance of the following machine learning algorithms: 3-nearest neighbors, J48, Support Vector Machine, and Naive Bayes. Test each of the algorithms with each dataset using a) The complete dataset for training and testing, and b) 10-fold-cross validation. (This means you will perform a total of 24 experiments).
4. Write a report describing the experiments you performed. Make sure to describe the algorithms used, the test options and parameters used. Include conclusions and insights (which algorithm you think is the best/worst and why, among other things). Do not copy/paste what WEKA gives as output of the experiments, although you can use the confusion matrix and the performance measures in your report.
5. The grade of this project will be based on the originality of the experiments, the analysis of results, and the overall quality of the report (grammar, references, soundness and flow).