Homework #4 – Neural Networks

CAP 5638, Pattern Recognition, Fall, 2015
Department of Computer Science, Florida State University

Due: Monday, November 16, 2015 (No late submission will be accepted as the solution will be available during the class on the due date for the midterm exam starting on November 18, 2015)
Submission: Hardcopy (including programs) is required and is due at the beginning of the class on the due date.

Points: 80 points

Problem 1 (10 points) Problem 1 of Chapter 6 of the textbook

1. Show that if the transfer function of the hidden units is linear, a three-layer network is equivalent to a two-layer one. Explain why, therefore, that a three-layer network with linear hidden units cannot solve a non-linearly separable problem such as XOR or n-bit parity.

Problem 2 (10 points) Problem 3 of Chapter 6 of the textbook

3. Consider an $d - n_H - c$ network trained with n patterns for m_e epochs.

(a) What is the space complexity in this problem? (Consider both the storage of network parameters as well as the storage of patterns, but not the program itself.)
(b) Suppose the network is trained in stochastic mode. What is the time complexity? Since this is dominated by the number of multiply-accumulations, use this as a measure of the time complexity.

(c) Suppose the network is trained in batch mode. What is the time complexity?

Problem 3 (15 points) Problem 11 of Chapter 6 of the textbook

11. Generalize the backpropagation to four layers, and individual (smooth, differentiable) transfer functions at each unit. In particular, let x_i, y_j, v_l and z_k denote the activations on units in successive layers of a four-layer fully connected network, trained with target values t_k. Let f_{1i} be the transfer function of unit i in the first layer, f_{2j} in the second layer, and so on. Write a program, with greater detail than that of Algorithm 1, showing the calculation of sensitivities, weight update, etc. for the general four-layer network.

Problem 4 (15 points) Problem 29 of Chapter 6 of the textbook

Consider a $d - n_H - c$ three-layer neural network whose input units are linear and output units are sigmoidal but each hidden unit implements a particular polynomial function, trained on a sum-square-error criterion. Specifically, let the output of hidden unit j be given by

$$o_j = w_{ji}x_i + w_{jm}x_m + q_jx_ix_m$$

For two prespecified inputs, i, and m≠i.

(a) Write gradient descent learning rule for the input-to-hidden weights and scalar parameters q_j.

(b) Does the learning rule for the hidden-to-output unit weights differ from that in the standard three-layer network described in the text?
(c) What might be some of the strengths and the weaknesses of such a network and its learning rule?

Problem 5 (15 points) Computer Exercise 1 of Chapter 6 of the textbook

1. Consider a 2-2-1 network with bias, where the transfer function at the hidden units and the output unit is a sigmoid $y_j = a \tanh[b\, net_j]$ for $a = 1.716$ and $b = 2/3$.

(a) Suppose the matrices describing the input-to-hidden weights (w_{ji} for $j = 1, 2$ and $i = 0, 1, 2$) and the hidden-to-output weights (w_{kj} for $k = 1$ and $j = 0, 1, 2$) are, respectively,

$$\begin{pmatrix} 0.5 & -0.5 \\ 0.3 & -0.4 \\ -0.1 & 1.0 \end{pmatrix} \text{ and } \begin{pmatrix} 1.0 \\ -2.0 \\ 0.5 \end{pmatrix}.$$

The network is to be used to classify patterns into one of two categories, based on the sign of the output unit signal. Shade a two-dimensional input space $x_1 - x_2$ ($-5 \leq x_1, x_2 \leq +5$) black or white according to the category given by the network. Repeat with

(b) Repeat part (a) but with the following weight matrices:

$$\begin{pmatrix} -1.0 & 1.0 \\ -0.5 & 1.5 \\ 1.5 & -0.5 \end{pmatrix} \text{ and } \begin{pmatrix} 0.5 \\ -1.0 \\ 1.0 \end{pmatrix}.$$

Problem 6 (15 points) Computer Exercise 5 of Chapter 6 of the textbook

5. Write a basic backpropagation program for a 3-3-1 network with bias to solve the three-bit parity problem, i.e., return a +1 if the number of input units that are high is even, and -1 if odd.

(a) Show the input-to-hidden weights and analyze the function of each hidden unit.

(b) Retrain several times from a new random point until you get a local (but not global) minimum. Analyze the function of the hidden units now.

(c) How many patterns are properly classified for your local minimum? Explain.

Extra credit problem

Problem 7 (5 points) Problem 21 of Chapter 6 of the textbook, **part (b) only**

21. Consider a three-layer network for classification with output units employing softmax (Eq. 30), trained with $0 - 1$ signals.

(b) Repeat for the criterion function is cross-entropy, i.e.,

$$J_{ce}(\mathbf{w}) = \sum_{k=1}^{c} t_k \ln \frac{t_k}{z_k}.$$