Machine Learning 19-20

Group 55
Mahima Raut, 97569
Sarah Seifi, 97658
Codelab Link

***Abstract*— This report entails the description and explanation of neural networks used in experiments as a part of an assignment given in the class of Machine Learning, a presentation of their implementation, evaluation and results.**

I. Introduction

For this homework, a neural network architecture was built using Tensorflow and Keras, intended to classify the famous MNIST[1] data set into 10 classes of digits (0-9). We tried different ways of optimizing, e.g. changing the no. of hidden layers and using different methods of regularisation to achieve the highest performance on the final test set.

II. Dataset

The MNIST database of handwritten digits has a training set of 60,000 examples, and a test set of 10,000 examples. The digits have been size-normalized and centered in a fixed-size image. It is a good database for people who want to try learning techniques and pattern recognition methods on real-world data while spending minimal efforts on preprocessing and formatting, which is why it's widely used for introducing image classification problems, comprising of the following attributes:

- 70k examples of handwritten digits
- Image size: 28x28
- 1 channel
- 10 classes: [0-9]

Throughout the implementation of various neural networks, we have taken 20% of the training set to serve as the validation set.

[1] MNIST data set (http://yann.lecun.com/exdb/mnist/)

III. Models

To classify the images of digits, we will require neural networks. In particular, for the current assessment we will use Feed Forward Neural Networks *(FFNN)* and Convolutional Neural Networks *(CNN)*.

A. Feed Forward Neural Network

1) Structure: Feed-Forward Neural Networks (FFNN) or also known multilayer perceptrons are the foundation of most deep learning models. The main goal of a FFNN is to approximate some function $f*$. For example, a regression function $y = f*(x)$ maps an input x to a value y. A FFNN defines a mapping $y = f(x; \Theta)$ and learns the value of the parameters that result in the best function approximation.

The reason these networks are called feed-forward is that the flow of information takes place in the forward direction.In this, if we add feedback from the last hidden layer to the first hidden layer it would represent a recurrent neural network.

2) Application: Today, FFNNS are one of the most used machine learning techniques for solving regression and classification problems in supervised learning. Networks like CNNs and RNNs are just some special cases of FFNNs. These networks are mostly used for supervised machine learning tasks where we already know the target function.

B. Convolutional Neural Network

1) Structure: The structure of CNNs are inspired by biological processes, mainly the visual cortex of the human brain. Fundamentally, it consists of multiple convolutional layers followed by a pooling layer. This composition can occur multiple times and is then called a Deep Convolutional Neural Network.

2) Application: CNNS are referred to as State-Of-The-Art method for various applications in the field of classification. They are especially successful and therefore widely used in fields as image recognition or processing of audio data. Therefore, in this paper this network is used to classify the MNIST dataset.

C. Regularization

While adding layers to the network, regularization can also be included using three different parameters: Kernel Regularizer, Bias Regularizer, and Activity Regularizer. The first applies regularization to the weights of the layer, the second to its bias, and the last to its output. As part of this assignment, a L2 regularization was applied to the weights of the FFNN with two hidden layers and a dropout regularization to the CNN with two hidden layers.

IV. Implementation

A. Feed Forward Neural Network

First of all, a basic FFNN with only an input and output layer was implemented. As a second step, a FFNN with one hidden layer and one with two hidden layers were implemented.

B. Convolutional Neural Network

Additionally, a CNN was implemented with no hidden layer and with two hidden layers, respectively. As regularization, a dropout regularizer was applied on the model with two hidden layers.

V. Results

A. Feed Forward Network

In the following table, the accuracy values of the respective models are shown.

	Baseline	1 HL	2 HL	2 HL L2
accuracy	0.9207	0.9701	0.9711	0.9719

The performance of the four models are graphically illustrated in the provided codebook.

B. Convolutional Neural Network

In this table, the accuracy values of the CNN model, the CNN with two hidden layers and CNN model with two hidden layers and an added dropout layer is presented.

	CNN	2 HL	2 HL Drop
accuracy	0.9806	0.9830	0.9851

The performance of the three models are graphically illustrated in the provided codebook.

VI. Conclusion

A. Feed Forward Neural Network

Although there is significant improvement in the accuracy when 1 hidden layer is introduced in contrast with the baseline model, we do not observe substantial improvement when hidden layers are increased from 1 to 2 in an FFNN. It does take double the time for every epoch approximately on an average, when 2 hidden layers exist instead of 1. When adding a L2 regularization, we can see that the model converges faster and plotting the evolution shows, that there are fewer oscillations. In short, FFNN with two hidden layers can represent functions with any kind of shape. There is no reason to use FFNN with any more than two hidden layers. In fact, for many practical problems, there is no reason to use any more than one hidden layer.

B. Convolutional Neural Network

Looking at the accuracy of the different CNN models, we can see that increasing the number of hidden layers slightly increased the accuracy of the model. The idea behind dropout is to disable a percentage of randomly selected neurons during each step of the training phase, in order to avoid overfitting. While looking at the graphs on the model, we can see that the dropout method removes any sign of overfitting and therefore provides the best model.

C. General Conclusion

Generally speaking, we can see that for classifying images like in the MNIST dataset, a CNN with added hidden layers and a dropout regularization method provides the best results.