CS/CNS/EE 156 a — Caltech
Learning Systems — Fall 2000

Homework # 4

DUE THURSDAY, NOVEMBER 16, 2000, AT 2:30 PM

Collaboration in the sense of discussions is allowed, but you should write the final solutions alone and understand them fully. Do not read class notes or homework solutions from previous years at any time. Other books and notes can be consulted, but not copied from. You should justify your answers, at least briefly. Definitions and notation follow the lectures.
The handouts and data for the homeworks can be found at:

http://work.caltech.edu/cs156/00/homeworks.htm

1. Nearest-Neighbor Classifiers

Consider the following training set generated from a target function $f : X \rightarrow \{-1, +1\}$ where $X = R^2$

$\mathbf{x}_1 = (1,0), y_1 = -1 \quad \mathbf{x}_2 = (0,1), y_2 = -1 \quad \mathbf{x}_3 = (0,-1), y_3 = -1$

$\mathbf{x}_4 = (-1,0), y_4 = +1 \quad \mathbf{x}_5 = (0,2), y_5 = +1 \quad \mathbf{x}_6 = (0,-2), y_6 = +1$

$\mathbf{x}_7 = (-2,0), y_7 = +1$

(i) Plot the decision boundary of the nearest-neighbor rule.

(ii) If we replace the training points in each class by their sample mean (μ_{-1} and μ_{+1}) and use the nearest-neighbor rule based on these two points only, what will the decision boundary be?

(iii) How many examples in the original training set would be misclassified if the rule in (ii) is used?

2. Radial Basis Functions

Consider the training set of problem **1**. We use the exact interpolation version of radial basis functions with unit-variance Gaussians

$$g(\mathbf{x}) = \sum_{n=1}^{7} w_n e^{-0.5||\mathbf{x}-\mathbf{x}_n||^2}$$

and then classify the point $\mathbf{x}$ as $+1$ if $g(\mathbf{x}) \geq 0$ and as -1 if $g(\mathbf{x}) < 0$.

(i) Find the values of w_n's (numerically).

(ii) Plot the resulting boundary between the $+1$ and the -1 regions in the plane.

3. Experiment with Nearest Neighbors

Write a program to implement the K-Nearest Neighbor algorithm. Use the training and test data from HW# 1 (with target values in $\{-1, 1\}$).

(i) Classify the test points according to the Nearest Neighbor ($K = 1$) classifier based on the training data. Compare your resulting error to your results from HW# 1.

(ii) Compare the results of classifying the test points using the K-Nearest Neighbor classifier based on the training data for $K = 1, 3, 5, 7, 9, 11, 13, 15$.

(iii) Compare the actual out-of-sample errors you obtain in (ii) to their in-sample, leave-one-out, estimates.

3. Experiment with Radial Basis Functions

Obtain the training and test sets from

http://www.work.caltech.edu/cs156/00/hw/hw4/train.dat

and

http://www.work.caltech.edu/cs156/00/hw/hw4/test.dat

The first two columns are the 2-dimensional input and the third column is the (continuous) output. Implement a RBF network that approximates the function. Use Gaussian basis functions, keep σ fixed, and use K-means clustering to find the centers. Find the second layer weights by computing the pseudo-inverse of the Φ matrix. Try using 2,3 and 4 basis functions and σ = 0.5,1.0,1.5,2.0. Explain the results.