Project 3 Writeup

Andrea Bajcsy   Michelle Cody   Charles Parker

05/10/2016

WU1 The normalized eigenvalues are shown in Figure 1 below. One needs 81 eigenvectors to account for 90% of the variance and 135 eigenvectors to account for 95% of the variance.
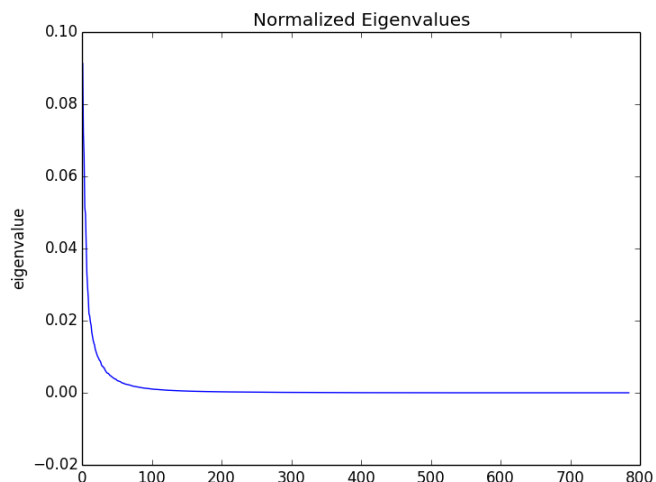

Figure 1: Normalized eigenvalues of the covariance matrix for the digits data

WU2 The plots of the top 50 eigenvectors are shown in Figure 2 below. Besides the first few eigenvectors (those corresponding to labels 0 - 5) which very roughly resemble digits, the images do not look like digits. They should not look like digits because the eigenvectors are the directions of most variance. In terms of the image, this corresponds to the weighted combination of pixels at which there is the highest variance. Since all of the digits are represented in the set and some digits share elements of appearance (such as a vertical bar at the center of the image), it would be very unlikely that a direction of large variance would correspond exactly to one particular digit.
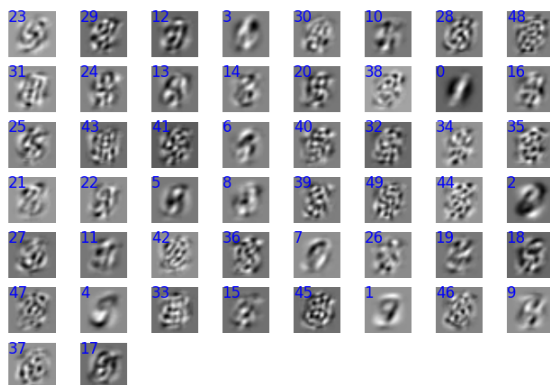

Figure 2: Top 50 eigenvectors

WU3 No, you could not have fewer support vectors. The 3rd support vector included ensures that the line symbolizing the decision boundary has the same margin from each of the three points and is therefore the one correct decision boundary for these input parameters. With only 2 support vectors, there are infinitely many decision boundaries that maintain the margin. These decision boundaries are those that pass through the midpoint of the line between the two support vectors.

WU4 You get these little blobs because you are summing up the normal distributions around each point. As we turn up gamma, the variance increases and the kernel essentially vanishes. When we up gamma up to 2000, we get a little decision boundary around each example (i.e. each decision boundary surrounds exactly one example) (see Figure 3).
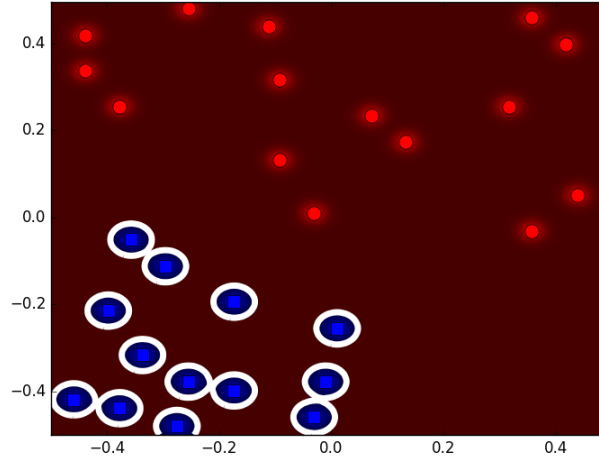

Figure 3: SVM with RBF kernel using gamma = 2000.

WU5 There are a lot of red support vectors on the blue side of the decision boundary because this data is not linearly separable.

WU6 The 0/1 loss on the training data is not zero because it misclassifies one blue point (in lower right hand corner). The hinge loss on the training data is also not zero because there are many points inside the margin.

WU7 When training an RBF kernel on this data, the smallest gamma for which we could get a good decision boundary was gamma = 2. A good decision boundary is defined by a low number of points being used as support vectors and a wider margin (see Figure 4).
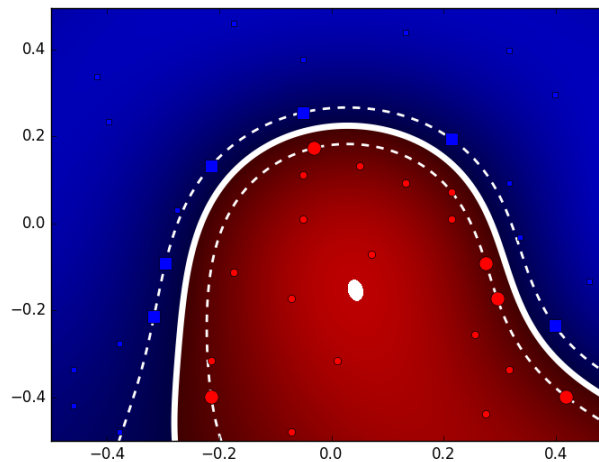

Figure 4: Complex dataset with RBF kernel using gamma = 2.