Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
and Geometry Lab

Computer Vision Lecture

Assignment 1 Report

Autumn Term 2021

Author:
Pascal Lieberherr

Contents

Chapter 1

Report

1.1 Q2.3

I obtain a accuracy of 48.6111%. Yes, this is an expected results because the data is not linearly separable. The capacity of the neural net is limited to linear functions since it only has one linear layer without any non-linearities. Since the data is not linearly separable the neural net can not learn the training data set.

1.2 Q2.4

I achieve a accuracy of 100%. This is because the capacity of the neural net now also allows to learn non linear decision boundaries. This is we have included non-linear activation functions (ReLU).

1.3 2.5

We could switch to polar coordinates by applying the following feature transform.

$$r' = \sqrt{x^2 + y^2} \tag{1.1}$$

$$\phi' = atan2(y, x) \tag{1.2}$$

Now, the data is linearly separable simply because the radius of the two classes differs. With this feature transform I obtain an accuracy of 99.4048% with a linear classifier.

1.4 Q3.3

The accuracy with one linear layer is: 91.76 % The accuracy of the MLP with one hidden layer plus ReLU and then a final prediction is: 94.84 %

1.5 Q3.4

I obtain a testing accuracy of: 98.54 %

1.6 Q3.5

MLP one hidden Layer
This is a fully connected layer with 28*28 nodes in the input layer, 32 nodes in the hidden layer and 10 output nodes. Thus, we have 28*28*32*10 = 250'880 weight parameters. In addition there are 32 bias weights for the first layer and another 10 for the output layer. As a results, we end up with 250'922 parameters.

CNN
Please note that ReLU and max pooling dont have parameters. The general formula for computing the parameters of a convolutional layer with a 2D Kernel is:

$$p = (n_k^2 + 1) * c_{in} * c_{out} \tag{1.3}$$

p	number of parameters
n_k	size of kernel/filter
c_{in}	number of input channels
c_{out}	number of output channels

The +1 is for the bias.

$p_{layer1} = (3^1 + 1) * 1 * 8 = 80$
$p_{layer2} = (3^1 + 1) * 8 * 16 = 1280$
$p_{layer3} = (3^1 + 1) * 16 * 32 = 5760$

The last layer is a fully connected layer with 36 input nodes and 10 output nodes. This layer has 36*10 weights plus another 10 for the bias. As a result, we have 370 params for the last layer.
In total we have 80+1280+5760+370 = 7490 parameters.

1.7 3.6

As one can see we have a high accuracy represented by the high values on the diagonal and low values on the off diagonal entries. This means we have a high accuracy in general. When looking at specific digits we can see that the digit 2 is often classified as a 7.

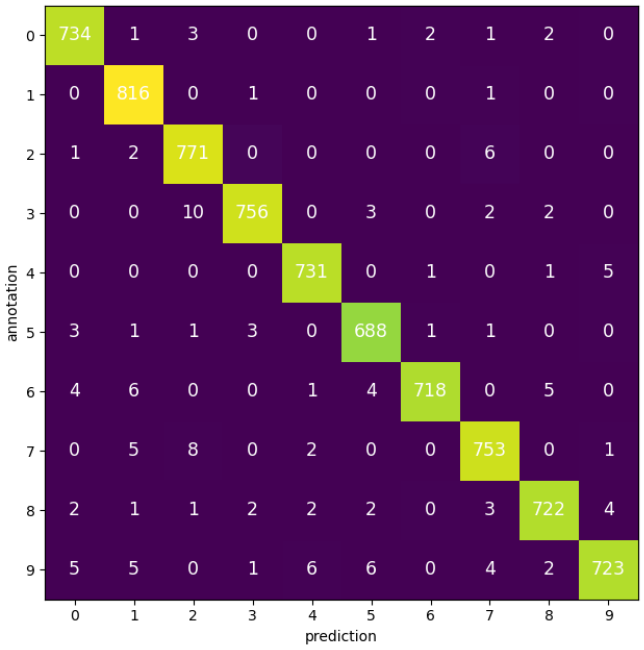

Figure 1.1: Confusion Matrix