Assignment 2
Multinominal Classification

1 Objective

- Building a model for multinominal classification task.
- Knowing the effect of using CNNs in models.

2 Problem Statement

In this assignment we will be addressing the task of multinominal classification of handwritten digits from the famous MNIST dataset.
The MNIST dataset consists of 60,000 training images and 10,000 test images. Our classes are the digits 0-9.
You are required to build 2 models to solve this problem:

1. A simple using fully connected layers.
2. A model using Convolutional Neural Network (CNN) before applying the simple model.

you should observe the difference in accuracy between the 2 models.

3 Requirements

1. Build a neural network, using tf.keras, consisting of 2 fully connected layers and apply this to the digit classification task, Our network will ultimately output a probability distribution over the 10 digit classes (0-9).
2. Compile your model, you'll want to experiment with both the choice of optimizer and learning rate and evaluate how these affect the accuracy of the trained model.
3. Evaluate accuracy on the test dataset, what is your observation on the accuracy on the the test dataset and the accuracy on the training dataset?
4. Build a CNN, using tf.keras, composed of two convolutional layers and pooling layers, followed by two fully connected layers, and ultimately output a probability distribution over the 10 digit classes (0-9).
5. Compile the CNN model with an optimizer, learning rate and regularizer of choice.
6. You can find the architecture of the required models in the next section.

7. You are required to show the running time, number of parameters and number of multiplication of your model in estimation and in training over the same number of epochs.

8. What is the highest accuracy you're able to achieve using the CNN model, and how does the accuracy of the CNN model compare to the accuracy of the simple fully connected network?

9. What is the result of using Relu and tanh activations, which of them produced a better result?

10. Try different convolution filter sizes, stride lengths, and pool layers. Give plots for the effect of those changes.

11. Use the CNN model to predict the digit in the first image of the test set and compare it to the first test label to check if the model prediction is correct.

Important Note: The comparison between different hyper parameters or any model changes should be measured in accuracy not in loss.

4 Model Architecture

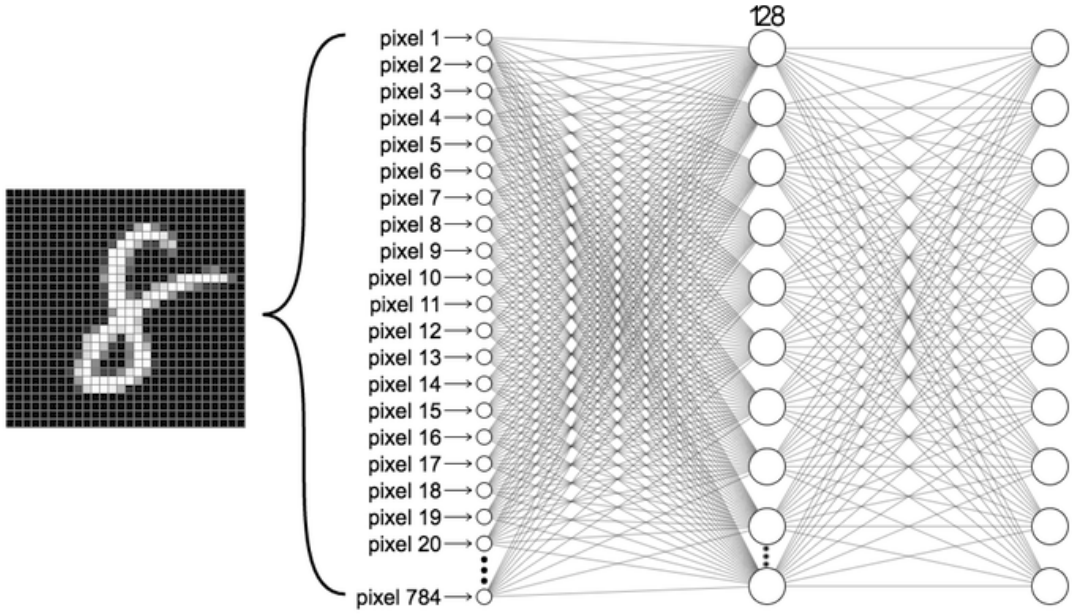

Figure 1: Simple Model

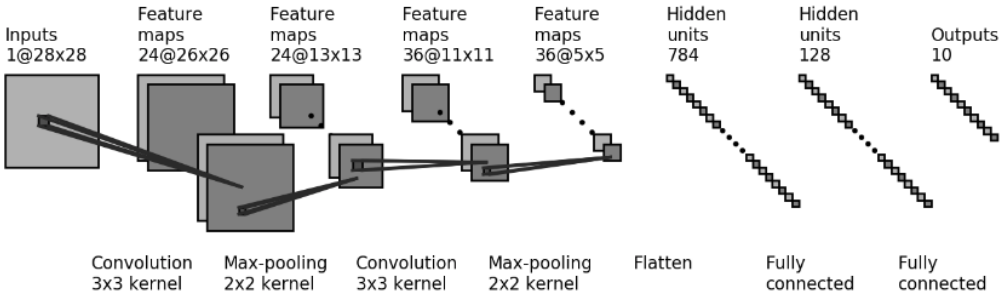

Figure 2: CNN Model

5 Notes

- Parts of this assignment are based on MIT 6.S191: Introduction to Deep Learning course.
- The starter code for the problems can be found in the resources section in Piazza.
- You should deliver a PDF report that has tables summarize all your work and experiments.
- Cheating will be severely penalized (for both parties). So, it is better to deliver nothing than deliver a copy. Any online resources used must be clearly idenified.