Deep Learning Lab, Report for Submission 2

## Exercise 2, Learning Rates
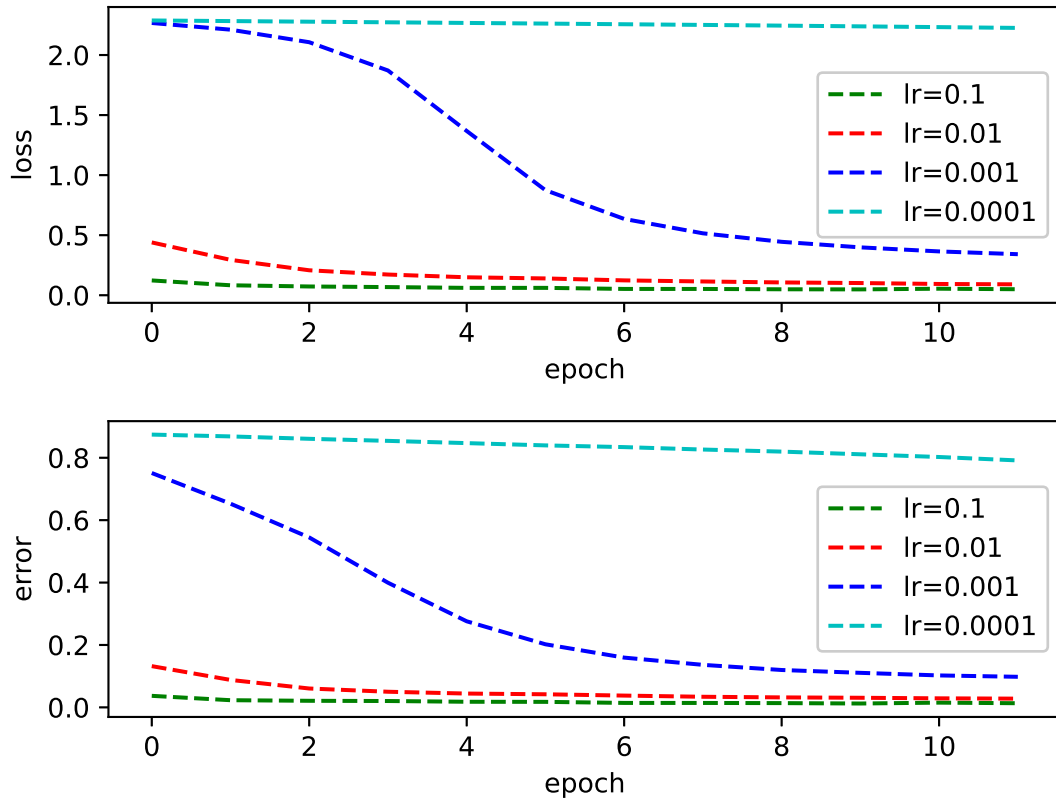
From that figure we can observe that the bigger learning rates lead to better results for 12 epochs. The two lower values (0.0001) also help with decreasing the loss and the error but lead to too small changes in the parameters such that the model is far from convergence after 12 epochs. With the two higher values (0.01 and 0.1) the model has mostly converged to a local minimum after 12 epochs and thus is finished with the training. With higher learning rates the performance again could get worse if the steps are too big and we jump out of a locally optimal region. However we can not observe this behavior in this experiment. Our Random search from exercise 4 suggests that the optimal learning rate (with SGD and no annealing schedule) lies slightly below 0.1 for this scenario.

Exercise 3, Filter Sizes
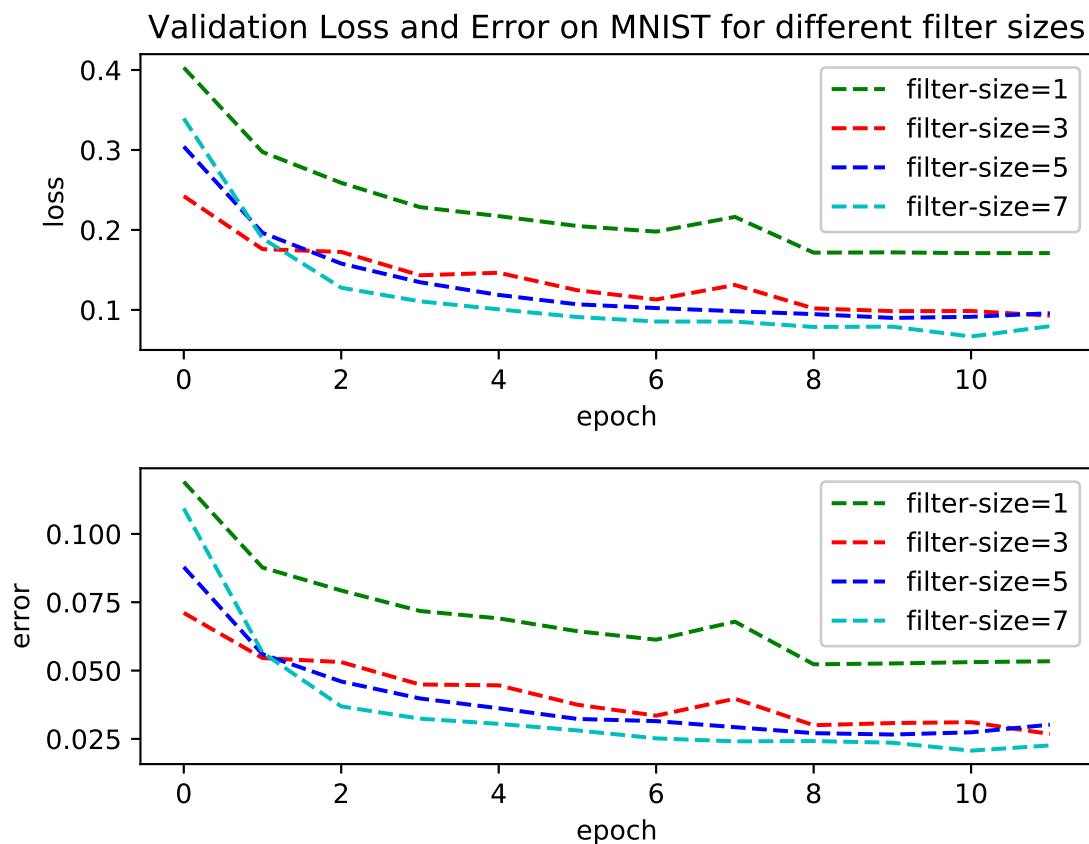
This picture was created with a batch size of 128 and a learning rate of 0.03. It suggests that bigger filters perform better on the mnist dataset. However the best performance from the random search in ex. 4 was achieved by a 3x3 filter. I would assume that smaller filters can perform better if the input resolution is lower, s.t the dependency or connection between neighboring pixels is of a short range. But this is just an assumption.

Exercise 4, Random Search

The best configuration that was found during my experiments with the given settings was:

- learning rate: 0.08847

- batch size: 46

- filter size: 3x3

- number of filters per layer: 63

After training this setup for the full 12 epochs it achieved an error rate on the test set of **0.98%** which is much better than what I could achieve with the feed forward networks from two weeks ago. The following plot shows the validation error of the currently found best configuration of random search (budget of 6 epochs)
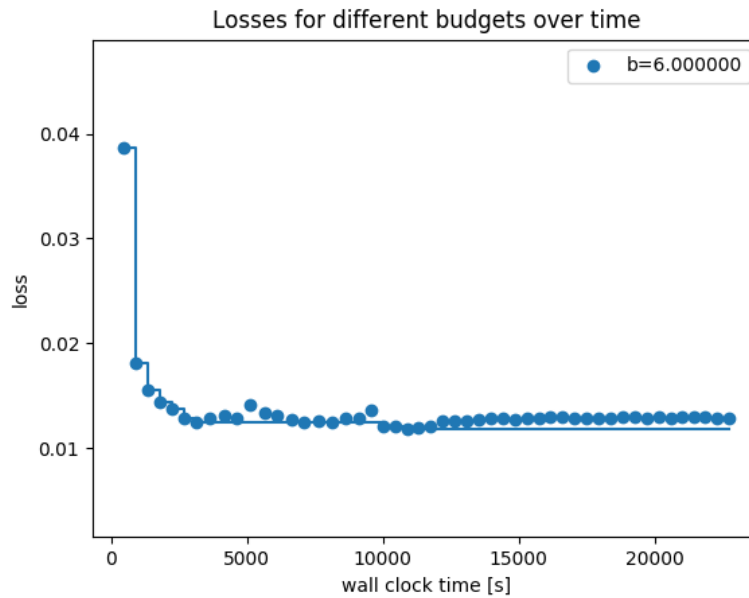
And the following graph shows the performance on the validation set of the best model during 12 epochs of training.

Validation Loss and Error on Best Model