CS 550 -- Machine Learning
Homework #1

Due: 9:30 (class time), March 25, 2014

1. Implement a decision tree classifier that uses postpruning approach. In your implementation, you will use your selected splitting criterion and postpruning technique. You should give the details of your selection. You should also list all parameters that are used by your implementation.

   In this question, you will conduct your experiments on the Iris data set, which is available on the course web page (*iris_train.txt* and *iris_test.txt*). The Iris data set contains 3 classes, each of which corresponds to a type of iris plant (*Iris Setosa, Iris Versicolour, and Iris Virginica*). Each instance is represented with 4 attributes (sepal length, sepal width, petal length, and petal width). Using your implementation, you are asked to
   a. Select the parameter values using 3-fold cross validation on the training samples,
   b. Get training and test set accuracies, and
   c. Make sensitivity analysis on the test samples (i.e., explore how test accuracies change as a function of your parameters).

   In this question, you may use any programming language that you would like. You are asked to report the details of your implementation and experimental findings. In this question, you SHOULD NOT give the screen shots or outputs of your program but you should summarize what you have found at the end of your runs.

2. This time, use a machine learning toolbox (e.g., PRTools, Weka) for a decision tree classifier. In this question, you will explore decision tree classifiers with different options, which are provided by your selected toolbox. Conduct your experiments for the Iris data set (*iris_train.txt* and *iris_test.txt*), give a list of what you try, and report your observations including the training and test set accuracies.

   Additionally, you should select another dataset from the UCI Machine Learning Repository and report your observations on this dataset as well. Your selected dataset should contain both discrete and continuous features. If there are no separate training and test sets for your selection, use 10-fold cross-validation. You should also report the details of the dataset you select.

   Again DO NOT directly give the screen shots or outputs of those programs but give a summary of what you have obtained at the end of your runs.

3. Why is the following statement true?

   Occam's razor: Prefer the simplest hypothesis that fits the data

   Make a brief survey using different sources (books, papers, internet, etc.) and write down why this is true *in your own words* (min 150, max 300 words excluding references). Give the list of all your references.

4. Compare the candidate elimination and the decision tree algorithms *in your own words* (min 150, max 300 words excluding references).

5. Derive delta rules (ΔW_{ji} and ΔV_{kj}) of multilayer perceptrons for multi-class classification (where you have more than two classes). Derive these rules using the following notations and selections. Show all your work.

    a. Notation: x_i (inputs), y_j (hidden units), o_k (outputs), W_{ji} (input-to-hidden weights), V_{kj} (hidden-to-output-weights), d (number of input units), h (number of hidden units), c (number of output units [classes]),
    b. Derive rules for the stochastic algorithm,
    c. Use the squared error function,
    d. Use sigmoid as the activation function of the hidden units, and
    e. Use softmax as the threshold function of the output units.