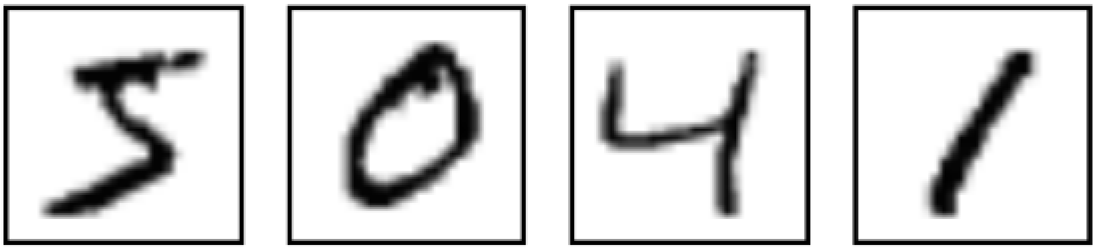
Machine Learning

MNIST is a standardized computer vision dataset that consists of images of handwritten digits. The following images represent example images in this dataset.



The dataset is commonly labeled with the associated digit for each image. For example, labels for the above images are 5, 0, 4, and 1.

Tasks

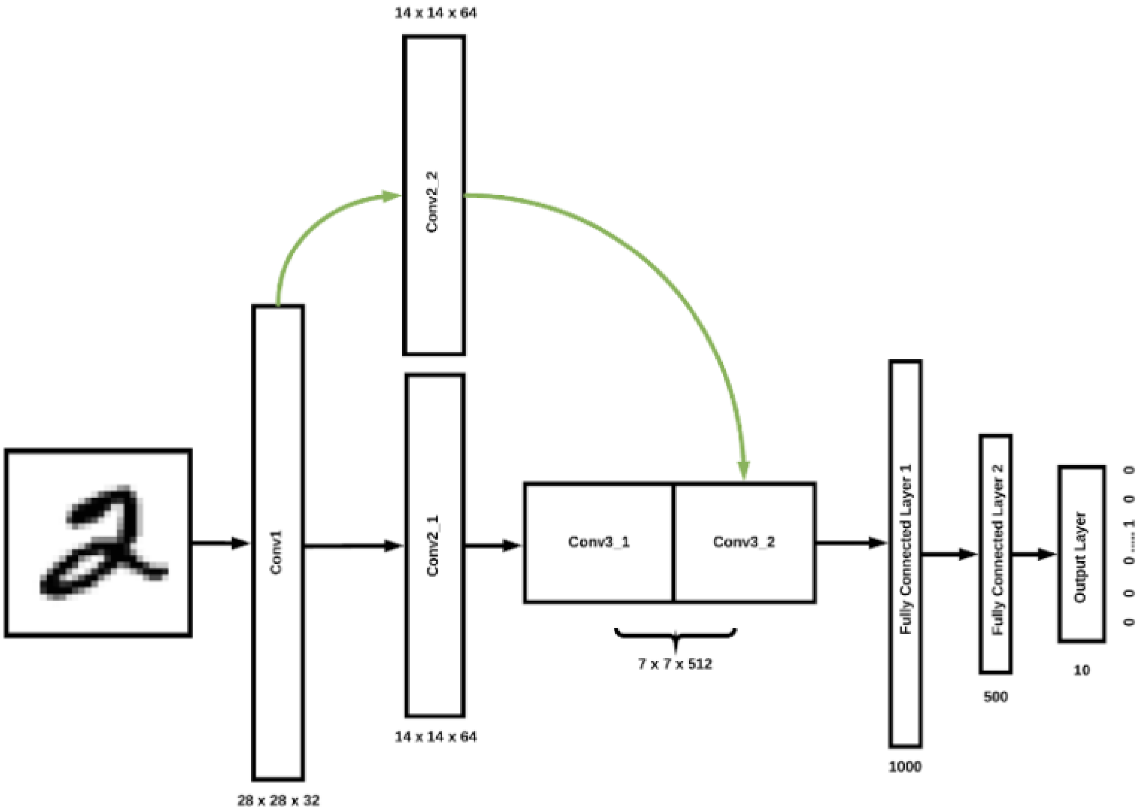
1. Develop a simple neural network training pipeline (with frameworks such as Tensorflow, Keras, Pytorch, etc.) for MNIST handwritten digit classification.
2. Implement an inference server with the trained model developed in the previous step. The server should accept raw images as input and predict the number with a probability.

Requirements

For MNIST datasets, you can use one of below resources:

- Download directly from original host website
- Tensorflow API
- Tensorflow Keras API
- Pytorch MNIST API

The architecture of the neural network should follow the following guidelines:



Additional details for this NN:

- It contains three convolutional neural network layers (shown above as conv1, conv2_1, conv2_2, conv3_1 and conv3_2), two fully connected layers (named as fc1and fc2) and one output layer (named as output)
- All nonlinear activations should be implemented using Rectifier Linear Unit (ReLu) and for downsampling should be implemented using max-pooling
- conv3 is a combined layer from conv3_1 and conv3_2
- You can define the default value for hyper-parameters that are not listed above

Expectations

- Consider the quality of the software implementation as shippable for production
- Consider the robustness of the training model
- Consider the performance of the inference implementation
- Provide test coverage, where appropriate
- Document with instructions on how to reproduce the model and run the inference server on additional data
- Provide a zip/tarball of your code along with details about how to build and run it