

Neural Networks Task 5

Lassi, Lehtonen 11805955
Group partner: Jens, Mellberg, 11806523

January 20, 2019

1 Training an LSTM with one hidden layer

We begin the training with a learning rate of 0.0001 and batch size 40. We find that adjusting the batch size does not improve the performance of our network, so we let the batch size remain 40. However, increasing the learning rate by small increments seems to improve the performance and decrease the convergence time. As seen in the plots, the convergence time decreases and we find our optimal performance when the learning rate is 0.00035. Increasing the learning rate more, the performance of our network will decrease. We conclude that the optimal learning rate for our network is 0.00035 with the batch size of 40.

2 Training the same LSTM, but now with ADAM

Our training begins with a learning rate of 0.0001 and default values for other parameters. As seen in the plots, it requires 45 epochs for convergence. We proceed by increasing the learning rate gradually. The evolution of the training/validation error is depicted in the provided plots. We reach a good performance at learning rate 0.0025. We perform our ten learning runs with these parameters, the results are shown in the table below.

Increasing the learning rate obviously improves performances because the network will take bigger steps when decreasing the error.

Run	Epochs
1	2
2	2
3	2
4	2
5	1
6	2
7	2
8	2
9	2
10	2
Mean	1.9
STD	0.316228

Figure 1: Gradient descent with learning rate 0.0001

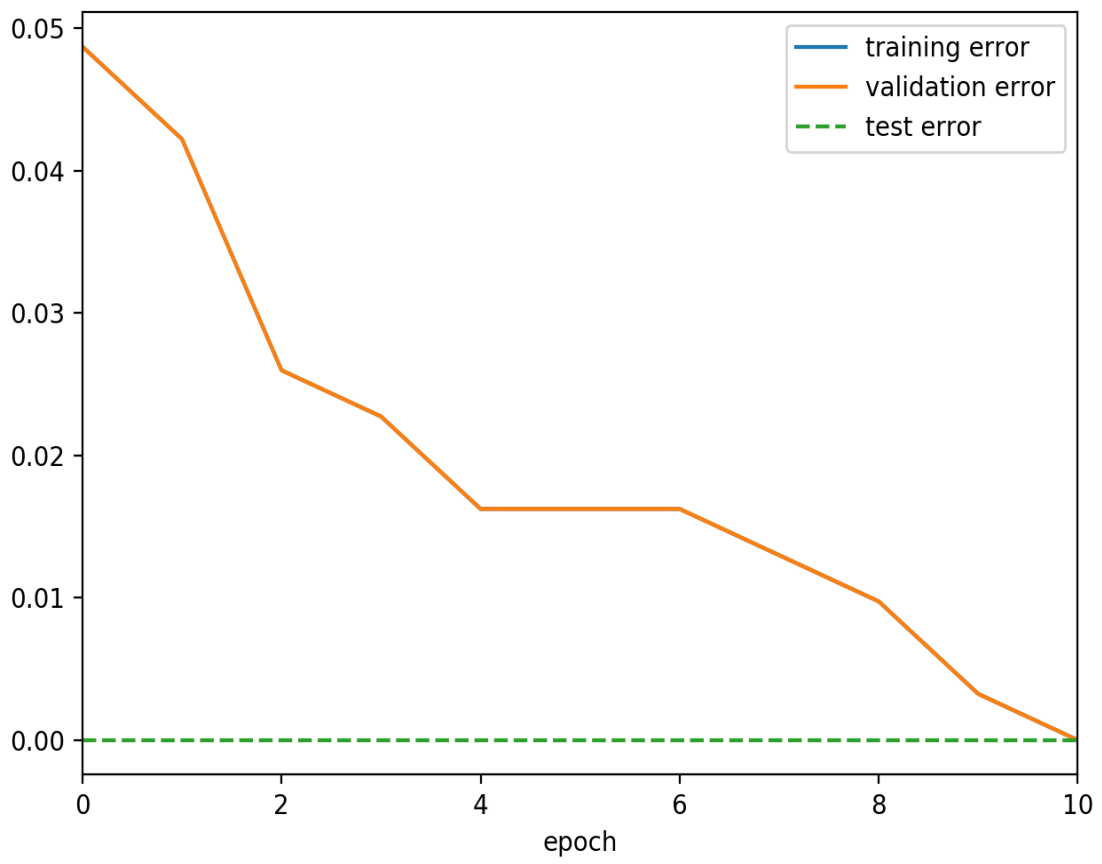


Figure 2: Gradient descent with learning rate 0.0002

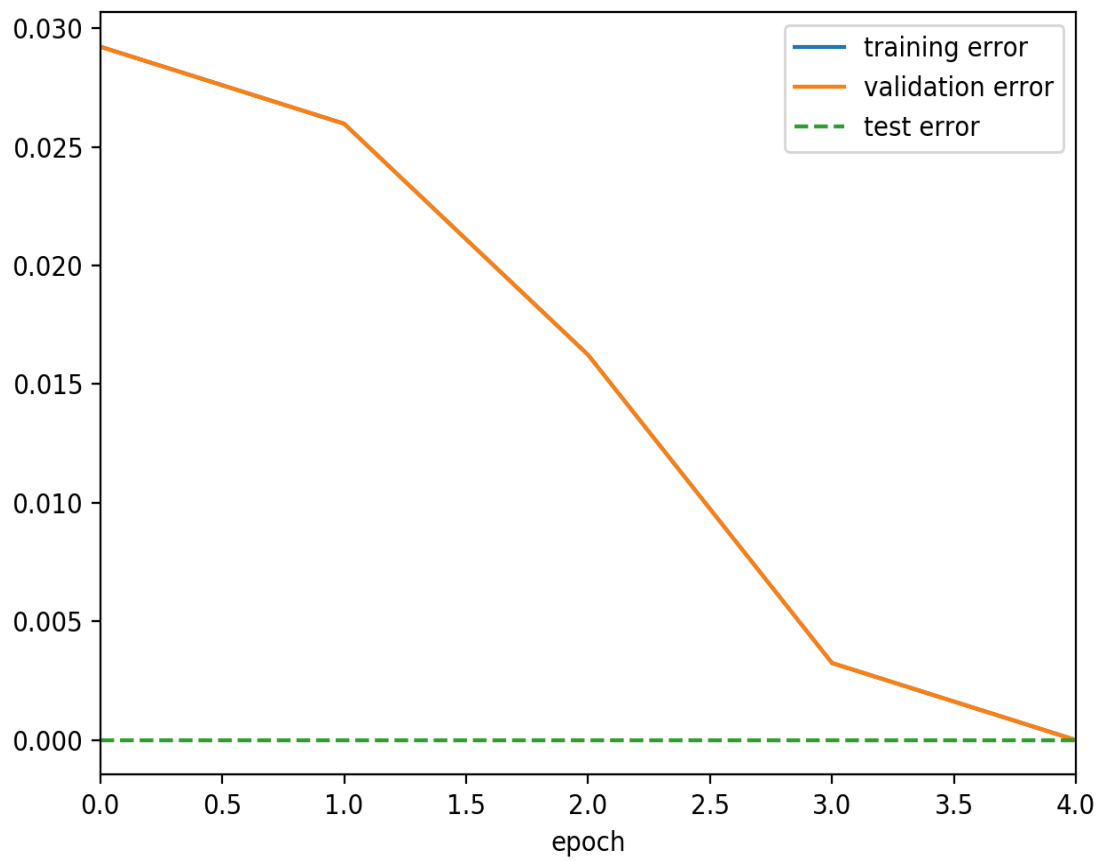


Figure 3: Gradient descent with learning rate 0.0003

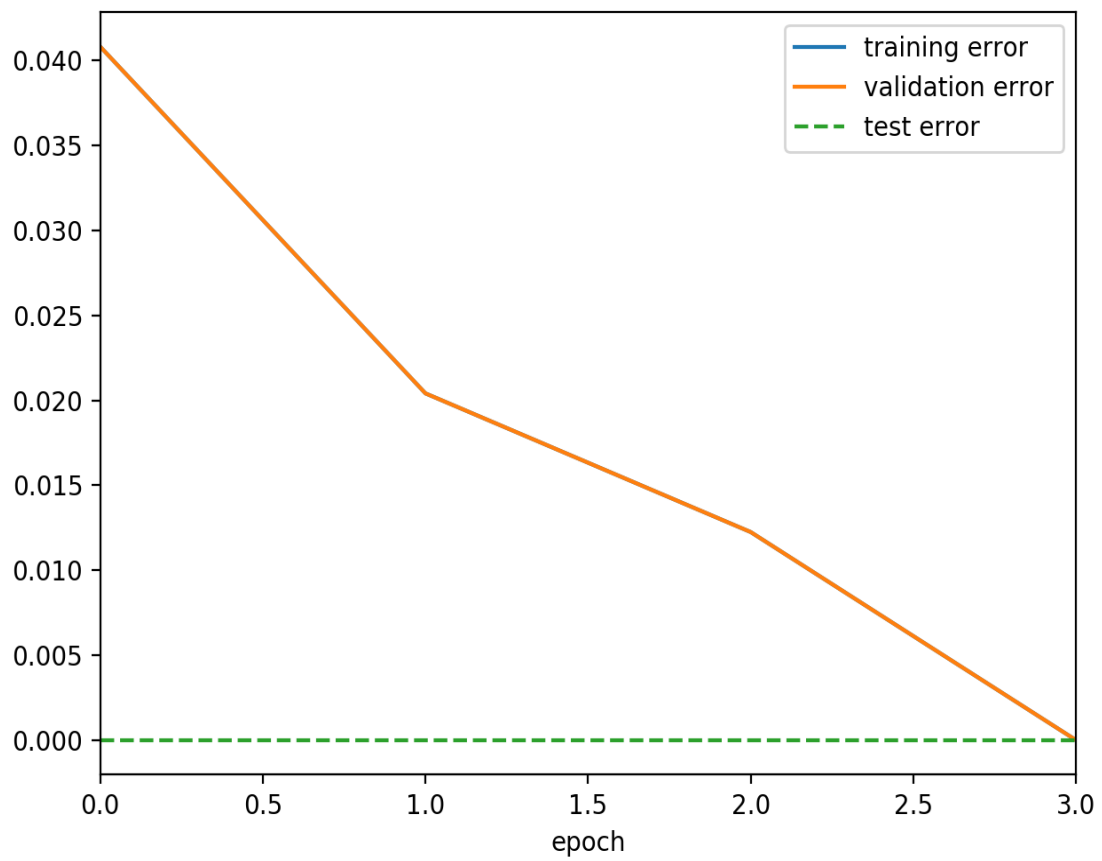


Figure 4: Gradient descent with learning rate 0.00035

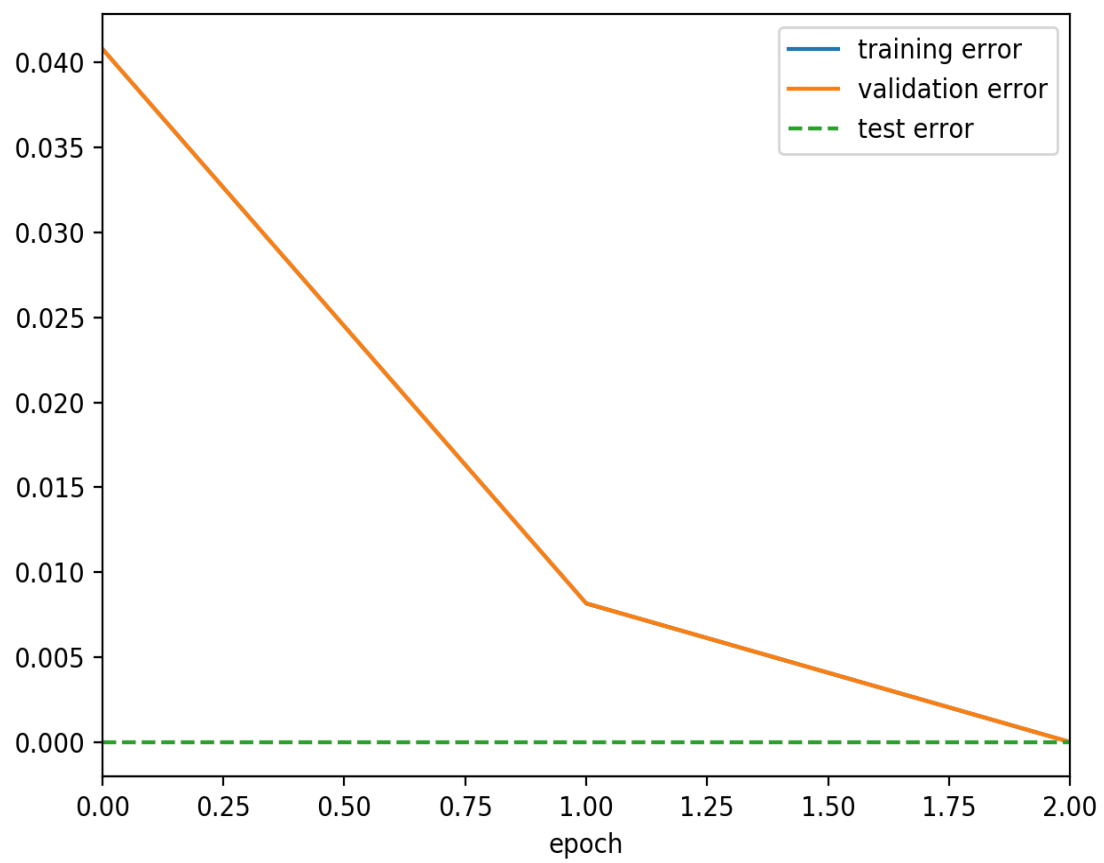


Figure 5: ADAM with learning rate 0.0001

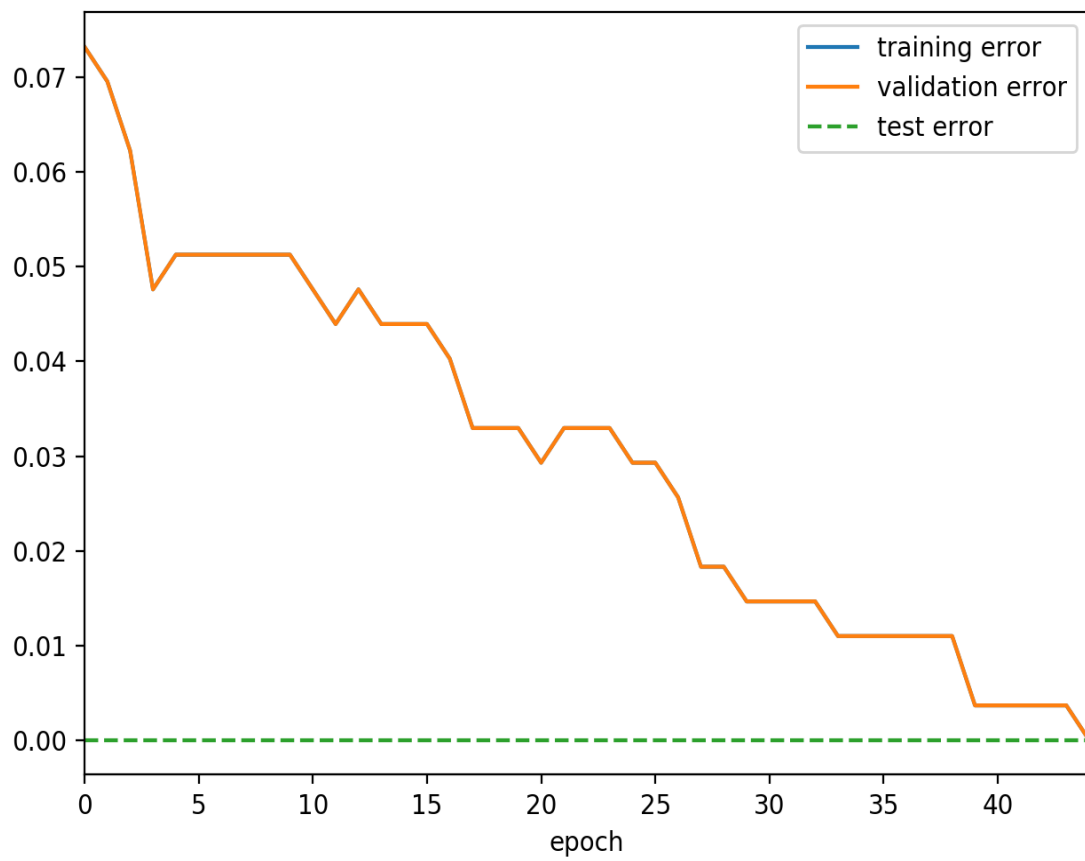


Figure 6: ADAM with learning rate 0.0002

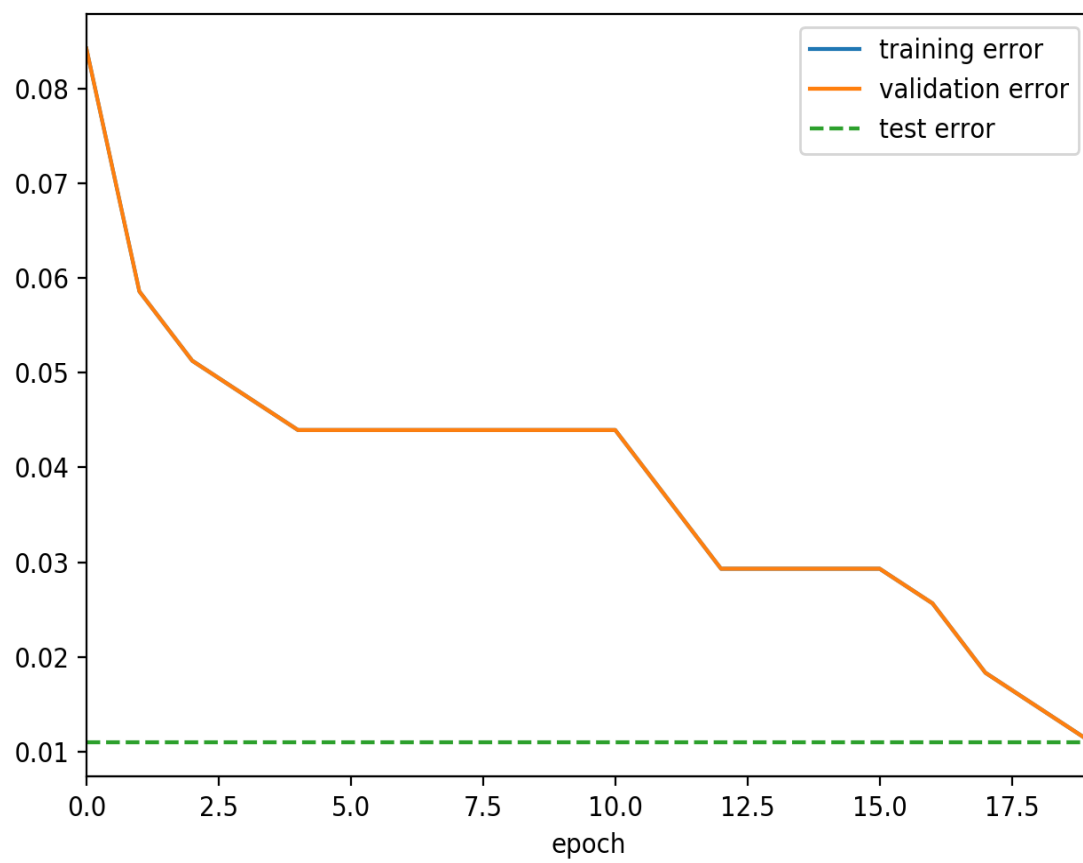


Figure 7: ADAM with learning rate 0.0004

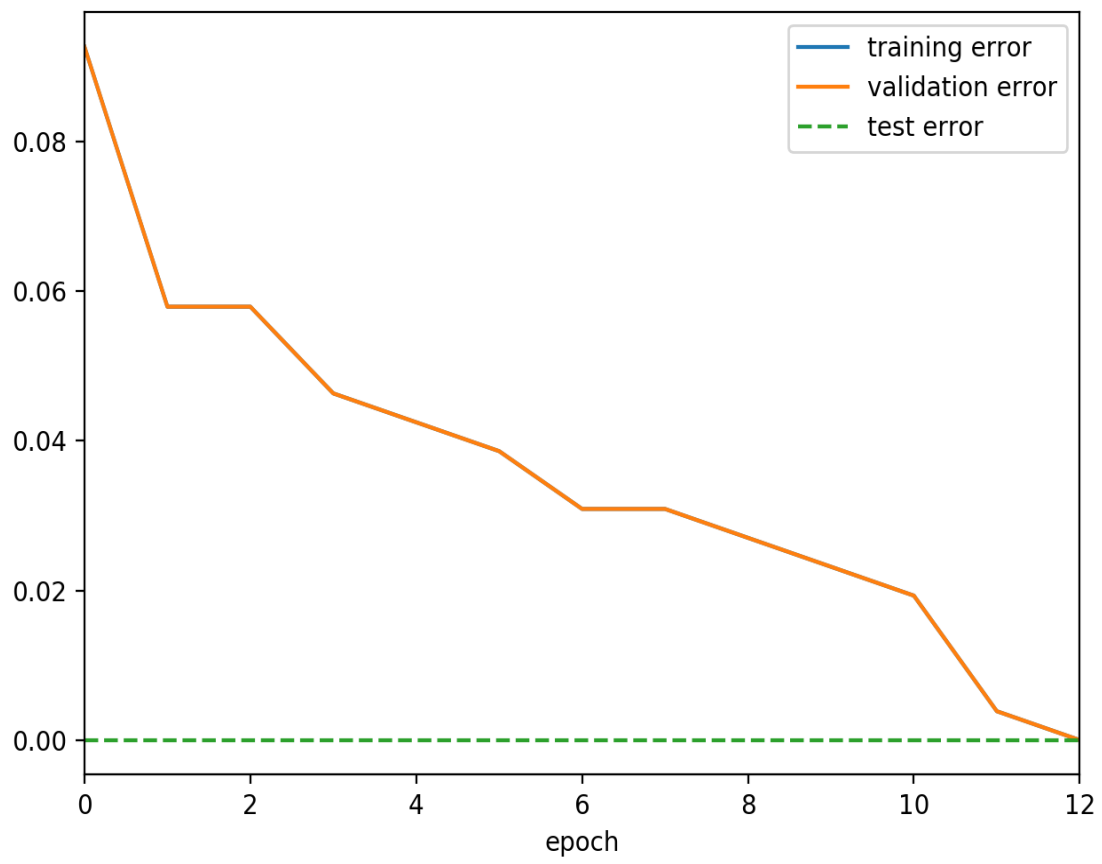


Figure 8: ADAM with learning rate 0.0006

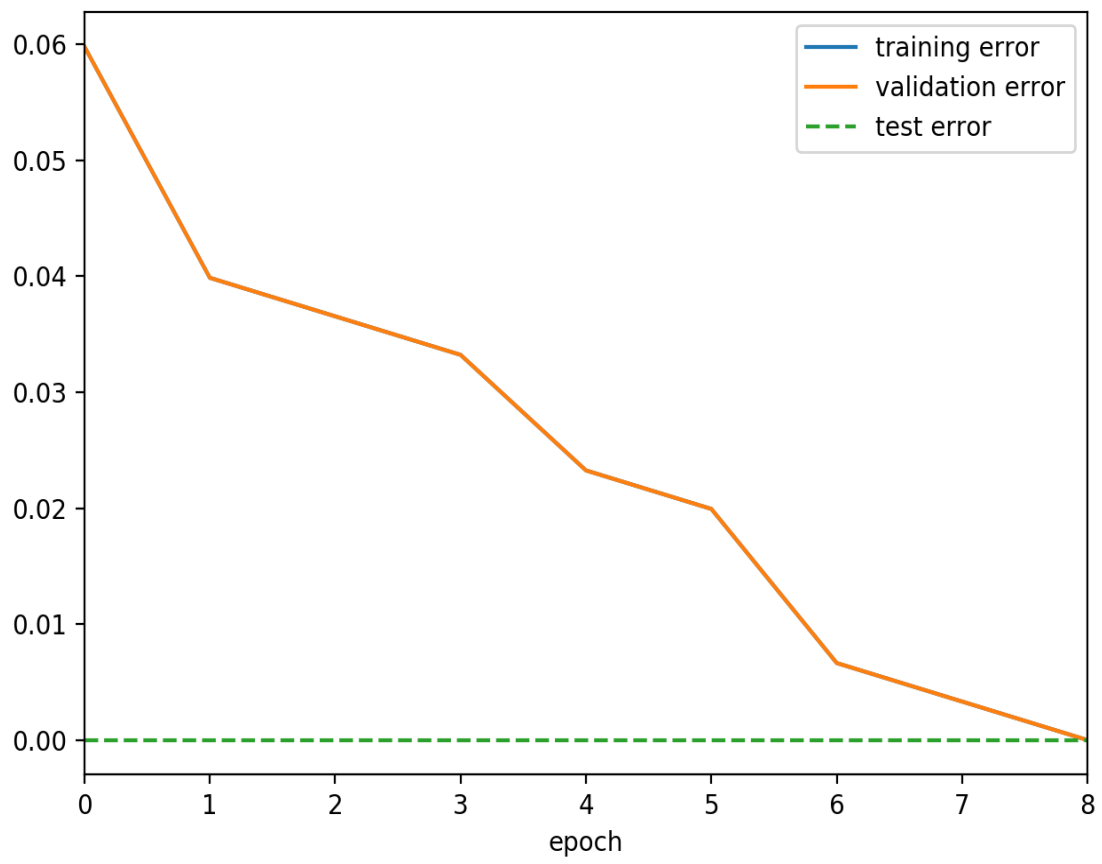


Figure 9: ADAM with learning rate 0.0009

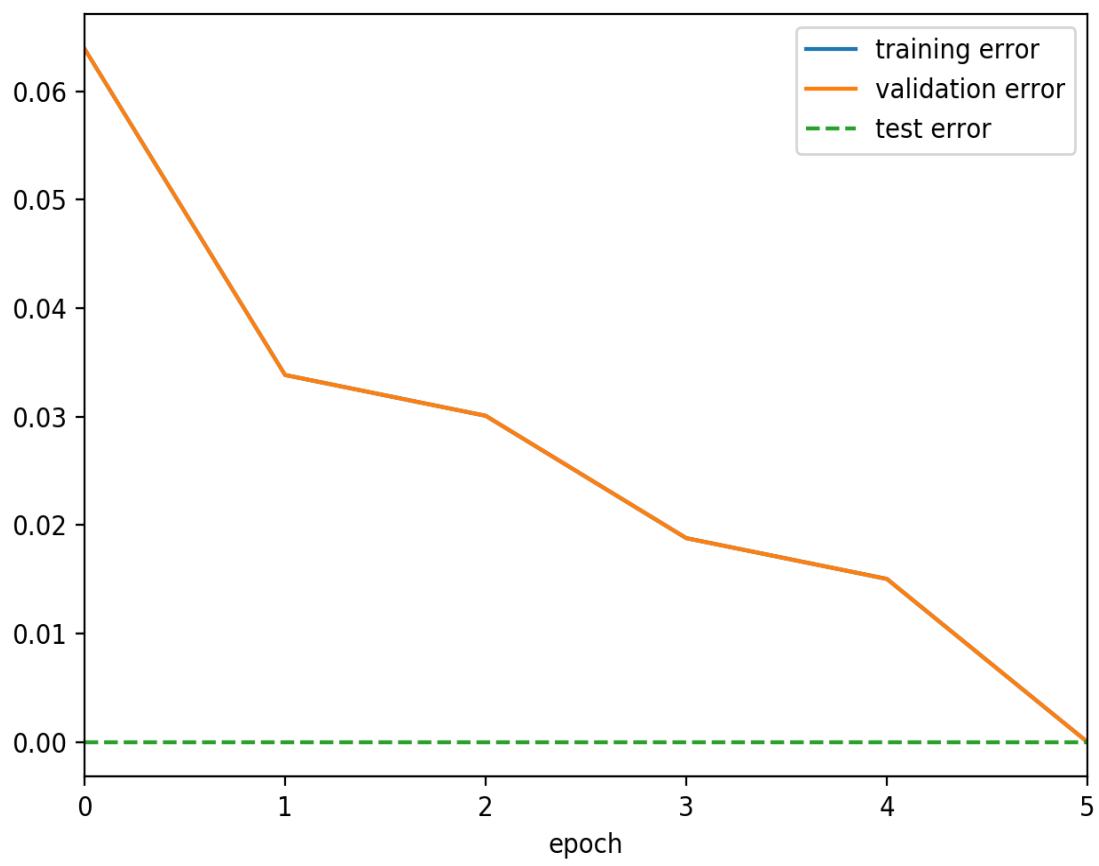


Figure 10: ADAM with learning rate 0.0015

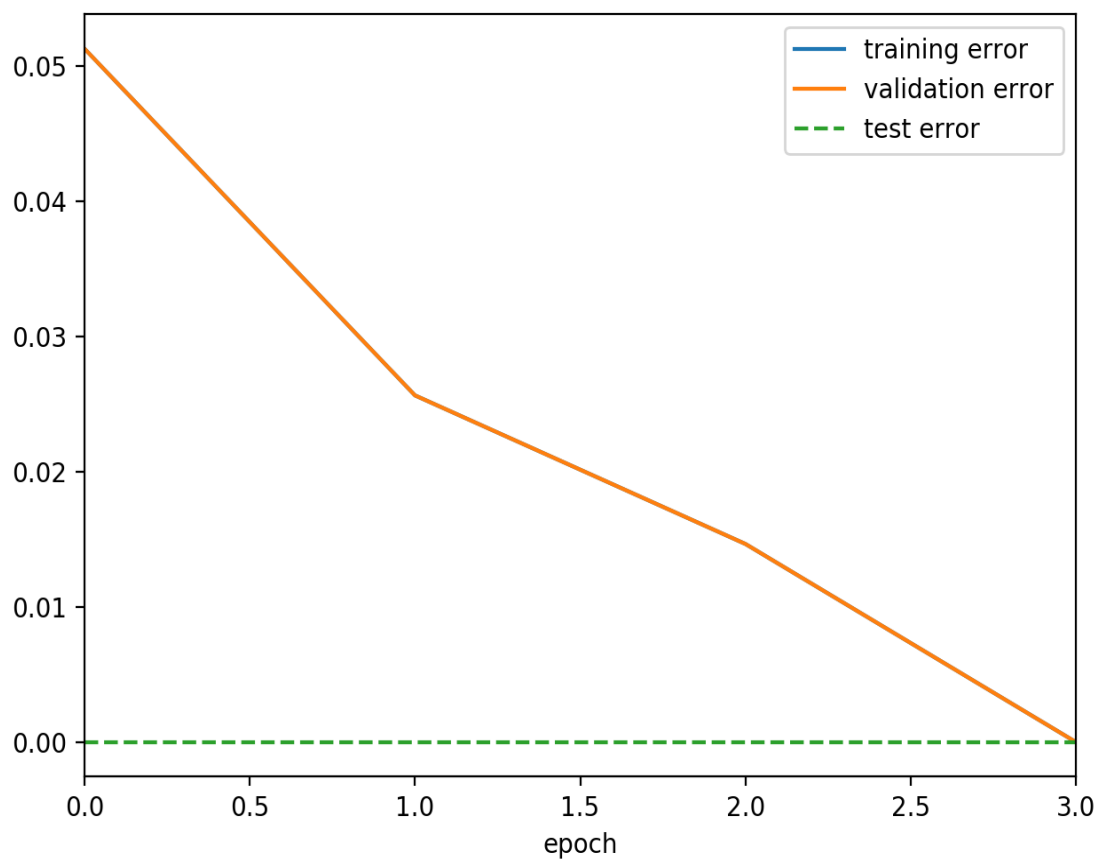


Figure 11: ADAM with learning rate 0.0025

