CSCE 496/896: Machine Learning	Fall Semester, 1999
Homework Assignment 2	
October 22, 1999	*Due Date: Monday, November 8*

This is due at 5:00 p.m. on the due date. **You may work with a partner on this homework. If you do, then I will grade more rigorously, especially on the written presentation of your results.**

1. (10 pts) Do problem 4.3 on p. 124 and *justify* your solution.

2. (10 pts) Do problem 5.3 on p. 152.

3. Do *either one* of the following. Note that you are expected to submit a clear, concise writeup of your results, no matter which problem you do. Your presentation will be part of your grade. If you do part (b) you should hand in your code using the handin program as with homework 1. If you do part (a) then you should hand in your code only if you made significant extensions to it.

   Each problem has components that are required only for students registered for CSCE 896. CSCE 496 students who do these will receive some extra credit.

   (a) (50 pts for 496, 65 pts for 896) Problem 4.11 from the text (p. 125). This involves applying a neural network to the problem of face recognition. The instructions, code, and data are available at `http://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/faces.html`. You need not download all the data (the datasets are very large); instead you may find it in `/home/classes/cse496/faces` and in `/home/classes/cse496/faces_4`. Everything else (code, instructions) you must get off of the web page. The instructions and code documentation are in `http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-8/faceimages/docs/hw97.ps`. CSCE 496 students must do Part I (Section 2.1) and may do Part II (Section 2.2) for extra credit. CSCE 896 students must do Part I and at least three suggested experiments from Part II.

   When you report your results, use error plots whenever possible, and compute 95% confidence intervals for the error results (you may incorporate the intervals into the plots if you wish).

   (b) (50 pts for 496, 65 pts for 896) Implement the Perceptron algorithm or Winnow and implement the bagging or boosting algorithm with it. Then generate a test set T of 500 examples uniformly at random from $[0, 100] \times [0, 100] \subset \Re^2$, where an example $(x_1, x_2) \in T$ is positive iff either (a) $x_2 \geq x_1 + 30$ or (b) both $x_2 \leq 2x_1 - 100$ and $x_2 \geq -3x_1 + 300$. I.e. (a) is a single halfspace, (b) is the intersection of two halfspaces, and the target concept is the union of (a) and (b). Now choose three values of m (each < 300), which will be the sizes of your training sets. For each value of m run 10 experiments, where each experiment involves generating a random training set of size m (the same way T was generated) and running 10 rounds of boosting or bagging on your classifier. At each round, average the training error of the 10 experiments and average the test error of the 10 experiments. Then plot your two error curves versus the number of bagging/boosting rounds. Do this for the three values of m, so you'll get three plots, each with two curves. Also, for each error value plotted, compute its 95% confidence interval (you may incorporate the interval into the plot if you wish). In your report, discuss the results, including the following questions. When did training error go to 0? Did overfitting occur after this point? When should you have stopped?

   CSCE 496 students need only choose one of the Perceptron or Winnow, and one of bagging or boosting. CSCE 896 students should run on both the Perceptron and Winnow and use both bagging and boosting. Thus these students will have more plots or more curves per plot.

   Hint: Those implementing Winnow should consider the extension that allows negative weights.

4. (5 pts) State how many hours you spent on each problem of this homework assignment.