Homework 11
CSCI4100

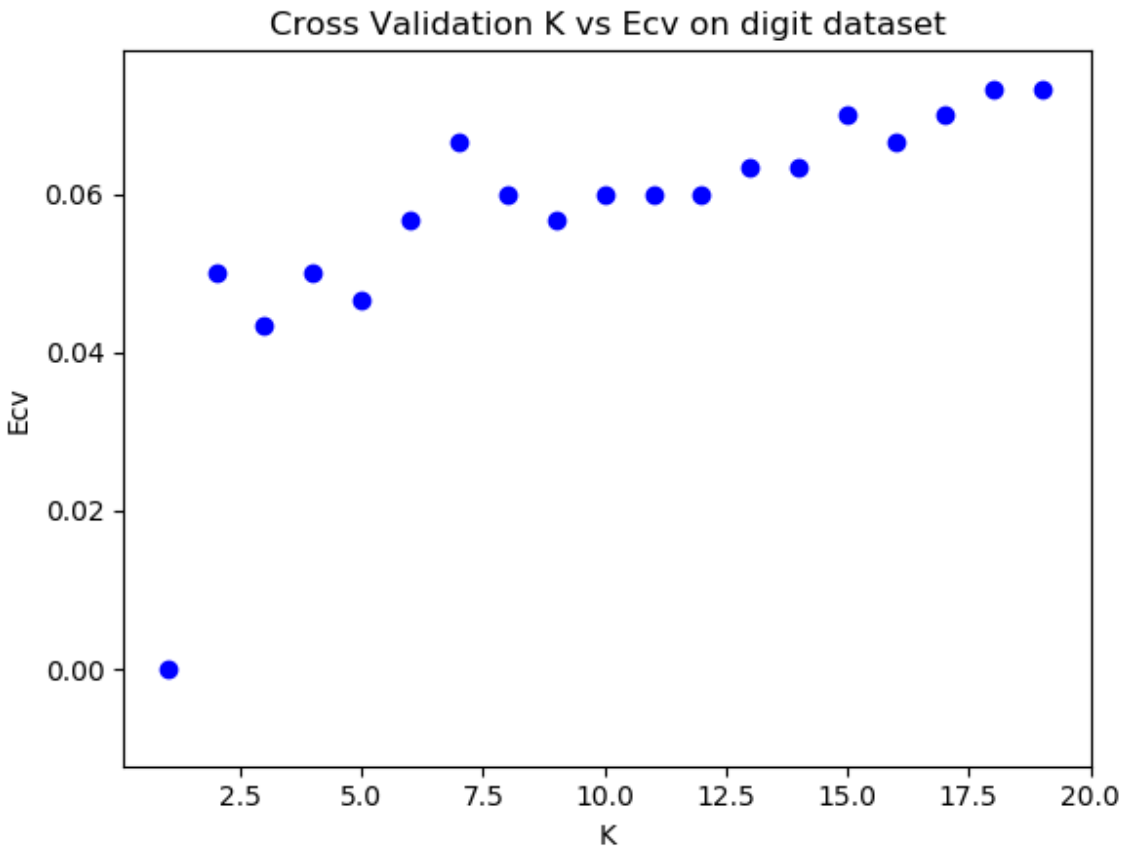
Han Hai
Rin:661534083
haih2@rpi.edu

November 27th 2018

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.

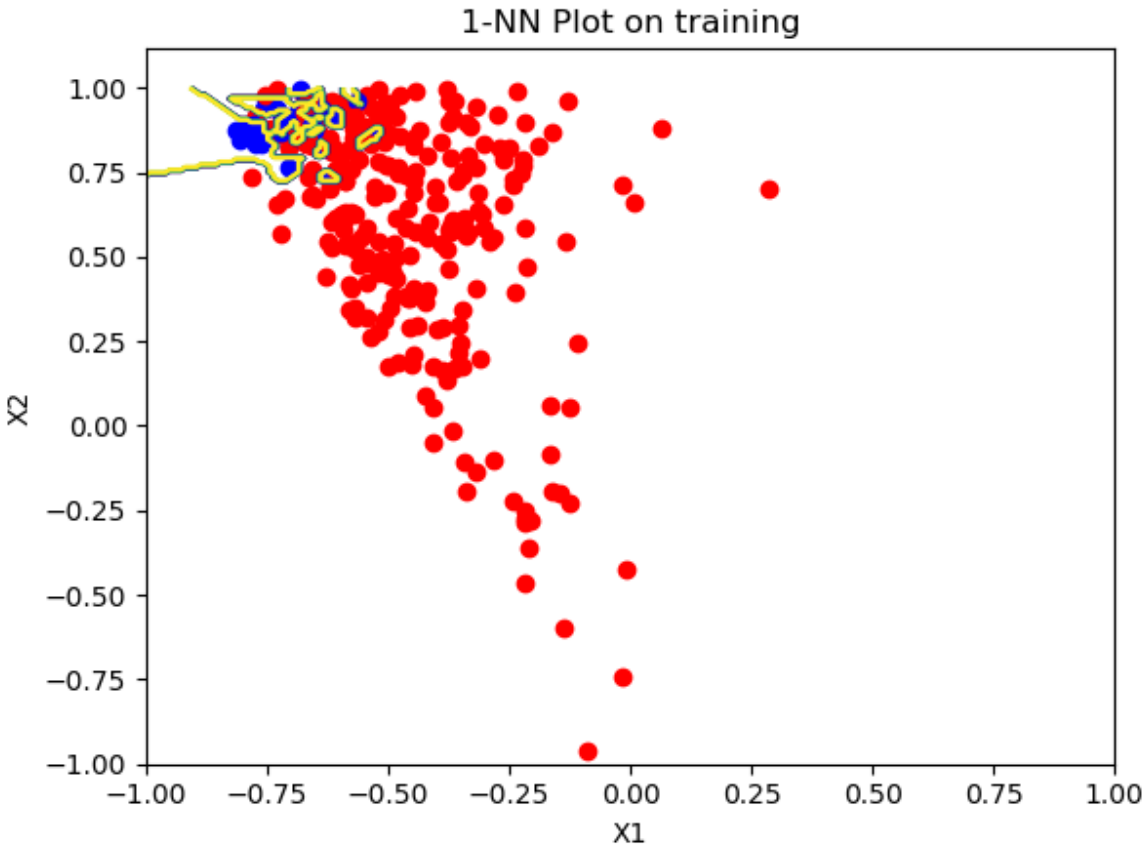
1. **(450) k-NN Rule**

   (a) Use cross validation with your training set to select the optimal value of k for the k-NN rule. Give a plot of Ecv versus k. What value of k do you choose.



According to the graph the best k is 1 with Ecv=0

(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.

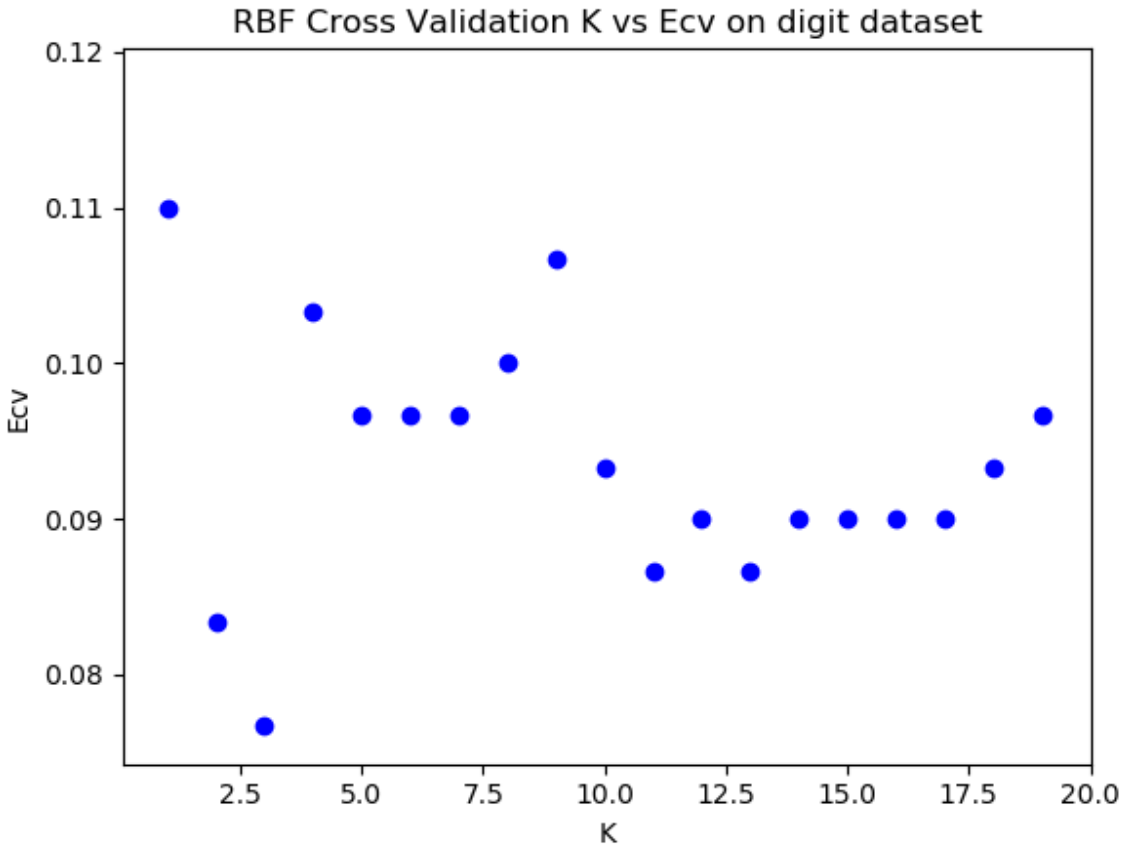


The in-sample error is 0.083 and the Ecv is 0

(c) Etest is 0.094

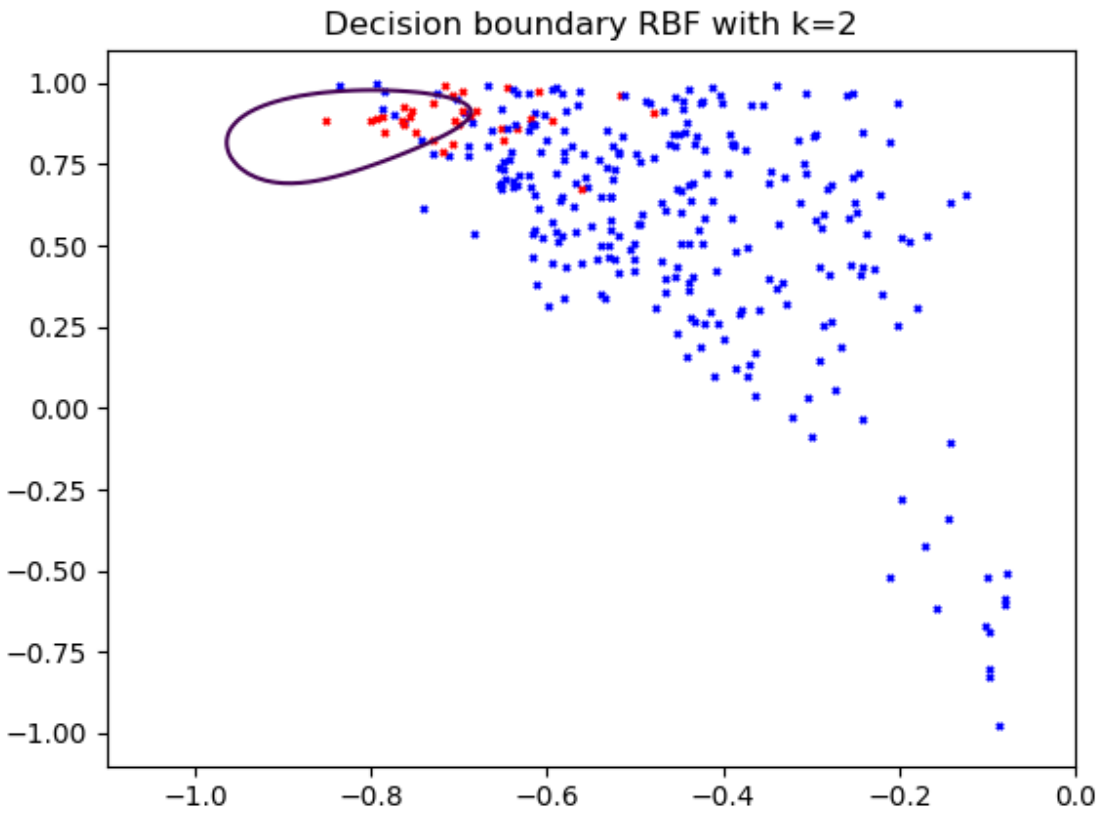
2. **RBF network**

(a)



According to the graph the best k is 3, at 3 Ecv is 0.073

(b)



At k=3, Ein is 0.083,

(c) Etest is 0.091

3. **Compare Linear, k-NN, RBF-network.**
KNN has test error of 0.094
RBF has test error of 0.091
linear model with regularization $\lambda = 1.35$ (from my data in homework 9) has test error 0.083
Apparently, linear model with regularization has the best Etest out of the three models. It is probably because that it's the simplest model.In a general sense, the simplest model often is the best model. KNN is the worst model here, while RBF is slightly better. It is probably because of overfitting, as we can see from the graph, some unnecessary extra boundaries are produced in KNN which probably increase test data errors.