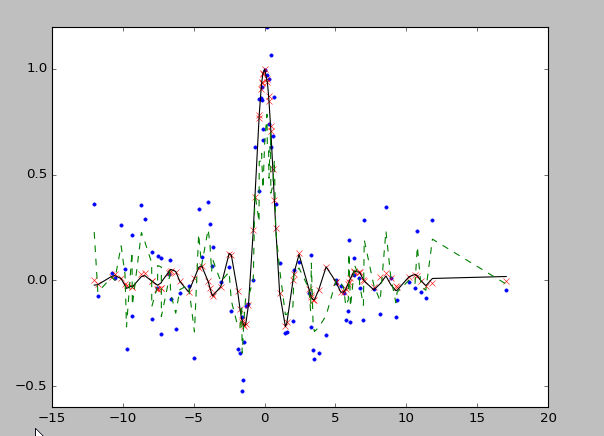
Jack Barry

AI final exam

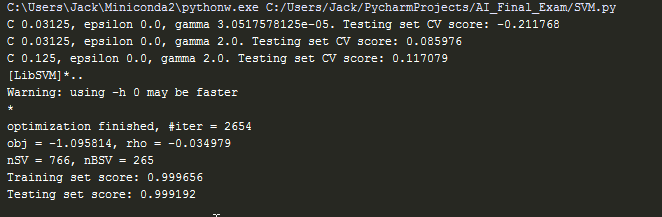
Pablo Rivas

12/16/2016

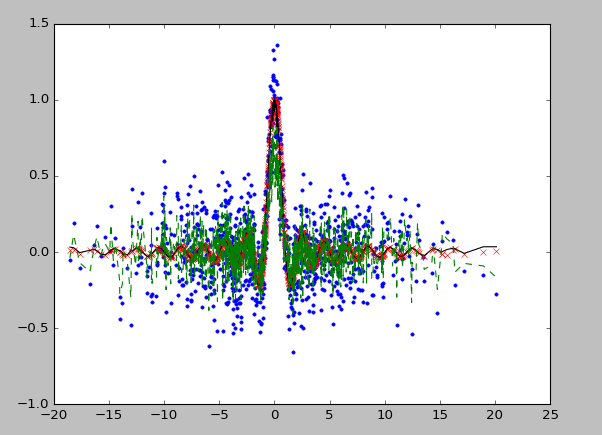
To test my code I to ensure I had the correct modules installed to run the code I did an initial run of 100 data points, yielding the following visualization. Once the code was working I let it run for several hours to find a solution for 1000 data points.



After running for several hours the SVM reached the solution below



A support vector machine is a discriminative classifier formally defined by a separating hyperplane. It provides a hyperplane that best separates data of multiple dimensions. SVM is a supervised learning model. The code outputs C, epsilon, gamma, and a cross validation score. These metrics are key for tuning an SVM model. C is a regularization parameter that controls the tradeoff between low error and having a plane that is skewed towards one dimension of the data more than the other. Our result of C = 0.125 was increased from the iteration before that which had a C of 0.03125. This is a large jump from one iteration relatively speaking but it is still a small C. Our small C shows that we have high bias and low variance. Our epsilon was 0 for every iteration. Gamma is the free parameter of the Gaussian rbf. We ended with a relatively small gamma of 2.0 which means most of our data points are being considered in how they affect the model or in other words the model has a large reach. The final scores for the models testing and training set are very good, showering a strong model for classification. Below is the visualization produced from N = 1000 of question 1.



For question 2 we see a complete change in the parameters as they go from small values to large values which shows how the model must be adapted to the data that is being worked with. The Large C means our model will have high bias and low variance, looking at the visual it makes sense because the data is fuzzy. Epsilon was 2.5. Gamma is 0.25 which is close to question one’s solution. The final scores for this model were not great. The data being so closely smashed together challenges the model to reach accuracy. All in all the final exam showed how important understanding the data you are working with is to conducting and analysis.

