Project 1 week 2

Opinion on partner's project

Zaoyi Chi

EE/CS 599

Partner: Hannah Caldwell-Meurer

The first figure Hannah chose is about the random forest. In specific, she wants to use the random forest learning method to classify the mixture dataset into two regions based on their labels. The random forest classifier is robust in dealing with the classification problem. She plans to reproduce the Figure 15.11 from the Hastie book. The figure is quite informative with its color-filled regions and boundaries. It also shows the value of the test, train, and Bayes error that could give me general ideas about the performance of this algorithm.

The second figure is from her first presented paper(*Gaussian Process Kernels for Pattern Discovery and Extrapolation*) in class. The idea behind this figure is that she wants to predict and fill the wave gap based on the partial observation generated by a given complex sinc function. In the original figure, it uses five different kernels and methods to fill out the gap. This figure shows the difference between the prediction of those different kernels pretty well; yet, I think it would be better to have one more graph to show what exactly the given function looks like in the gap part, and then we can know which kernel or method performs better in this case. Besides, the figure should also provide a legend to clarify each colored line.

I would suggest reproducing the first choice for the project because the random forest has a wide application realm. The first figure is almost the same as the figure I would like to choose for my SVM project. I think it would be better if we can do this kind of comparison between hers and mine in the end. Moreover, it would be better if the algorithm could be coded from scratch; that could offer a better understanding of this random forest project.