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

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

    (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.

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

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