Machine Learning Lab1

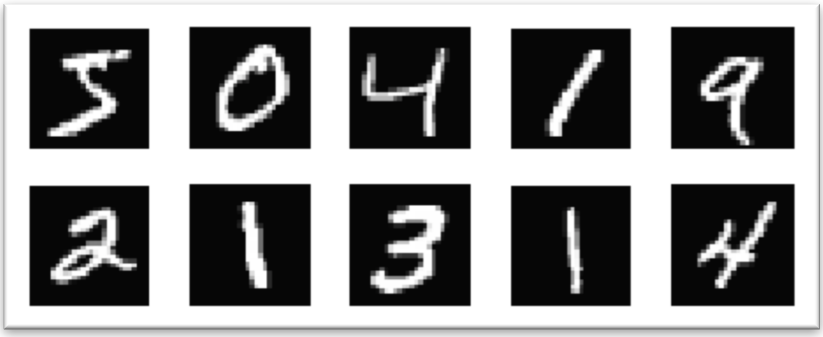
Throughout the lab sessions, we will continuously do experiments on the MNIST digits dataset. The ultimate goal is to build a digit recognition system using Matlab.

Task 1: Getting started with Matlab

Matlab, which is short for Matrix Laboratory, is probably the most widely used scientific and engineering numerical software. We have uploaded a few basic tutorials to help you get started. Please check the Moodle module page to get access to them.

Task 2: Load and display the MNIST dataset

1. MNIST digits recognition dataset is one of the most widely used datasets in machine learning. It contains 60,000 training samples and 10,000 test samples. Data files used in lab session are also provided in module page, please download all files contained in the "data" folder and unzip these files in your computer. Alternatively, you can download data directly from the original website http://yann.lecun.com/exdb/mnist/.
2. Please carefully read the instructions therein about the structure of how the data are stored. Try to load the data to Matlab and display them. The first 10 digits in the training set should be like this:



Tips:

1. In loading the MNIST data, probably you will rely on the following functions in Matlab: fopen, fread and fclose.
2. In displaying the data, you will need to use the following functions: reshape, imagesc and imshow.
3. Just type 'help XXX' in matlab to see how to use the above functions.
4. If you encounter the 'out of memory' problem in Matlab, which means the data is too large to fit in to the memory, and you have to delete unnecessary variables in the workspace or save them to the hard disk.