

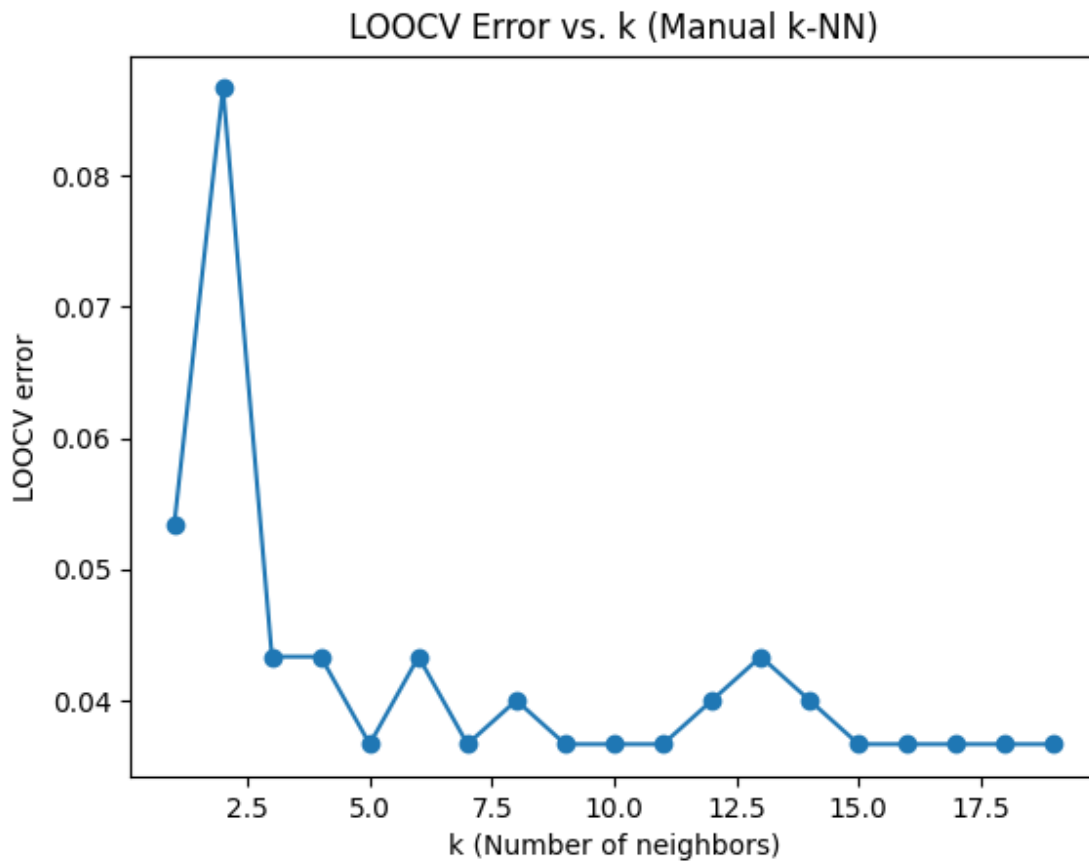
LFD Problem Set 11

John Cohen

November 25, 2024

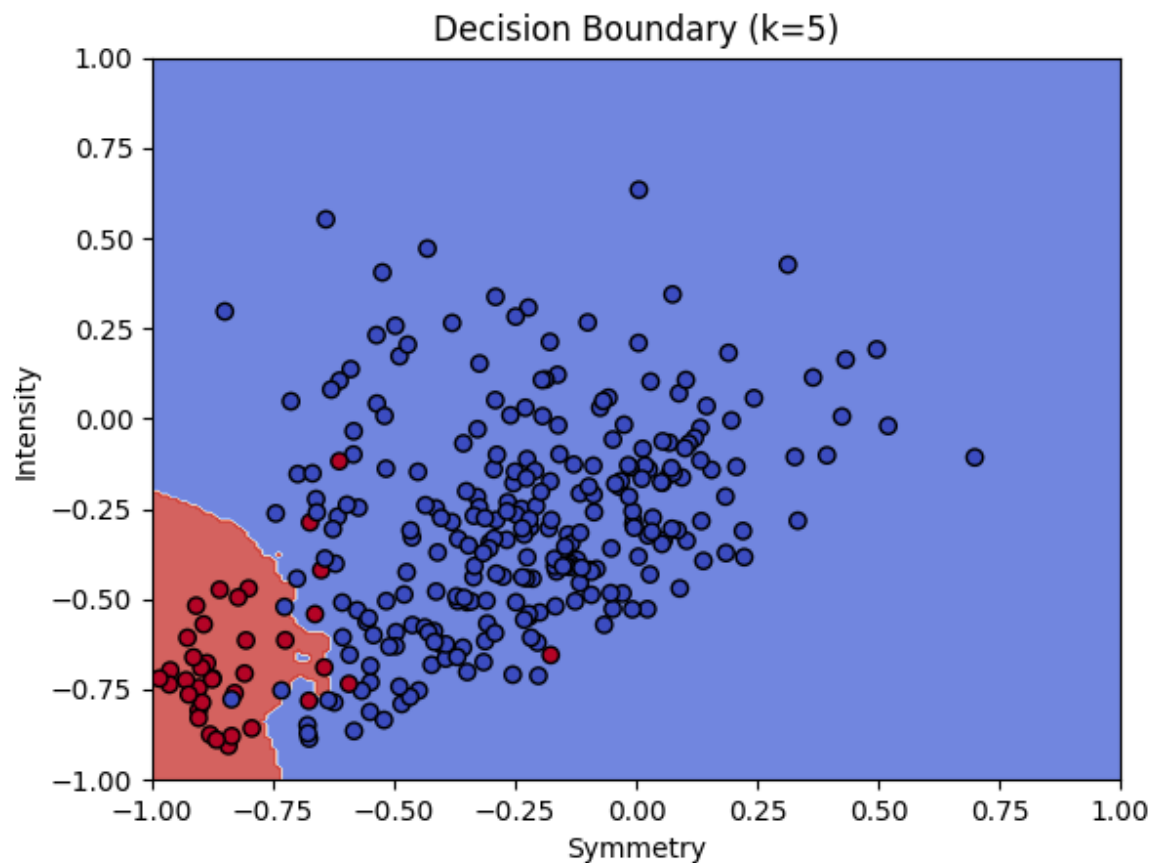
k-NN Rule

(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.



Optimal k : $k^* = 5$

(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.



$$E_{in}(k^*) = 0.0367$$

$$E_{cv}(k^*) = 0.0367$$

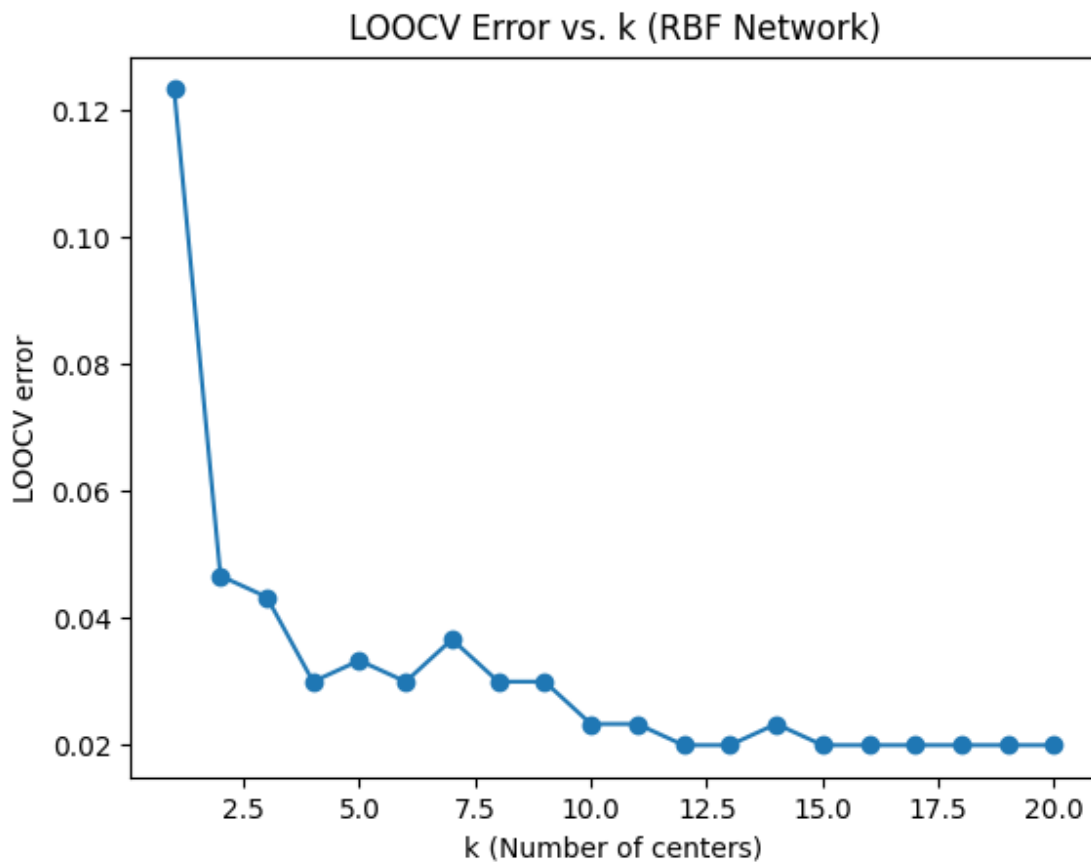
(c) What is the test error E_{cv} ?

$$E_{test}(k^*) = 0.0321$$

RBF-network.

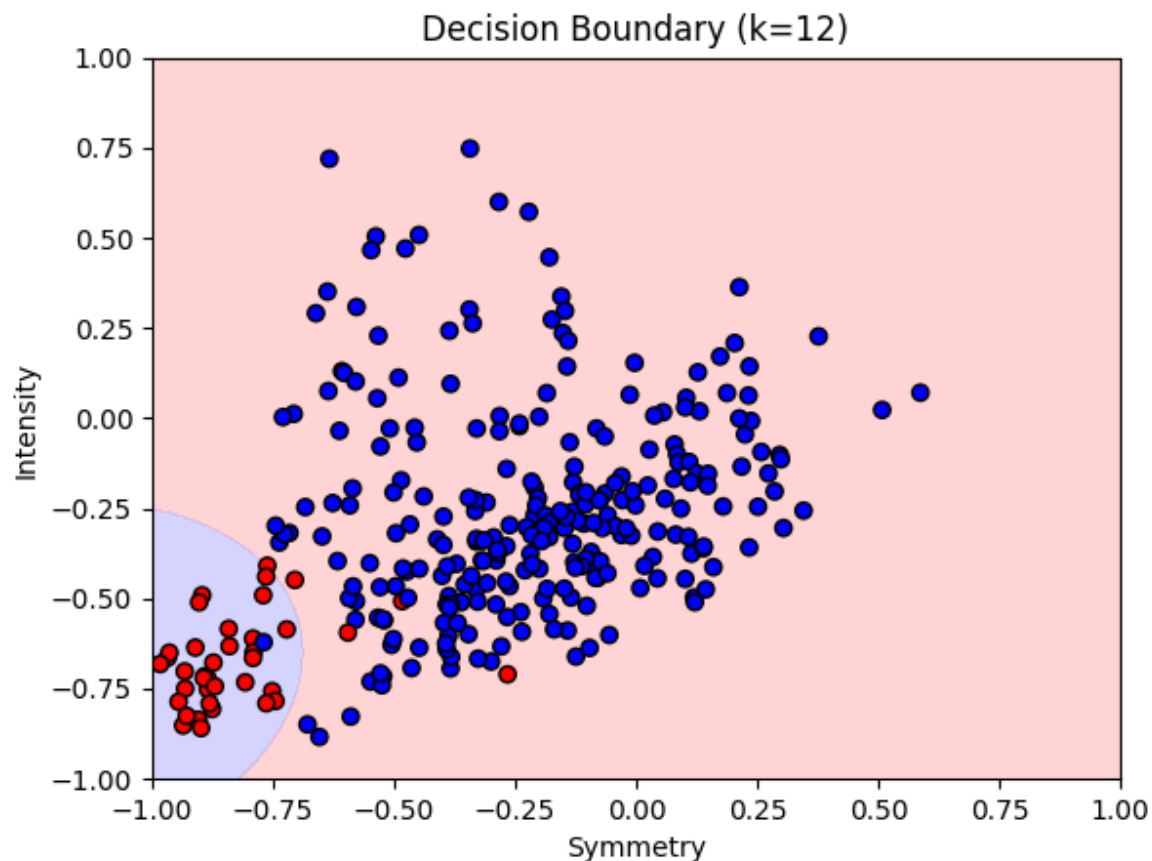
(a) For the RBF-network with Gaussian kernel, set the scale $r = \frac{2}{\sqrt{k}}$, where k is the number of centers. Use cross validation with your training set to select the optimal number of centers

k for the RBF-network. Give a plot of E_{cv} versus k. What value of k do you choose.



Optimal k: $k^* = 12$

(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.



$$E_{in}(k^*) = 0.0200$$

$$E_{cv}(k^*) = 0.0200$$

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

$$E_{test}(k^*) = 0.0335$$

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.

The k-NN method achieved the lowest test error, indicating it effectively captures the local patterns in the data. However, it requires storing all training data and can be computationally intensive for large datasets. The RBF-Network performed slightly worse but offers a balance between accuracy and efficiency by approximating the k-NN while using fewer resources during prediction. The Linear Model had the highest test error, suggesting that even with an 8th order polynomial, it may not capture the complex nonlinear relationships as effectively. Considering both performance and computational efficiency, the RBF-Network may be the preferable choice despite the k-NN's marginally better accuracy.