ECEN765 Machine Learning with Networks
Liuyi Jin

This project requires us to implement a SVM classifier and implement the EM algorithm based on the Gaussian Mixture Model

SVM

I used the sk-learn package in python to implement the SVM, the result is shown below.

Linear Kernel with C value varies from 0.25 to 4

As the C value increases, the training error decreases gradually, and the testing error increased instead. Also it can be seen that the testing errors are more than two times the value of training error. The support vectors also decrease with the increasing C values.
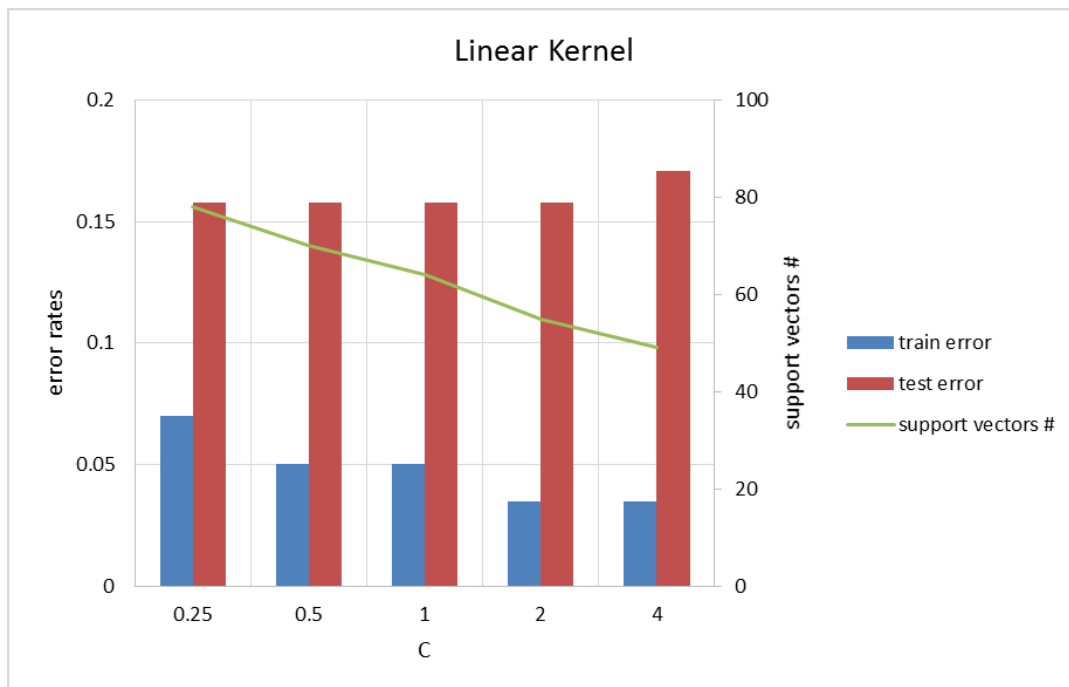

Figure 1 Linear Kernel Results

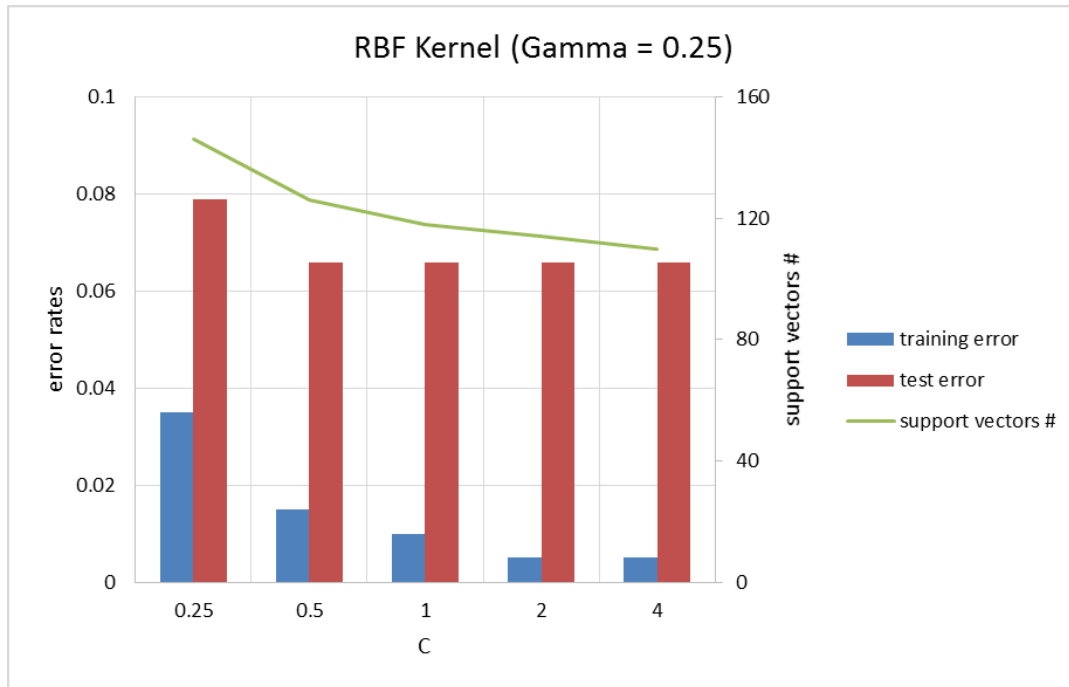The figure 2 – 6 showed the results of RBF kernel SVMs with varying C values.
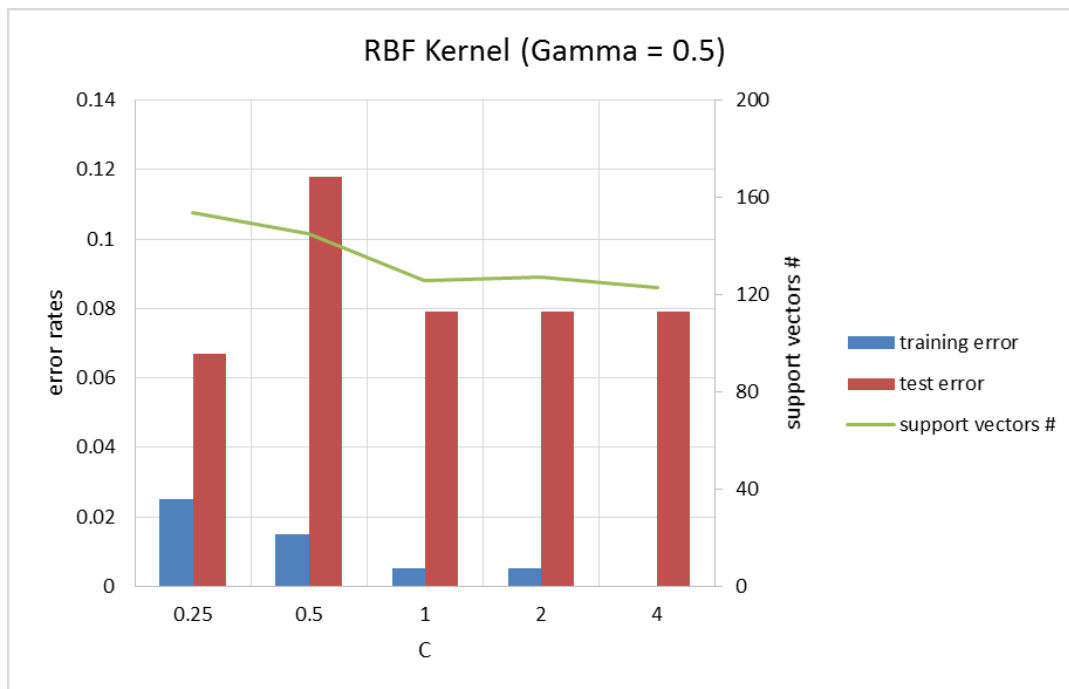
Figure 2 RBF Kernel (Gamma = 0.25)

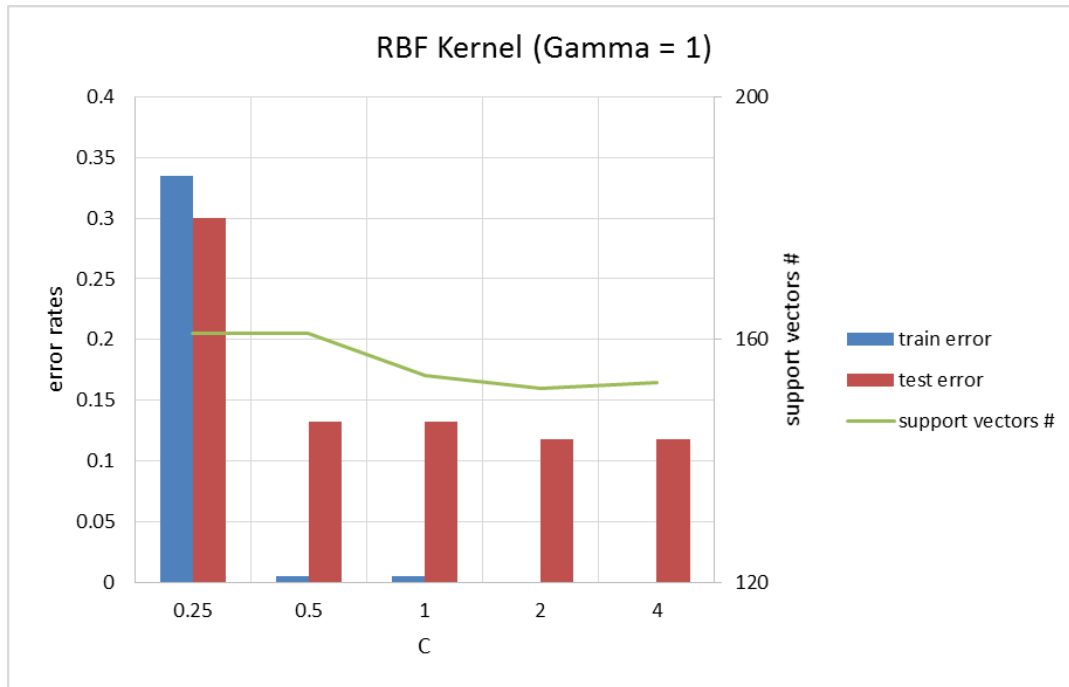

Figure 3 RBF Kernel (Gamma = 0.5)

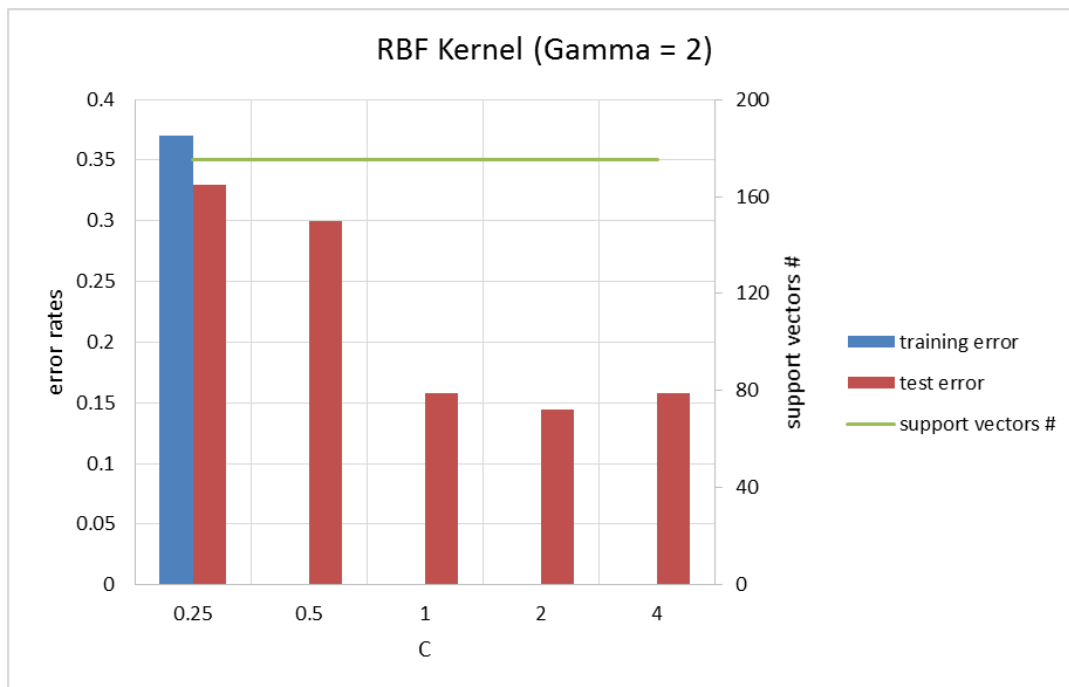*Figure 4 RBF Kernel (Gamma = 1)*
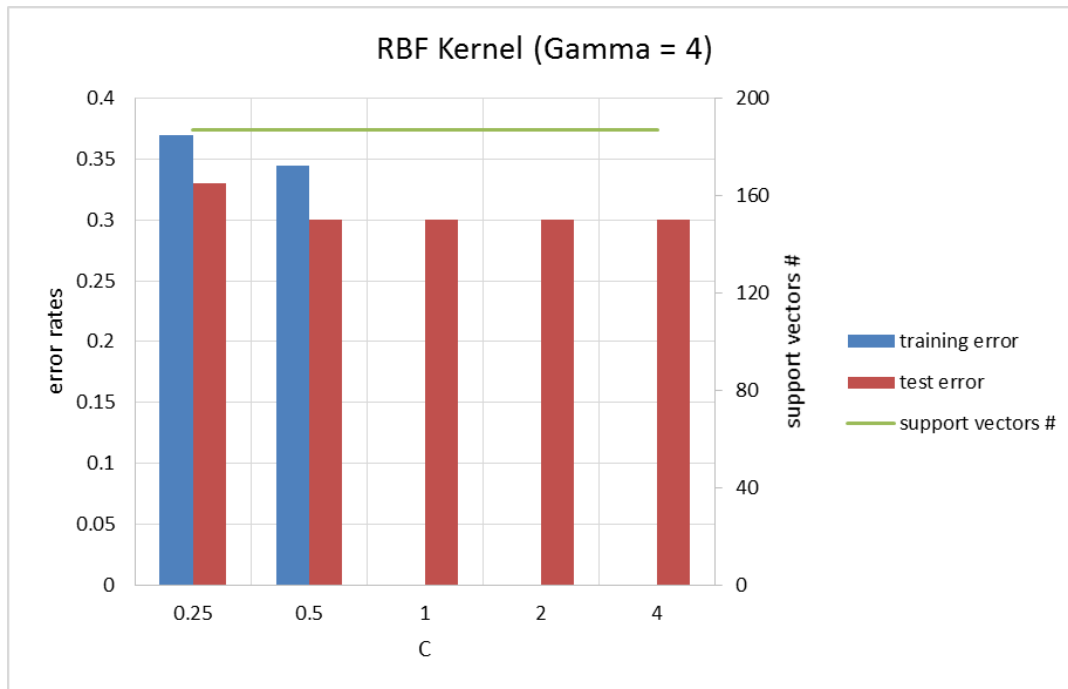

Figure 5 RBF Kernel (Gamma = 2)

Figure 6 RBF Kernel (Gamma = 4)

The graphs from figure 2-6 showed the similar error rates trend, except in figure 5 and figure 6, we have constant support vectors number as C values increases.

Normally speaking, the larger C value would result in less training error, however the larger C values may increase the risk of losing the generalization properties of the classifier in that the optimization process tries to fit as best as possible for the training points, including the possible errors in the training sample dataset. Additionally, a larger C would actually increases the time needed for training.

The smaller C value gives larger-margin hyperplane, which allow for more misclassified samples. So the support vector number should increase.

GMM

GMM based EM algorithm is a non-supervised learning algorithm. I have tried to clustered data given using EM algorithm, but it always gave a "singular matrix" error during my training process. So I just attached my python implementation without giving any results.