CSCE 496/896: Machine Learning **Fall Semester, 1999**

Homework Assignment 3

November 10, 1999 *Due Date: Monday, November 29*

This is due at 5:00 p.m. on the due date.

1. (25 pts) For this problem you will implement a k-nearest neighbor classifier for at least three significantly different values of k.

   The data for this problem is in `/home/classes/cse496/hw3/`, in four files: `pos.train`, `pos.test`, `neg.train`, and `neg.test`. Obviously `pos.*` contains positive examples and `neg.*` contains the negative ones, with training examples in `*.train` and testing examples in `*.test`.

   You are to run 10 experiments, where experiment 1 uses 100 of the 1000 training examples (drawn uniformly at random *without* replacement from the union of `pos.train` and `neg.train`), experiment 2 uses 200 of the 1000 training examples, etc. For each experiment. compute the average classification error over all the test examples as well as the average time to classify a single test example. Plot these results on two graphs (two graphs per value of k) and include 95% confidence intervals for your measurements. What conclusions can you draw? Are either of the attributes irrelevant? When does the increase in classification time outweigh the decrease in classification error? How does varying k affect the results? When drawing your conclusions, you might want to plot the data itself to see what you're training and testing on.

   For 5 extra points, also try distance-weighted nearest neighbor in the above experiments (in addition to the three values of k).

   For 10 extra points, also try locally weighted linear regression in the above experiments (in addition to the three values of k). In order to get binary output, you should threshold the output of $\hat{f}(x)$.

2. (5 pts) Do problem 6.1 on p. 198.

3. (15 pts) Do problem 6.5 on p. 199.

4. (5 pts) State how many hours you spent on each problem of this homework assignment (for CSCE 896 students, this includes the next problem).

   The following problem is only for students registered for CSCE 896. CSCE 496 students who do it will receive extra credit, but the amount will be substantially less than the number of points indicated.

5. (15 pts) Do problem 6.4 on p. 198.