Machine Learning - Assignment 3

Due Tuesday, October 15, 2002

1. [10 points] Mitchell, pg. 124 Problem 4.2.

2. [65 points] This question asks you to experiment with neural networks for the task of face recognition, using the code provided with our textbook. The code and documentation for using it are available from:

   `http://www-2.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/faces.html`

   (a) Get the training set and test set data *for the reduced-size images*

   (b) Do tasks described in Part I, items 2-4 in the documentation posted on the web page (which is in the form of a homework assignment).

   (c) Solve problem 4.10, pg. 125 from the book.

   (d) Change the backpropagation algorithm (file backprop.c) to implement the new update rule you derived.

   (e) Repeat the same experiment as before, using the weight decay algorithm (from item c). Plot a graph comparing the error (or accuracy) of the two algorithms on the training and test error, as learning progresses. Also, choose a couple of hidden units and plot their weights under the two update rules. Write a little report summarizing your results.

3. [25 points]

   One perceived drawback of neural networks is the difficulty of interpreting the weights. One way of producing an explanation of the network is to consider its behavior in the context of a particular example. Consider a multi-layer neural network N with k hidden units and one output unit. Suppose we have a single example $\mathbf{x} = (x_1 \ldots x_n)$ that we would like to explain. One way to produce an explanation is to study how small changes to each one of the input features $x_1, \ldots x_n$ will affect the output of the network. To do this, we want to compute:

   $$\frac{\partial N}{\partial x_i}, \forall i = 1, \ldots n$$

Show that a simple extension of the backpropagation algorithm can be used to do this. Hint: the derivation is almost the same as the derivation fro adjusting the weights of hidden units in the backpropagation algorithm.

Explain how this idea can be used to interpret the network.