Data Mining HW 3

Jeff Carney

March 1, 2017

Summary of Classification Results

As can be seen in the plots of purity found for each separating hyperplane in the 'one vs all' case, we were able to achieve a high level of accuracy for all values of lambda with each hyperplane. I then grabbed the optimal lambda for each separating hyperplane and then compute a 785 x 3 β vector and used that to classify all the points in our test X matrix and then evaluated the model with our test Y vector. The results can be seen in a visual representation in the confusion and misclassification matrices. The confusion matrix uses raw numbers for the correct predictions where as the misclassification matrix is normalized so all the values are proportional.