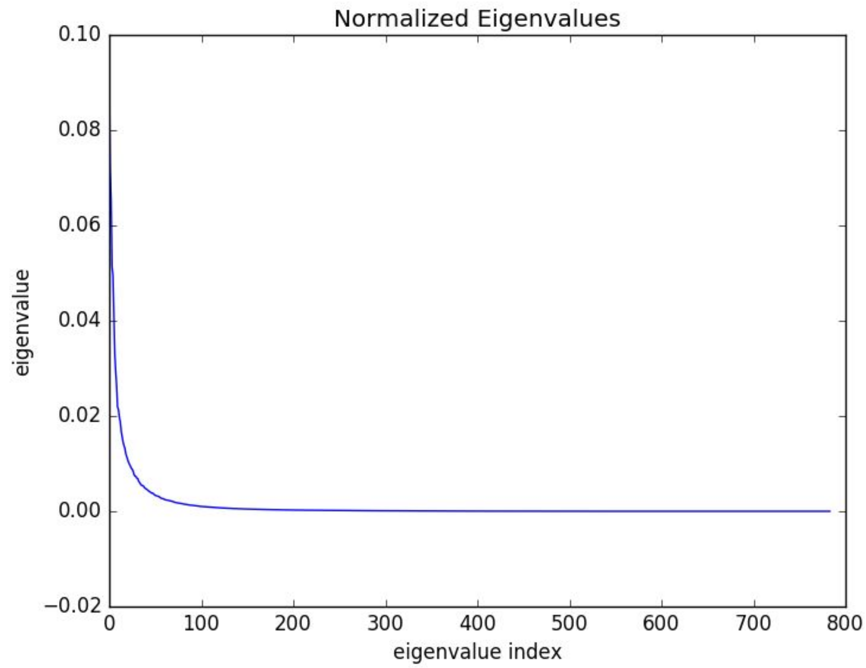
WRITEUP

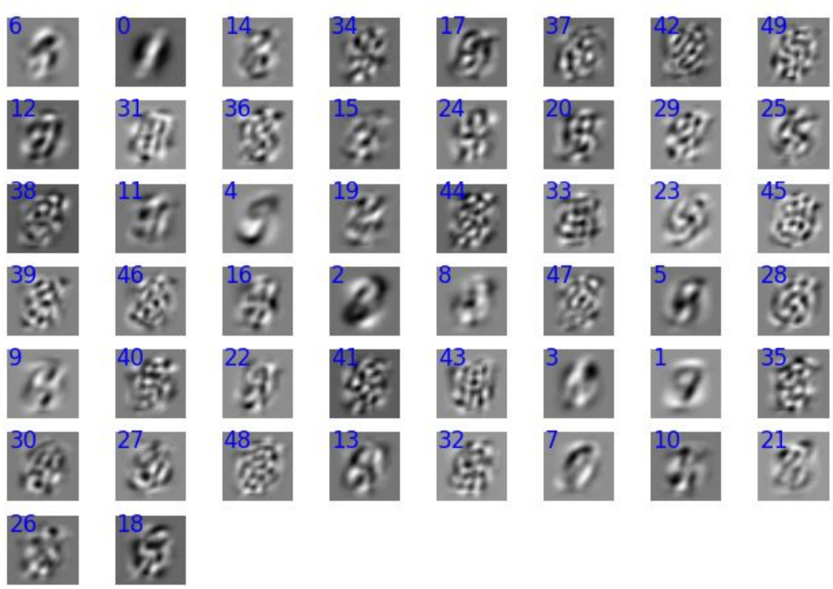
WU1

Need to include 81 eigenvectors to account for 90% of the variance and 135 eigenvectors to account for 95%.



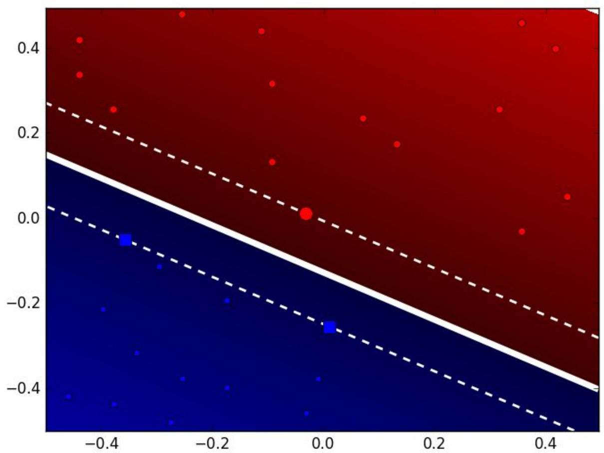
WU2

Most of the top 50 eigenvectors don't look like anything, but a couple seem to resemble handwritten digits, such as vector 4 looking like a 5 and vector 2 looking like a 0. The top eigenvectors account for most of the variance in the data so they likely encode the common basic structures of all the digits. Therefore we wouldn't be able to really tell what they look like individually.

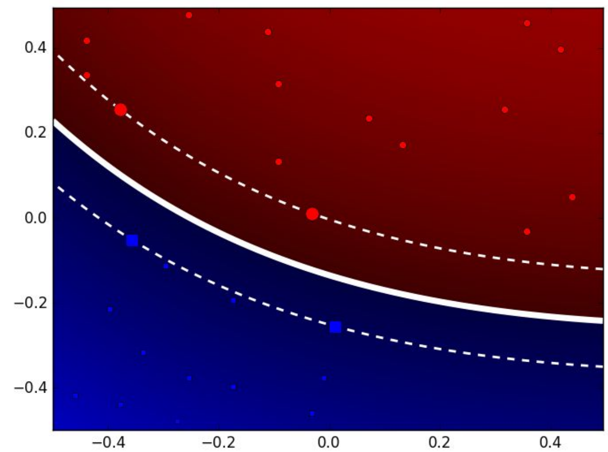


WU3

Linear SVM



10th Degree Polynomial SVM

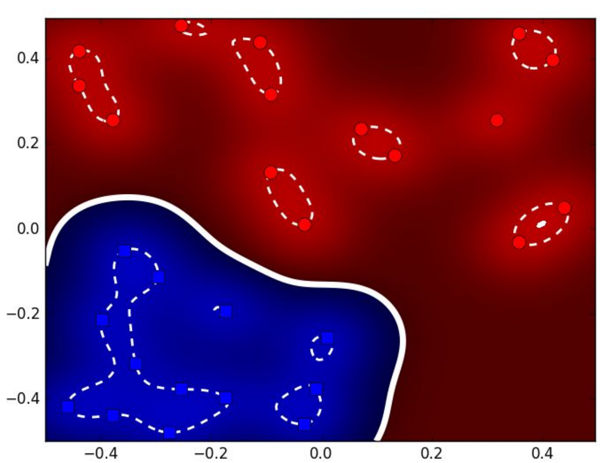


Since this dataset is linearly separable, it is possible to have fewer than 3 support vectors. However, the SVM algorithm should be able to find a hard margin, meaning it finds the margin that is maximally far away from any data point with no room for error. The support vectors of the hyperplane then should be a result of the best that the algorithm can do, meaning it is not possible for the maximum margin to be accompanied by less than the 3 support vectors found. Even increasing overfitting to C = 1000 produces the same 3 support vectors. Decreasing C means increasing the slack and allowing more points to stray inside the margin, creating more support vectors.

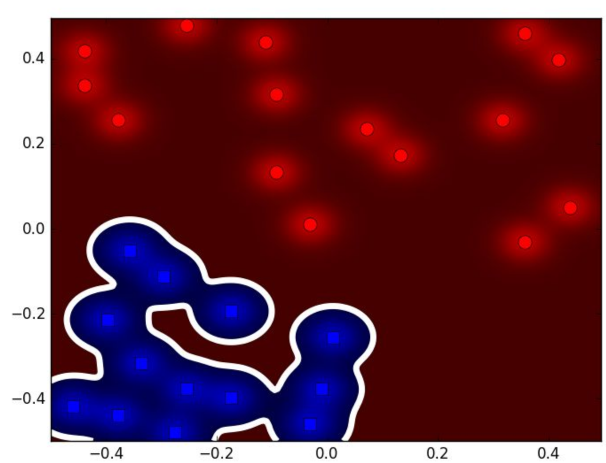
WU4

The “blobs” appear because a higher gamma value means that a data point must be closer to a support vector in order to have a non-zero kernel value. As such, the decision boundary “tightness” also increases, which gives the blob-like shape since the boundary wraps closer to the positive training examples. This can be visualized by the following plots produced using increasing values of gamma beyond 100:

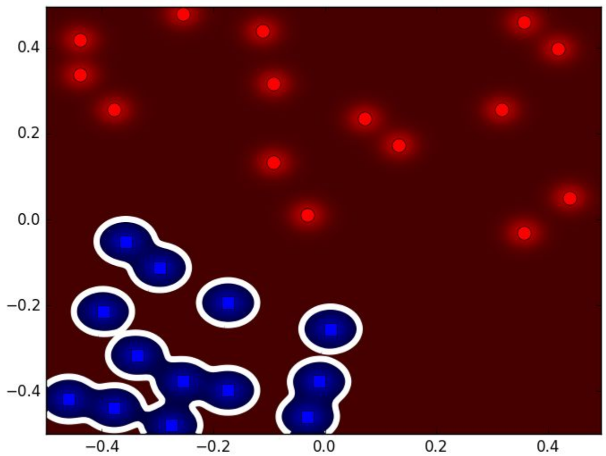
Gamma = 100



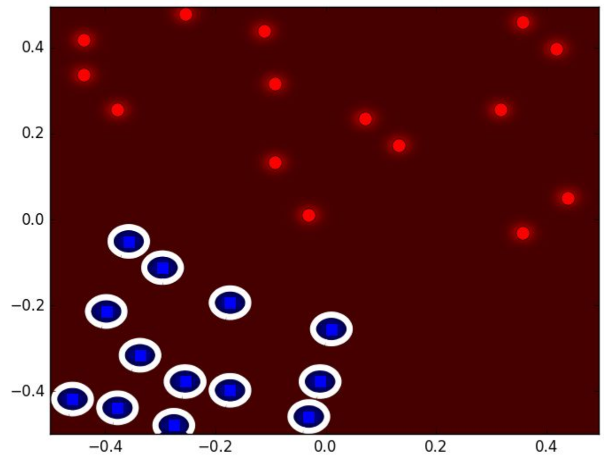
Gamma = 500



Gamma = 1000



Gamma = 2500



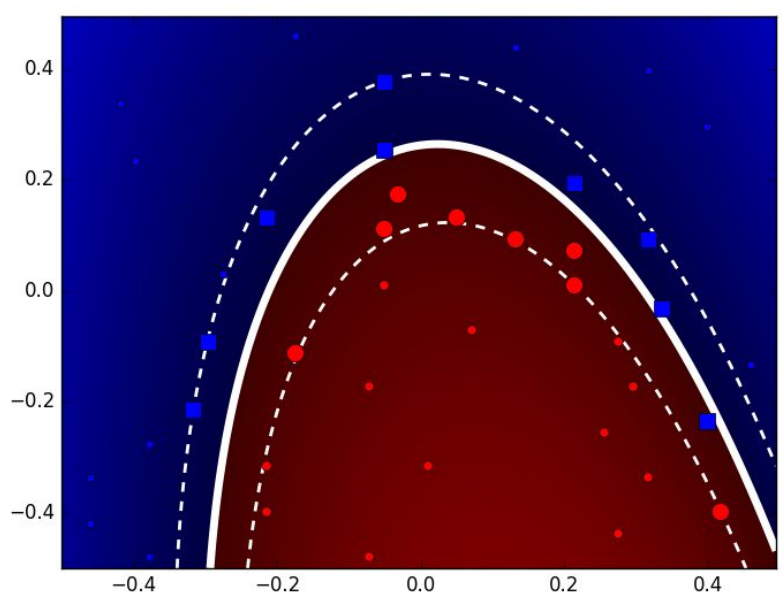
We found that for Gamma = 1000, decision boundaries drawn on the plot seem to isolate individual points but still overlapped, but using a value of 2500, it is clear that every point has its own decision boundary around it.

WU5

The data is not linearly separable so the SVM algorithm must find a soft linear margin that tolerates some error, i.e. red dots on the blue side and blue dots on the red side.

WU6

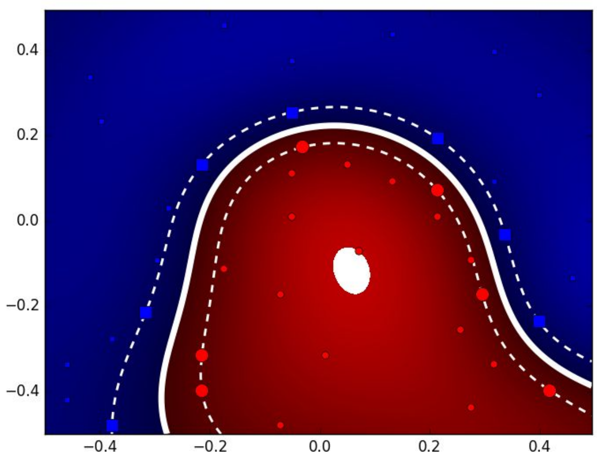
The 0/1 loss on the training data is nonzero since there is one blue point that made it inside the red decision boundary, at around (0.4, -0.2). The hinge loss is also nonzero since there are points inside the margin, meaning $\max(0, 1 - y_i(w^T x_i + b))$ is not 0.



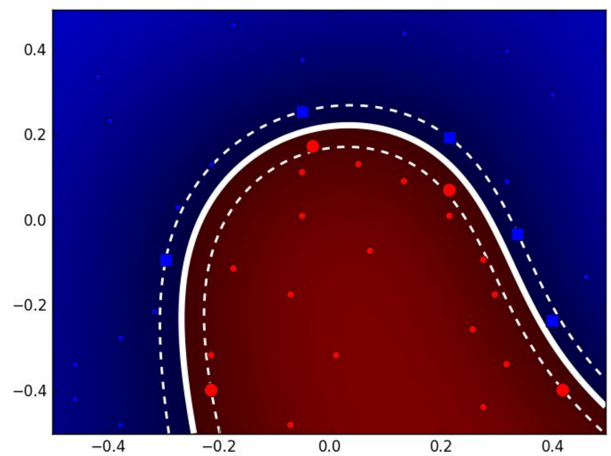
WU7

A good decision boundary should have few support vectors, minimal misclassification, and not be overfit/underfit. The smaller the gamma is the larger the variance, smaller the bias, and wider/smoother the decision boundary is. Out of the 4 small values of gamma plotted below, gamma = 2 seems to give the best decision boundary. It has the fewest support vectors, does not misclassify any data point, and because it's not super small it doesn't give each support vector too wide a range to influence other data points like for gamma = 0.1

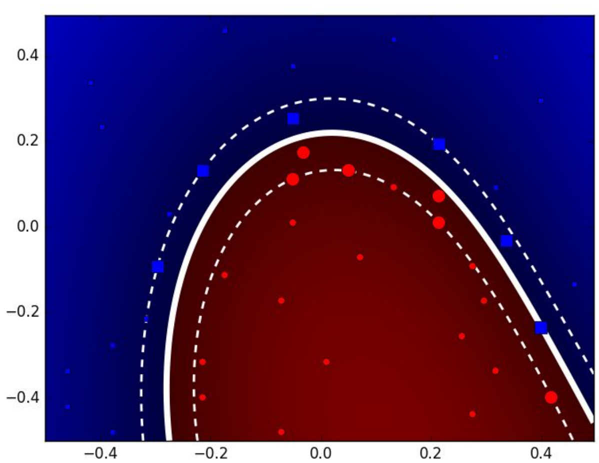
Gamma = 10



Gamma = 2



Gamma = 1



Gamma = 0.1