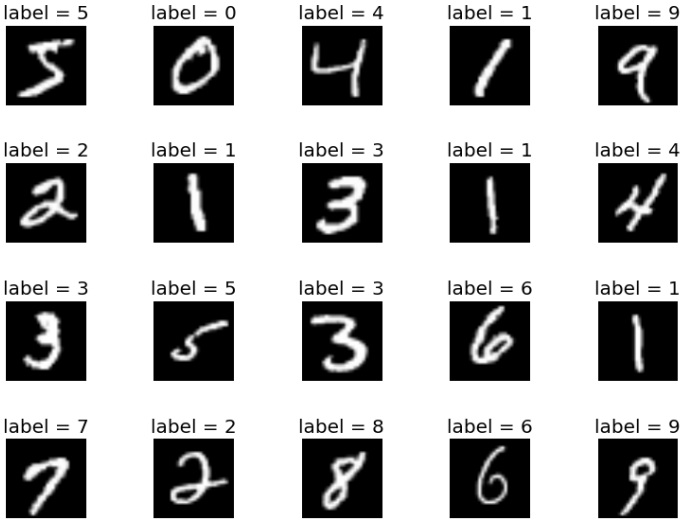
1. *Image classification by logistic regression.*

In this homework, we solve a real-world image classification problem by logistic regression.

Data Set and Data Format: The MNIST database of handwritten digits, available here, is a well-known database that contains a training set of 60,000 examples and a test set of 10,000 examples. Each example includes an image of a handwritten digit and the corresponding label (i.e., ground truth of which digit is in the image). Some samples of the data are as follows:



Each image has $28 \times 28 = 784$ pixels, each of which is represented by the gray level. We provide codes (`load_mnist` and `load_mnist_5_6`) that give you the data in the following formats:

- `train_image`: a 785-by-m matrix, each element of which is a real number in $[0, 1]$;
- `train_label`: a 1-by-m row vector, each element of which is an integer in $\{0, 1\}$;
- `test_image`: a 785-by-$\hat{m}$ matrix, each element of which is a real number in $[0, 1]$;
- `test_label`: a 1-by-$\hat{m}$ row vector, each element of which is an integer in $\{0, 1\}$.

The training data is contained in `train_image` and `train_label`, and the test data is contained in `test_image` and `test_label`. Note that:

- Each column of the matrices `train_image` and `test_image` represents the 784 pixels of an image. We add `1` as the first element of each column for normalization. Therefore, the length of each column is $1 + 784 = 785$.
- The functions (`load_mnist` and `load_mnist_5_6`) allow you to change m, namely the number of training examples.

Assignment: Your task is to fill in the blank in the code `logistic_regression`. Specifically, you need to compute the coefficient $x \in \mathbb{R}^7 85$ of the logistic regressor, using the m training examples. Then the code will output the accuracy of your regressor based on the evaluation from the $\hat{m}$ test examples. We will use CVX to solve the optimization problem for the coefficient x. More detailed descriptions of the tasks are as follows.

- Do the image classification task for digits 0 and 1, using $m = 10$ and $m = 50$ training examples, respectively. Observe the runtime of CVX and the quality of the solution (as indicated by the CVX solver). Report the accuracy.
- Do the above tasks for the image classification for digits 5 and 6.

More Details About The Codes:

If you use Matlab, please do the following to setup:

- download the data set (i.e., four files) from http://yann.lecun.com/exdb/mnist/;
- download `load_mnist.m`, `load_mnist_5_6.m`, and `logistic_regression.m`;
- fill in the blank in `logistic_regression.m` to compute the regressor.

If you use Python, please do the following to setup:

- download the data set (i.e., four files) from http://yann.lecun.com/exdb/mnist/;
- install `scikit-learn` in Python from http://scikit-learn.org/stable/install.html;
- download `load_mnist.py`, `load_mnist_5_6.py`, and `logistic_regression.py`;
- fill in the blank in `logistic_regression.py` to compute the regressor.