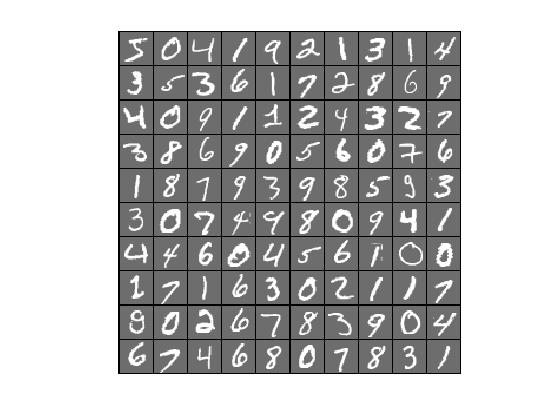
## Handwriting recognition

Handwriting recognition is a well-studied subject in computer vision and has found wide applications in our daily life (such as USPS mail sorting). In this project, we will explore various machine learning techniques for recognizing handwriting digits. The dataset you will be using is the well-known MINST dataset.

1. The MNIST database of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. (<http://yann.lecun.com/exdb/mnist/>)

Below is an example of some digits from the MNIST dataset.



The goal of this project is to build a 10-class classifier to recognize those handwriting digits as accurately as you can. All the assignments below should use the training data (60K examples) and test data (10 K examples) as given by the dataset.

Here are the basic requirements for this project.

### Assignment #1

Do a clustering of the training dataset with the cluster number K=10 first (without looking at the labels), and then assign a class label to each cluster based on majority vote of the cluster member’s known digital labels.

Based on that, **compute the training classification accuracy for 10 classes from clustering**. Then use each cluster’s centroid as the anchor point, and you will obtain 10 class centroids corresponding 10 anchor points.

To **compute testing accuracy**, first assign each test image to the nearest centroid with the corresponding centroid label, and then compute the test accuracy.

### Assignment #2

Build a number of non DNN based classifiers using all pixels as features for handwriting recognition. You need to use **at least two the following four techniques** we have learned from the class to do the work:

* Logistic regression
* SVM
* Decision tree, and
* Random forest.

For each technique, please use your own language to give a general description of the technique, its pros and cons, and why such a technique is suitable for solving the handwriting recognition problem. The goal is to make sure you know why you decide to choose this technique.

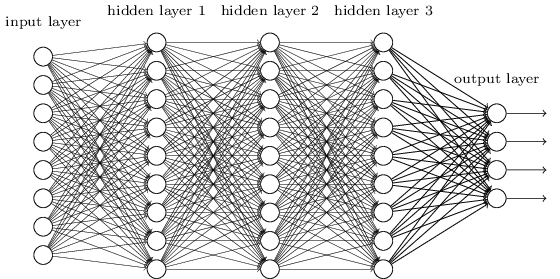
Please also **make a comparison table** among the four techniques as well as the method from Assignment #1 above.

### Assignment #3

In this assignment, we will explore various techniques related to neural network and deep learning to solve the 10-class classification problem.

Since there are many existing implementations to solve the MINST problem, we need to give some twists to this problem to make it worthwhile to do for our final project. (Please refer to the ranking list for MNIST at <http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html>.)

The basic network structure that we are trying to explore is something like the following (i.e., the fully connected deep neural nets). The number of hidden layers and the size of each hidden layer in terms of neurons are left as tuning parameters that you can explore.



In ‘Assignment4.py’ file, the guide of implementing CNN is provided. Finish the code and report the accuracy.

### Assignment #4

This assignment reflects the data collection process. Everyone is required to

* Hand write 5 styles of your own digits from 0 to 9 on a paper, and make sure your own handwritings are for sure recognizable by yourself. Please take a picture of each digit you write (so you have total 5 x 10 = 50 images), resize and convert it to the same data input format as the MNIST dataset. In other words, you have 50 new data points with labels.
* Treat these 50 images as “brand-new” test dataset and run your own ML models from Assignment #1 - #4 on these 50 images and report the achieved test accuracy. Note, the goal for this exercise is not for achieving “high” accuracy, but to show what potential gaps there may be between existing MNIST dataset and your own test dataset, a scenario you would encounter in real life.

In your report (not formal) and submission, the following is required for this assignment

* Show all 50 images you create with the corresponding labels you intend to assign,
* Make a table to show the test accuracy on these 50 images for each ML model you obtained from Assignment #1 to #4
* Submit in a subfolder with your 50 handwriting dataset in MNIST format.