Deep Learning

Theoretical Exercises – Week 13 – Chapter 10

Exercises on the book "Deep Learning" written by Ian Goodfellow,
Yoshua Bengio, and Aaron Courville.
Exercises and solutions by T. Méndez and G. Schuster

FS 2024

1 Exercises on Sequence Modeling: Recurrent and Recursive Nets

1. Given are the following update equations of a recurrent neural network

$$\begin{aligned}
\boldsymbol{h}_1^{(t)} &= \tanh(\boldsymbol{U}\,\boldsymbol{x}^{(t)} + \boldsymbol{V}\,\boldsymbol{h}_1^{(t-1)} + \boldsymbol{W}\,\boldsymbol{h}_2^{(t-1)} + \boldsymbol{b}) \\
\boldsymbol{h}_2^{(t)} &= \tanh(\boldsymbol{R}\,\boldsymbol{h}_1^{(t)} + \boldsymbol{S}\,\boldsymbol{h}_2^{(t-1)} + \boldsymbol{T}\,\boldsymbol{y}^{(t-1)} + \boldsymbol{c}) \\
\boldsymbol{y}^{(t)} &= \text{softmax}(\boldsymbol{P}\,\boldsymbol{x}^{(t)} + \boldsymbol{Q}\,\boldsymbol{h}_2^{(t)} + \boldsymbol{d}).
\end{aligned}$$

   (a) Draw the unfolded (unrolled) recurrent neural network that belongs to these equations.

   (b) Draw the folded recurrent neural network that belongs to these equations.

   Solution:

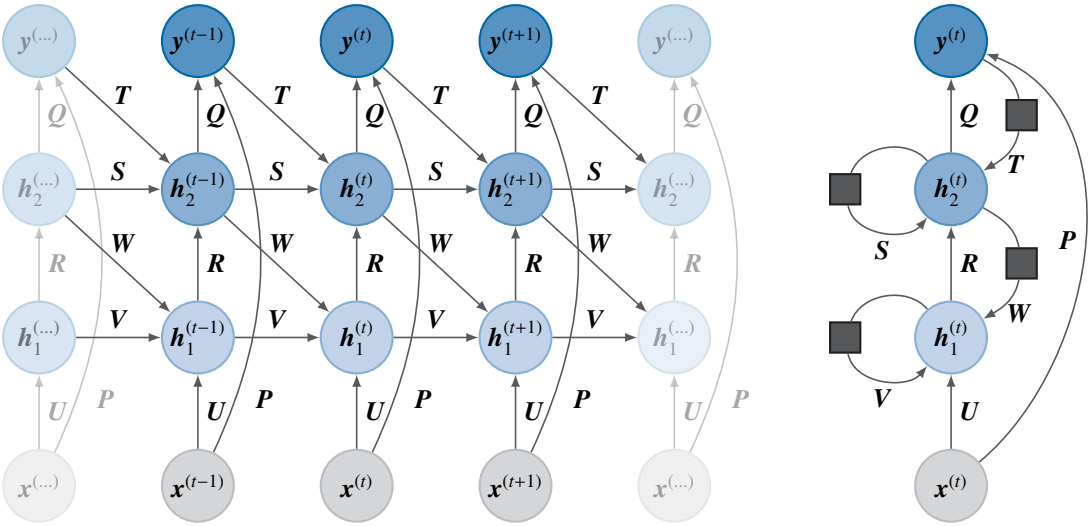
2. Explain the function of the different gates of the long short-term memory (LSTM).

   Solution:

   Forget Gate:
   The forget gate decides, what information is going to be thrown away from the cell state and what information is kept, by outputting a number between 0 and 1 for each element in the cell state. A 1 means that the information is completely preserved, while 0 means that the information is completely deleted.

   Input Gate:
   The input gate decides, which elements in the cell state are going to be updated, again by outputting a number between 0 and 1.

   Output-Gate:
   The output gate decides which parts of the cell state are going to be outputted, also by outputting a number between 0 and 1.



(a) Unfolded RNN (b) Folded RNN

3. Explain how the gated recurrent unit (GRU) differs from the (LSTM).

Solution:

The main difference to the LSTM is that in the GRU the forget gate and the input gate are combined into a single update gate, which simultaneously controls the forgetting factor and the decision to update the state unit. Thus, the decision what to forget and what to add is no longer made independently of each other. Only something will be forgotten if something new is added instead.

Also in the GRU the cell state and the output are the same. So instead of the output gate there is a reset gate which decides which parts of the cell state are used to compute the next cell state.