ASSIGNMENT 11

LFD is the class textbook

k-NN and RBFs on Digits Data

In this assignment, use the same data you created in Assignment 9 for the digits problem classifying 'Digit 1' from 'not Digit 1'.

So you combined all the data to one data set, normalized the data so that the range of both features is $[-1, 1]$ (there is mild data snooping here but, for simplicity, we will live with it) and selected 300 data points for your training set and the remaining are a test set.

1. **(450) k-NN Rule.**

(a) Use cross validation with your training set to select the optimal value of k for the k-NN rule.

   Give a plot of E_{cv} versus k. What value of k do you choose.

(b) For the value of k that you took, give a plot of the decision boundary. What is the in-sample error. What is the cross validation error.

(c) What is the test error E_{test}?

2. **(450) RBF-network.**

(a) For the RBF-network with Gaussian kernel, set the scale $r = 2/\sqrt{k}$, where k is the number of centers. Use cross validation with your training set to select the optimal number of centers k for the RBF-network.

   Give a plot of E_{cv} versus k. What value of k do you choose.

(b) For the value of k that you took, give a plot of the decision boundary. What is the in-sample error. What is the cross validation error.

(c) What is the test error E_{test}?

3. **(100) Compare Linear, k-NN, RBF-network.**
Compare the final test error from your three attempts to solve this problem:
   (i) Linear model with 8th order polynomial transform and regularization selected by CV.
   (ii) k-NN rule with k selected by CV.
   (iii) RBF-network with number of centers selected by CV.
Make some intelligent comments.