HW3: Problem 2

Ian Dover

1 Part 1

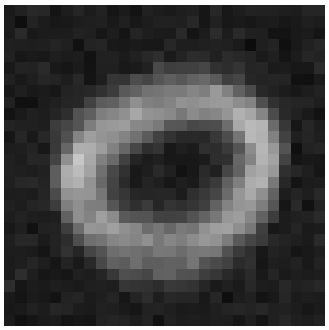

Figure 1: Vector Image of Label = 0

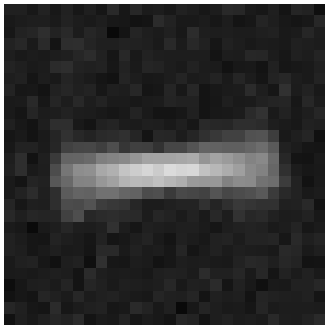

Figure 2: Vector Image of Label = 1

2 Part 2

The ALS algorithm was used to reduce images to a size of 20 x 20. In Python using sklearn.ensemble, a random forest classifier was trained with the specified hyper-parameters

3 Part 3

It appears that the random forest model is overfitting.
Two methods can be used to improve classication performance on the dataset:
(1) Further reduce the dimensionality of the data to remove noise from the image.
(2) Use models such as convolutional neural networks which are more capable of image classification.Accuracy on the training data: 1.0.
Accuracy on the test data: 0.5.
Accuracy on the noisy data: 0.5.