Lab Report- CNN for MNIST Hand-written Digits

emil.haglund@umu.se

January 2021

1 Aim

The aim of this lab is to implement convolutional neural network models for classification of hand-written digits from the MNIST dataset. Experiments will be conducted to study the effect of cnn architecture, size of training dataset and training length (the number of epochs) on the models performance.

2 Background

Convolutional Neural Networks (CNN) are a type of deep learning networks commonly used to analyze images and within computer vision.

CNN's take an image as input and assigns learnable weights to various objects or features of the image. Like a regular feedforward neural network a convolutional neural network consists of an input, an output layer and hidden layers. Hidden layers in a CNN can consist of convolutional layers, pooling layers and regularization layers. Convolutional layers use kernel multiplication to transform the image to a feature map. Pooling layers reduce the dimensionality of data to decrease the computational power required to process the data.

To train the CNN models, stochastic gradient descent (SGD) was used as the learning algorithm. An important hyperparameter for SGD is the number of epochs, which controls the number of complete passes through the training dataset. The larger the number of epochs, the more times the internal model weights are updated and adjusted to the training dataset. The number of epochs should be set to allow the learning algorithm to run until the error from the model on a validation dataset has been sufficiently minimized.

3 Method

The implementation and evaluation framework used for these experiments is largely based on a tutorial by Jason Brownlee. See: https://machinelearningmastery.com/how-to-develop-a-convolutional-neural-network-from-scratch-for-mnist-handwritten-digit-classification/.

Two different convolutional neural network models were trained. The architectures of the models are described below. The architectures are similar but in model 2, two convolutional and one pooling layer has been removed.

Model 1 architecture:

- input layer- convolutional layer- pooling layer- convolutional layer- convolutional layer- pooling layer-fully connected layer- output layer

Model 2 architecture:

- input layer- convolutional layer- pooling layer- convolutional layer- convolutional layer- pooling layer-fully connected layer- output layer

The same experiments were conducted on both models to study the effect of the size of the dataset and training length on the models performance measured in classification accuracy. Models were trained on 60 thousand (60k) and 30 thousand (30k) datapoints that were randomly sampled from the mnist training dataset. The models were tested on 10 thousand datapoints from the mnist test dataset. Classification accuracy was recorded after 1, 5 and 10 training epochs.

4 Results

4.1 Model 1

	epochs=1	**epochs=5**	**epochs=10**
30k datapoints	97.00%	98.12%	98.74%
60k datapoints	97.56%	99.03%	99.09%

Table 1: Accuracy scores for model 1 trained with 30k and 60k number of datapoints. Accuracy scores were recorded at epoch 1,5, 10

4.2 Model 2

	epochs=1	**epochs=5**	**epochs=10**
30k datapoints	96.02%	98.00%	98.27%
60k datapoints	97.55%	98.67%	98.85%

Table 2: Accuracy scores for model 2 trained with 30k and 60k number of datapoints. Accuracy scores were recorded at epoch 1,5, 10

5 Discussion

The implementations of convolutional neural networks were able to accurately classify hand-written digits from the MNIST dataset. Model 1 was able to predict the correct hand-written digit in over 99% of cases when training on 60 thousand datapoints, see Table 1. Decreasing the number of training samples by half from 60 thousand to 30 thousand reduced accuracy scores for all test cases. Accuracy scores were still high however, model 1 trained on 30 thousand datapoints had an accuracy of 90.74% after 10 epochs of training. That fewer datapoints increased generalization error is an expected result.

As part of the experiments the length of training was varied between 1, 5 and 10 training epochs. For both models trained with 30 and 60 thousand training datapoints 10 epochs produced the highest accuracy scores. This could indicate that more epochs would further reduce the test accuracy scores. However, more epochs could also lead to overfitting on the training set. Early stopping is a technique to select the number of epochs for training which stops training once the model performance stops improving on a hold out validation dataset.

In the comparison between models, model 1 outperformed model 2 on all test cases. Model 1 is a more complex model than model 2 with two additional convolutional layers and one additional pooling layers. This complexity allowed the model to capture more of the nuances of the hand-written digit images resulting in a higher classification accuracy.