

Peter Smith

PES71

ECE1395

Problem Set 8

- 1) For my SVM classifier my training data achieved a correct prediction rate of 88.4%. On the other 4 training sets it achieved correct prediction rates of 79.5%, 81.8%, 78.3%, 81.4% and for testing data I achieved a correct prediction rate of 80.8%. This result was a little lower than what I expected given the accuracy of the SVM classifier. For my KNN classifier my training data achieved a correct prediction rate of 94.1%. On the other 4 training sets it achieved correct prediction rates of 88.6%, 91.3%, 87.7%, 89.5% and for testing data I achieved a correct prediction rate of 89.2%. These results are about what I expected for the KNN Classifier. For my logistic regression classifier my training data achieved a correct prediction rate of 99.7%. On the other 4 training sets it achieved correct prediction rates of 89.2%, 87.9%, 88.5%, 88.4% and for testing data I achieved a correct prediction rate of 89.4%. This is the result I was expecting from my SVM classifier. I believe that if I trained the SVM classifier with this data set I would have gotten much better results from it. My decision tree classifier was by far the worst performing. On the training data it got a 100% correct prediction rate. On the other training sets I got prediction rates of 68.6%, 66.1%, 67.5%, 67.3% and the testing data received a correct prediction rate of 69.4%. The decision tree was really good on the training set but not to much else. My random forest achieved a prediction rate of 93.1% on the training set. On the other four training sets I achieved prediction rates of 80.8%, 81.6%, 82.6%, 81.1% and 81.4% on the testing dataset. The result I got from the majority voting was 91.6% accuracy. This accuracy was moderately better than the results I got from any of the individual models. I feel like bagging may have hurt my overall results in this instance, for example I feel like training a model using the SVC with all of the samples would have yielded better results. But overall I feel like bagging gave me an accurate model to use to accurately predict the results.
- 2) The results I achieved from this showed how clearly the restarts gave me cleaner looking images no matter how many clusters I was doing. The first image here is an image with 7 k's, 30 iterations and 20 restarts.



The next image has 7 k's, 30 iterations and 5 restarts.



I believe the first image shows a cleaner version of the image.

I don't think that doing many iterations after 20 really made too much of a difference in the images. The next image below has 5 k's, 20 iterations and 20 restarts.



The next image has 5 k's, 30 iterations and 20 restarts.



As far as k's go, it really depends on the image. I think image 2 looks better with 5 k's while image 3 looks better with 7 k's.