



Exercise Sheet 10

NNTI Tutorial

Deadline: 18.01.2023 08:00

Exercise 10.1 - Recurrent Neural Networks

(0.5 + 0.5 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**.

- a) What is the key difference between CNN and RNN in terms of the modelling of the problem? Reference Material: [RNN, LSTM, and GRU](#)
- b) 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 10.2 - Transformers

(0.5 + 0.5 + 1.5 points)

Attention is all you need revolutionized the ML research space by providing similar architecture(s) for research in different modalities i.e. text, image. In this exercise, we will go through some foundational knowledge concerning Transformers. Please limit your answers to 100 words for each question.

- a) Transformers are used to model the sequential data inference. However, transformer actually would not have a recurrence (RNN and LSTM) nor convolution (CNN) to describe a relative positional information of the tokens (words). How do Transformers describe the positional information of the tokens at the input (and/or output) of the neural network architecture? Does this kind of encoding in Transformers offer a unique and deterministic mapping (between the input and output)?
- b) Should this encoding have trainable parameters or is it unnecessary? As an open-ended question, you are *encouraged* to bring up your own arguments and offer reasonable references for your answers.
- c) **Multi-Head Attention:** Explain the difference between Single-head attention and Multi-head attention and why latter performs better. [Paper](#) mentions that we can use single-head attention in specific settings to get performance boost over multi-head attention. Briefly explain what they are proposing and what is the reason for the performance boost.

please see the attached *ipynb* notebook

Submission instructions

The following instructions are mandatory. If you are not following them, tutors can decide to not correct your exercise.

- Please submit the assignment as a **team of two to three** students.
- Write the Microsoft Teams user name, student id and the name of each member of your team on your submission.
- Hand in zip file containing a **single** PDF with your solutions and the completed ipython notebook. Do not include any data or cache files (e.g. `__pycache__`).
- Important: please name the submitted zip folder and files inside using the format: **Name1_id1_Name2_id2**.
- Your assignment solution must be uploaded by only **one** of your team members to the 'Assignments' tab of the tutorial team (in **Microsoft Teams**). Please remember to press the **Hand In** button after uploading your work.
- If you have any trouble with the submission, contact your tutor **before** the deadline.