Machine Learning Design Methodology

Worth: [20 marks]
Due Date: 10pm Friday, October 19th 2007

Using the Weka system, run the following four algorithms on the following 5 datasets. The algorithms used should be unpruned decision trees, pruned decision trees, naïve bayes and instance-based learning. The five datasets are credit-g, heart-c, hepatitis, vowel, and zoo.

Run your experiments using what you learned in the "experimental design and evaluating hypothesis" lecture notes comparing multiple algorithms on a dataset. Make an argument showing that your results are valid based on your experimental methodology. This would include items such as: training versus testing splits, number of runs, statistics taken, etc.

The weak system and all the datasets can be found at www.cs.auckland.ac.nz/~pat/weka/
Extra documentation can be found at www.cs.waikato.ac.nz/ml/weka

You should hand in a report presenting your experiments. Include what parameters you used for each algorithm and briefly describe the algorithms, the datasets, and parameters. In describing the experimental results explain why you think certain algorithms did better or worse then others. The report must be in MS Word, Postscript, pdf or plain text file.

Marking

Marking is based on the report which should describe the machine learning algorithms, datasets, parameters used, experimental methodology, and results.