Alexander Palomba

Artificial Intelligence

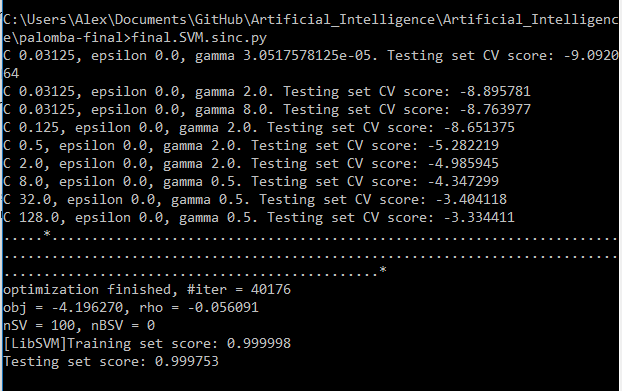
Dr. Rivas

16 December 2016

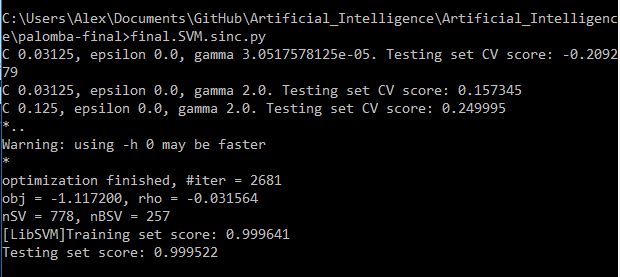
Final Exam

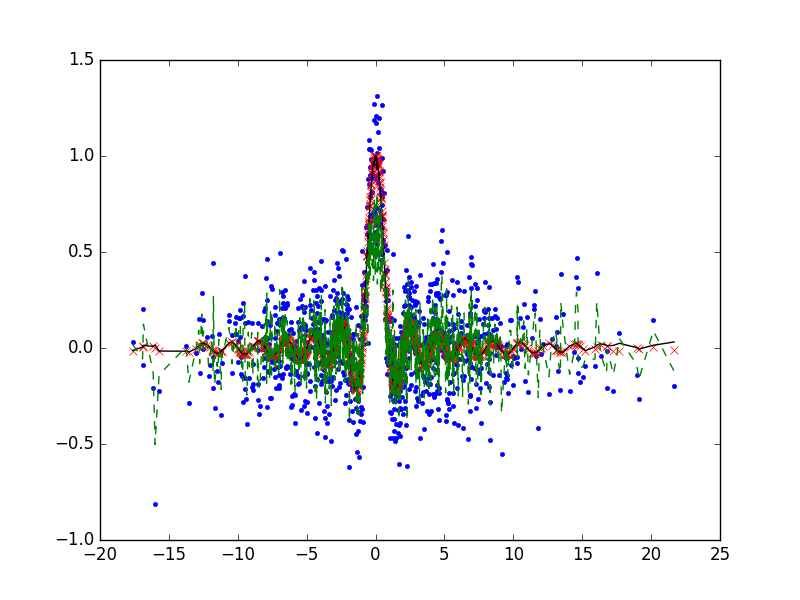
1.

a-b. After successfully downloading the two python files, I ran the sinc function to make sure it was working properly.



c. Ran the program for N = 1000 samples

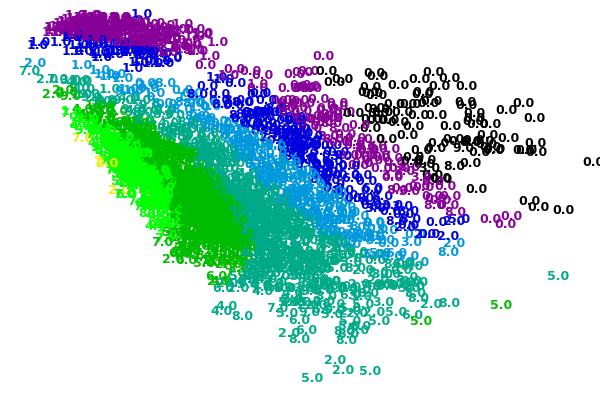


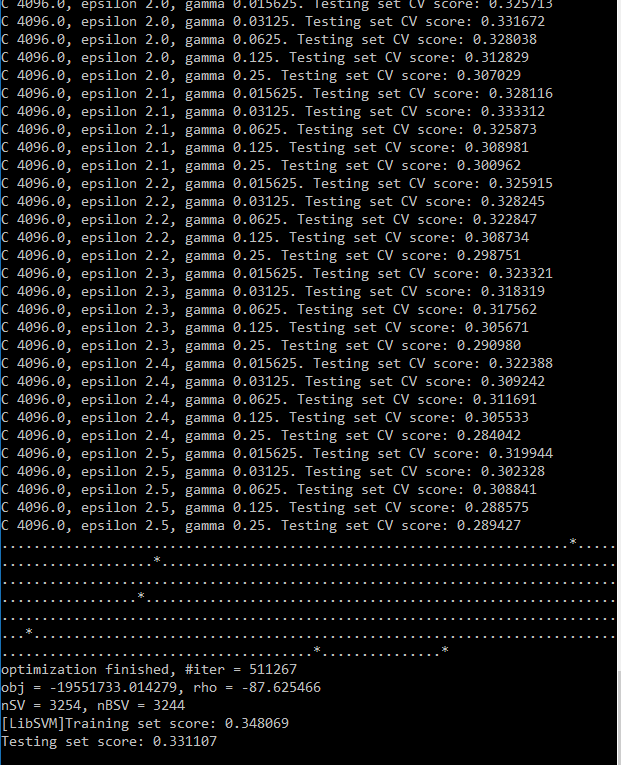


d. The best penalty value for C was comparatively very low, meaning that the output would be smoother but many of the points could be misclassified as a result. When compared to the results of the test in part a/b, the value for gamma is on the larger side, but was still relatively close to the average. This means that the model should be a good representation of the data without the hazard of overfitting. For every iteration thus far, epsilon has stayed at 0, so no proper conclusions can be drawn on a correlation between C and epsilon at this time.

2.

a-b. Saved the files, and let the function run overnight.





c. After 50,000+ iterations, the best penalty that the model could produce was C = 4096, with an epsilon of 2.5, and a gamma of 0.25. The training and testing scores were both right around 0.3, which is pretty low. Based on these results, and what we know about a high C and low gamma (i.e. high probability of overfitting), one could expect a rather high amount of misclassified points in the diagram. Looking at the diagram itself, it appears that there are in fact a lot of points that have been classified in the wrong clusters (e.g. 8s grouped in with the 0s and vice versa).

Resources:

<http://reev.us/cmpt404f16/hw/> - python code and datasets

<http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html> - general information on SVM parameters