Deep Learning 2018 - Assignment 1

David Rau (11725184)
University of Amsterdam
`david.rau@student.uva.nl`

Vanilla RNN versus LSTM

Vanilla RNN in PyTorch

Question 1.1

Question 1.3

The plot that shows the accuracy versus palindrome length of the Vanilla RNN can be found in the answer of Question 1.6.

Question 1.4

Long-Short Term Network (LSTM) in PyTorch

Question 1.5

Question 1.6

For this exercise the long range dependencies of a simple Vanilla RNN and an LSTM models are probed. We train on the Palindrome task that is to predict the last character of a symmetric sequence of characters. The plot that shows the accuracy plotted against the sequence length of both models can be found in Fig. 1. Each of the models was trained with the default parameters (learning rate: 0.001, batch size: 128, hidden units: 128, max. steps: 10000) initially. As it can be seen in Fig. 1 the LSTM with the default learning rate outperforms the Vanilla RNN but the performance collapses from 100% to around 30% for sequence length 20. Knowing that LSTMs are, at least in theory, able to keep long distance relationships indefinitely long and that the gradient vanishes with an increasing sequence length, increasing the learning rate should boost the performance. The results show that the LSTM with learning rate 0.01 yields a perfect performance (accuracy 100%) for all sequence lengths, whereas the Vanilla RNN fails to predict the last character (accuracy $< 30\%$) from length 11 on.
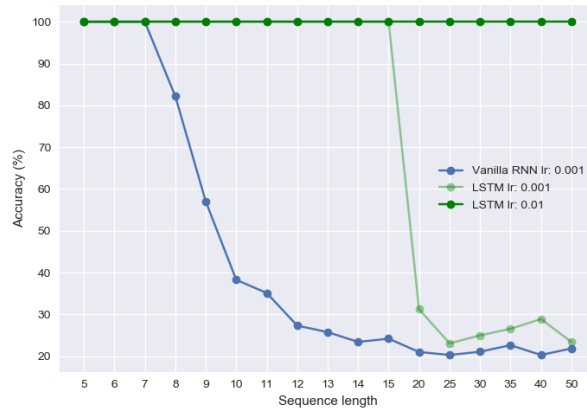
Figure 1: Accuracy (%) of the Vanilla RNN (blue) and LSTM (green) plotted against different Sequence lengths. The LSTM model was trained with two different learning rates 0.01 and 0.001.