Project 1

Each group should

- Submit one lab report containing procedures and results to Canvas by April 12, midnight. See Report Guidelines below.
- Include a thoughtful and reflective paragraph from each team member summarizing what has been learned.
- Include a section describing the work done by each member of the group.
- Attach your well-documented code in the appendix of the report.
- You can use any programming language to do the project. MATLAB is preferred.

Report Guidelines

1. The Cover Page identifies the course number and name, the Project number, group member's names, and the report delivery date.

2. Make sure all steps in the assignment are followed.

2. Use words/sample_code to describe the procedures (How you did it). Explain the steps taken if needed.

3. Show and evaluate the results. Use words to describe the results and provide explanations if needed.

4. Number all the images/figures with captions. Refer to these images/figures in the write-up.

5. Neatness and a professional presentation count!

Feature Analysis and Selection for Handwritten Digit Recognition

The MNIST (Modified National Institute of Standards and Technology) dataset is a collection of handwritten digits used in OCR (optical character recognition) and is considered as one of the benchmark datasets for pattern recognition and machine learning. An image containing all training samples of digit 0 is shown below. Note each sample has a standard size of 28x28 pixels arranged in a regular grid. Electronic copies of all ten images can be found on Canvas.

1. In this project, you will examine various features extracted from the sample digits and select two best features for digits classification of 0 and 1.
2. For each image, do the following:
    a. Load the image and convert it to a binary image.
    b. Find samples of the digit by extracting the regular grids in the binary image.
    c. Extract shape measurements (features) from each sample digit. (You can use regionprops() in MATLAB. You can also consider other derived features from these measurements.)

3. Analyze the extracted features by building and comparing the histograms. For this step you don't need to show all shape features. Choose a few.
4. Based on the results in Step 3, select two best features. Plot the training data in 2-D scatter plot. Do you think a good classifier can be built based on these two features?

5. Now find the best feature(s) to classify 0 and 2. Repeat steps 3 and 4.