The work done on this assignment must be your own work. Think carefully about any problems you come across, and try to solve them yourself before you ask anyone else for help. Under no circumstances should you work together with another student on Java code used in assignments.

Assessment

Due: Monday 26th September 2005 9.00 am
Worth: 10% of total marks for 367

Aim of the assignment

This assignment will give you practical experience implementing a machine learning system using weka for java.

Part 1: numerical data   (40 marks)

Using the dataset ***breastw.arff*** construct a weka-based program to classify instances of possible breast cancer. You program should take as command line arguments a single string containing the information to be classified, in the same order as the information is present in the arff file.
Example: `java –cp weka.jar breastw "5,1,1,1,2,1,3,1,1"`
Ideally, you should add the newly classified instance to the training set. (do not worry about offline storage of the classifier).
Output is to the console, and is just the classification of the data.
You must use the weka class `weka.classifiers.lazy.IBk`

Question 2 mixed nominal and numerical data   (40 marks)

Implement a C45 tree for the dataset ***hepatitis.arff*** using the weka classifiers in package `weka.classifiers.trees.j48`. You also must use the C45 model selection.
Again, input is a single command line string containing information in the same form as the arff file
Example: `java –cp weka.jar hepatitis`
`"30,male,no,no,no,no,no,no,no,no,no,no,no,1,85,18,4,?,no"`
Output is to the console, and is the classification of the data by the tree.
You must use pruning.

Question 3 Accuracy   (20 marks)

Test your algorithms (either using weka or by hand), choosing an experimental method, to compute the accuracy of your program.
Make sure you discuss the experimental method and the results.
Full marks are only given in cases where you perform some kind of anaylsis of each algorithms' peformance.
The report need not be large: a paragraph should suffice.

Submitting Files

Submit all your files in a single ZIP file via the computer science assignment dropbox https://adb.ec.auckland.ac.nz/adb/