

PROTOCOL EXERCISE 1 DEEP LEARNING LAB 2018/2019

Max Fuchs (Matrikel-Nr.:4340529)

This protocol will shortly present and evaluate the results from exercise 1.
The assignment was to build a neural network and evaluate the results for different parameter update methods and parameter settings.

The network design:

Input layer:

Number of units: number of samples

Hidden layer 1:

Number of units: 400

Standard deviation: 0.01

Activation function: Relu

Hidden layer 2:

Number of units: 10

Standard deviation: 0.01

Activation function: Linear

SoftMax output layer

This design was chosen as it performed quite good on the training data.

In the second step the training and validation set were used together for training and the neural net was then tested on the test data set

On the given network layout, two parameter update methods were tested.

In gradient descent the whole dataset is fed to the input layer at once, and the parameters get updated after every epoch. With the given network topology and a learning rate of 1 the test error could be decreased to about 7 %. As a convergence usually came after about 100 epochs no further optimizations have been done on the gradient descent as it took too much time to compute the different parameter sets.

In stochastic gradient descent the dataset is split into batches that are fed to the network one after another. The parameters are updated after every batch.

Several measurements with differing learning rates and batch sizes have been tested on the data.

In general, it could be seen that small learning rates took longer to converge as bigger ones, but too big rates were likely to miss a minimum and did not converge. Whereas small batch sizes did often not converge and too big batch sizes tend to take very long until they find a minimum. Here as well a good trade-off had to be found.

The final parameters for stochastic gradient descent were

Learning rate: 0.09

Batch size: 120

Network topology as described in the beginning of the report.

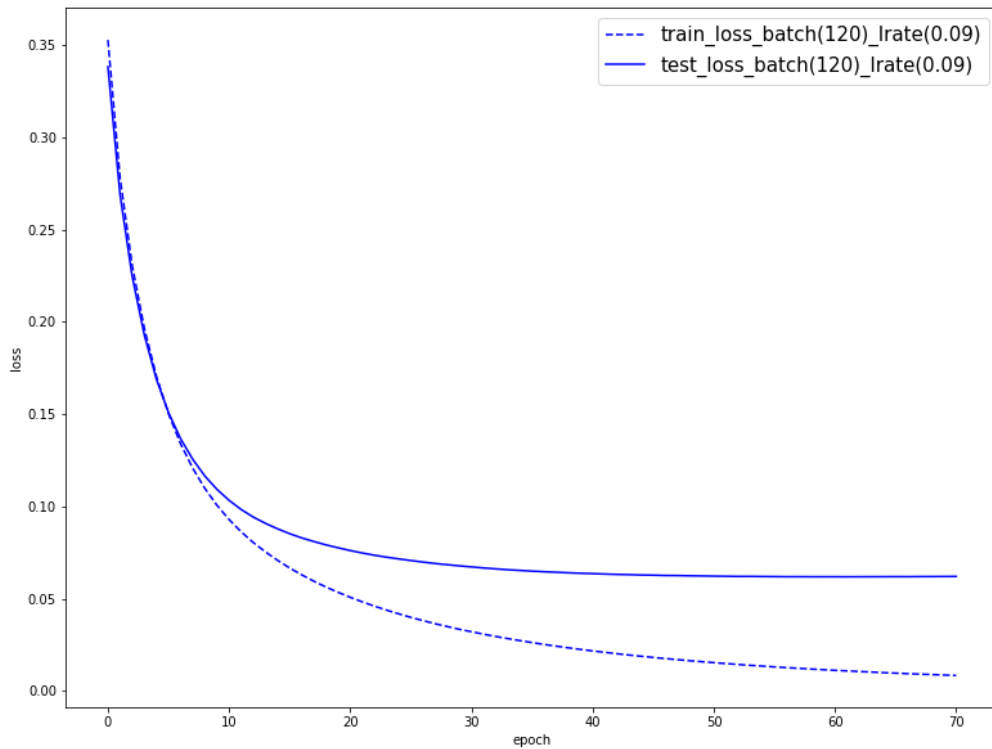


Figure 1: Training and test loss over epochs

Figure 1 shows the result for the final parameters. After about 30 epochs the test loss decreases much slower than the test loss. This is because further training will more precisely fit the parameters for the training data. And do not generalize as well for the test data as in earlier epochs.

The final achieved test error was: 1,73 %