Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNTI Winter Term 2024/2025

Exercise Sheet 11

Deadline: 29.01.2025 23:59

Guidelines: You are expected to work in a group of 2-3 students. While submitting the assignments, please make sure to include the following information for all our teammates in your PDF/python script:

Name:

Student ID (matriculation number):

Email:

Your submissions should be zipped as **Name1_id1_Name2_id2_Name3_id3.zip** when you have multiple files. For assignments where you are submitting a single file, use the **same naming convention** without creating a zip. For any clarification, please reach out to us on the **CMS Forum**.
Note that the above instructions are mandatory. If you are not following them, tutors can decide not to correct your exercise.

Exercise 11.1 - Sequence Modeling (0.75 + 0.75 points)

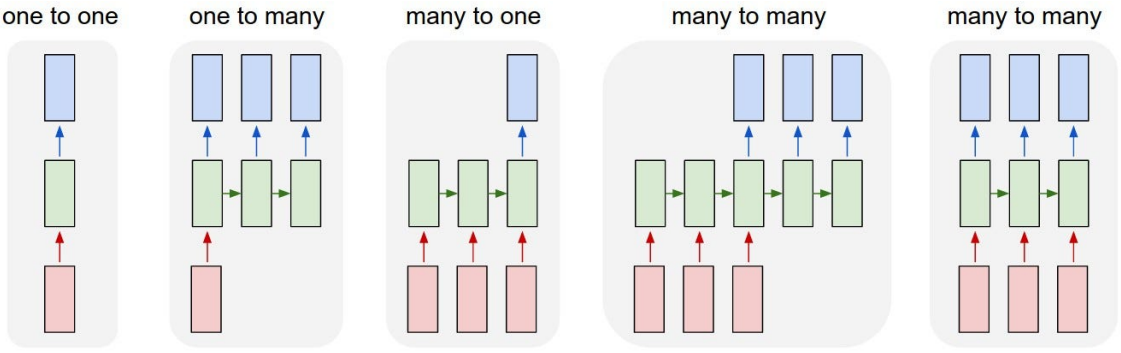
The aim of this exercise is to build upon the theory presented during the lecture on *RNNs*. Please limit your answers to 100 words for each question. Feel free to use any additional reading material for the same, but always remember to **cite your sources**.

1) What is the key difference between CNN and RNN in terms of the modelling of the problem? Reference Material: RNN, LSTM, and GRU

2) What is the mathematical difficulty in training the RNN, and how does LSTM mitigates the same? Assume that you use *Tanh* or *logistic sigmoid* activation functions. Does LSTM have similarities to residual neural networks? ResNet

Exercise 11.2 - Sequence Modeling (0.75+ 0.75 + (0.5 + 0.5) points)

Consider the following architectures (not all being RNNs):



1) Give an example application for each of the architectures above and describe its input and output. [max 2 sentences each].

2) A recursive NN is not in the drawings above. Draw the architecture of a recursive NN and describe it [3-5 sentences].

3) Consider sequences each of length T and a vanilla RNN model with the following operations:

$$\mathbf{z}_t = \mathbf{W}_t \mathbf{h}_{t-1} + \mathbf{U}_t \mathbf{x}_t + \mathbf{b}$$
$$\mathbf{h}_t = \sigma(\mathbf{z}_t)$$

Using the chain rule, we can arrive at:

$$\left|\left| \frac{\partial \mathcal{L}}{\partial \mathbf{h}_t} \right|\right| \quad \leq \quad \left|\left| \frac{\partial \mathcal{L}}{\partial \mathbf{h}_T} \right|\right| \, ||\mathbf{W}||^{T-t} \prod_{k=t}^{T-1} ||\mathbf{J}_{k+1}|| \tag{1}$$

a) What is $\mathbf{J}_{k+1}$ in the above equation? Give the expression.

b) What issues can occur because of equation (1)? Mention the two factors that cause the said issue.

Exercise 11.3 - Practical (6 points)

See the accompanying jupyter notebook.