

Sheet 1 (Code and theory: See https://github.com/itzeck/deep_learning)

Modeling

We have tried a pretrained ResNet50 model but it took very long to train due to us not having access to a powerful GPU. Therefore we build a very simple model ourselves. It consists of 2 convolutional layers with kernel size 3x3, max pooling with kernel size 3x3 and ReLU functions. And 4 fully connected layers with a ReLU activation functions and a LogSoftmax at the end.

We've also added early stopping because when using Adam we reached a very high test accuracy very quickly and we wanted to prevent overfitting. Therefore we stop after three consecutive epochs where the test_loss doesn't reach a new minimum.

Furthermore we normalized the input images to have values between -1 and 1.

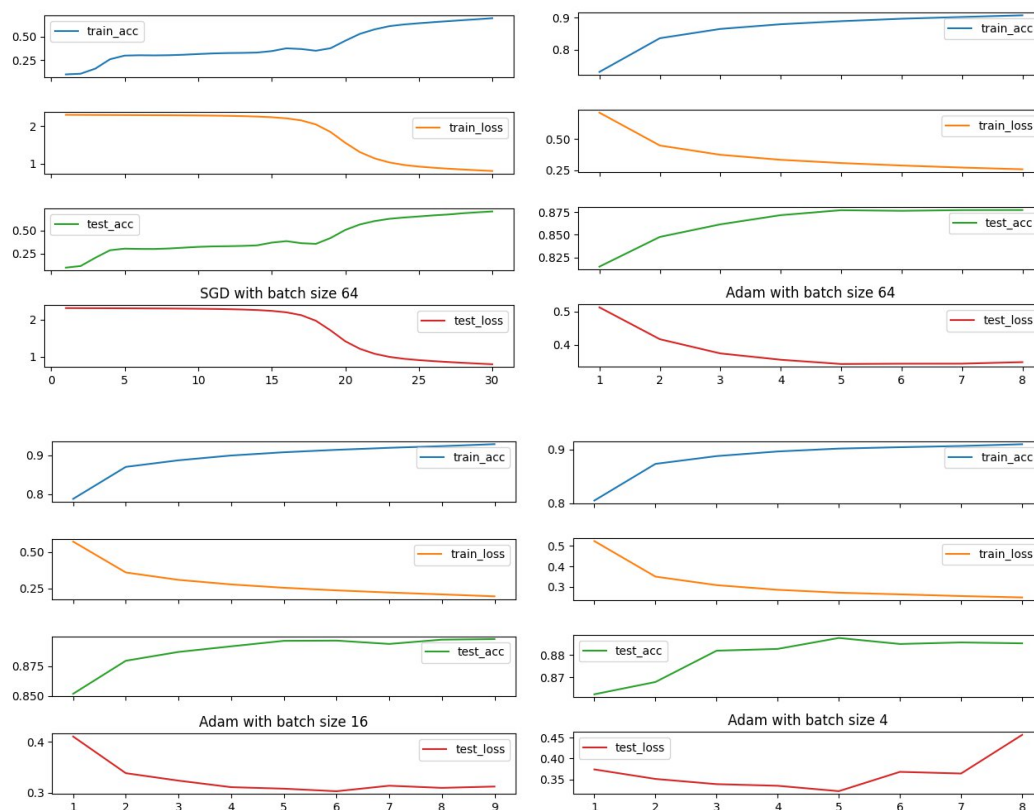
Training Parameters

We used a learning rate of 0.001 and up to 30 epochs.

Training

Using Adam early stopping was triggered after 7 epochs at most. Using SGD we trained for the whole 30 epochs. We used batch sizes of 2, 16 and 64.

Results



Interpretation

Using SGD we did not reach the same accuracy / loss as when using Adam even after 30 epochs. Also when using Adam the model started due to the weight initialization of Adam with very high accuracy. Therefore the early stopping was triggered for all batch sizes. SGD however started having approx. 10% accuracy. Between the different batch sizes there is not too much difference. A batch size of 16 seems to have worked best, however the differences in the metrics are marginal.