Prof. Dr. Andreas Hotho, Albin Zehe, Jonas Kaiser
Data Science Chair

26./27.06.2024

8. Assignment in "Machine Learning for Natural Language Processing"

Summer Term 2024

1 General Questions

1. Describe two tasks that are better suited for RNN models than for MLP models and provide a short explanation.
2.
   - What are the problems in Machine Translation with simple RNNs that can be solved by Sequence to Sequence models?
   - What additional problems can be solved with the Attention Mechanism?
3. What are the key parts of the self-attention module in the transformer architecture? Explain each of these components.

2 Attention

Given a Sequence to Sequence model with an encoder and decoder RNN, that translates text from English to German. Apply the (Luong-)attention mechanism to the hidden states of the encoder RNN.

For the input "I love Language Processing", the encoder produces the following hidden states h_i:

$$h_1 = \begin{pmatrix}1\\0\\1\end{pmatrix}, h_2 = \begin{pmatrix}1\\2\\0\end{pmatrix}, h_3 = \begin{pmatrix}0\\2\\2\end{pmatrix}, h_4 = \begin{pmatrix}0\\2\\1\end{pmatrix}.$$

The current output of the decoder is "Ich", the decoder's hidden state is

$$s_1 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

Calculate the context vector c, which is used to predict the next word.

3 Recap: Backpropagation

Given the network depicted below, perform backpropagation to get the gradients for all weight matrices!

The input is the *row vector* $x = [1, 0.5, 0.75, 0.25]$, the target output is $t = 0$.

The initial weights are

$$W_1 = \begin{pmatrix} 0.1 & 0.2 & 0.6 & 0.9 \\ 0.7 & 0.5 & 0.3 & 0.6 \\ 0.5 & 0.75 & 0.6 & 0.3 \\ 1. & 0.1 & 0.1 & 0.2 \end{pmatrix}$$

$$W_2 = \begin{pmatrix} 0.6 & 0.6 & 0.6 & 0.2 \\ 1. & 0.3 & 0.1 & 0.1 \\ 0.2 & 0.5 & 0.1 & 0.7 \\ 0.75 & 0.3 & 0.5 & 0.9 \end{pmatrix}$$

$$W_3 = \begin{pmatrix} 0.6 \\ 0.7 \\ 0.2 \\ 0.9 \end{pmatrix}$$

No bias is used in the network. As loss function, we use the squared error, $L = \frac{1}{2}(t-y)^2$, where t is the true label.

Use the vectorised version of backpropagation introduced in the lecture.

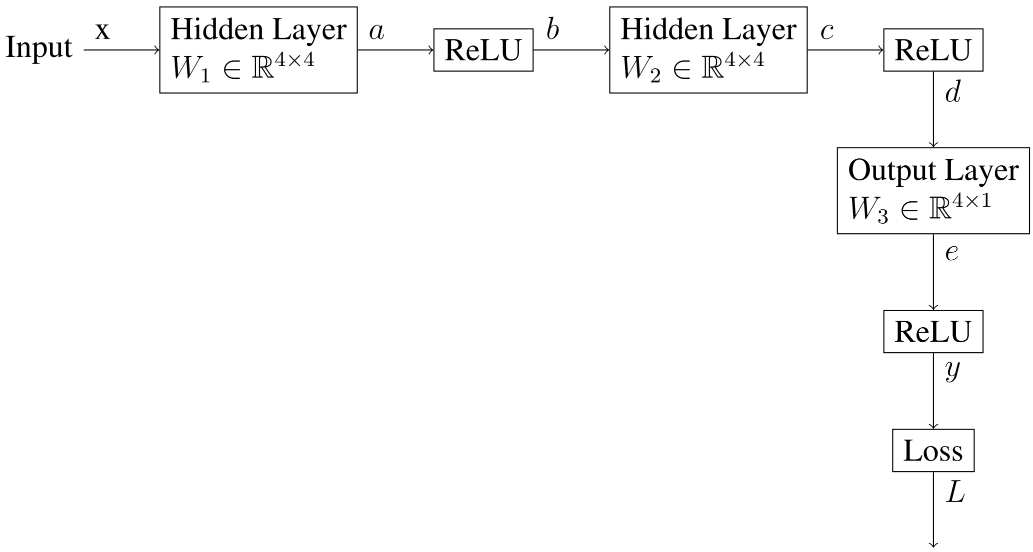

Figure 1: Neural Network with two hidden layers