Project

CSE5052 - Fall 2024

Due Date: Dec/11/2024

Handwritten Image Classification 5 pts

Download the file mnist_49_3000.mat from Canvas. This is a subset of the MNIST handwritten digit database, which is a well-known benchmark database for classification algorithms. This subset contains examples of the digits 4 and 9. The data file contains variables x and y, with the former containing patterns and the latter labels. The images are stored as vectors.

To load the data use the following python code:

```
import scipy.io
import numpy as np
data = scipy.io.loadmat('mnist_49_3000.mat')
x = np.array(data['x'])
y = np.array(data['y'][0])
y[y==-1] = 0
```

To visualize an image, type the following:

```
from matplotlib import pyplot as plt
index = 0 #change the index to show different images
image = x[:,index].reshape(28,28)
plt.imshow(image, interpolation='nearest')
plt.show()
```

Implement a binary classifier (any ML algorithm discussed in the class would be fine). Use the first 2000 examples as training data, and the last 1000 as test data.

(a) Create a video (5 minutes or shorter) and explain how your code is working. In the video, please report the test accuracy and show 5 images that have been misclassified in the test dataset.

(b) To receive the full credit, please upload the code and the video on Canvas as a single zip file.