PCA on MNIST Digits

Joe Shymanski

April 14, 2022

1 Data Set

I used the MNIST data set for handwritten digit recognition. I utilized all 60,000 training samples for training the model and the other 10,000 samples for testing the model.

2 Model

I used my implementation of logistic regression from homework three. It utilizes the one-versus-all method to work for the 10 different digit classes. All samples were standardized before undergoing PCA or training the logistic regression model.

3 Experimental Results

# Comps.	Training Time (s)	Train Accuracy (%)	Test Accuracy (%)
N/A	100.38	89.41	88.82
150	30.46	88.90	89.30
100	23.07	85.47	85.64
50	18.92	82.36	82.98

Table 1: Performance with varying number of PCA components.

4   Discussion

As you can see from 1, the model classified the digits with an accuracy of about 90% without any PCA. Using 150 PCA components, the accuracy remained the same, but the model took 70% less time to train. The training time continued to decrease as the number of components dropped, but the accuracy also started to decline.