

## NEURAL NETWORKS – exercise 7

Exercise 7 will concern recurrent networks. Recurrent networks can be useful for sequential data, where the order of the data in time is important. The recursive layer accepts a sequence of vectors as an input and also returns a sequence of vectors (however, we are often interested in operating on the last vector of this sequence and, for example, torch by default returns both the full sequence of outputs and the last state of the layer). Regarding action, it is important that:

- The layer applies the same transformation to each step in the sequence
- in step  $n$ , the input is step  $n$  of the input sequence, and the output from step  $n-1$
- specific types of recursive layers (RNN/GRU/LSTM) differ in an additional transformation imposed on the output from the previous step

In addition to the recursive layers themselves, there are two important new features:

- Embedding layer, whose purpose is to convert values from a finite set into vectors. For language problems, a set can be, for example, the set of all words of a language. Such a layer is simply a matrix (number of words) per (vector dimension), from which, just like any other matrix within the network structure neuronal, we are able to calculate the gradient. And this means that we can optimize it, which ends up finding a vector for each word that somehow reflects the "meaning" (basically: connections between similar meanings, encoded in the similarity/distance of the vectors).
- Padding sequences to the same length. The texts in the IMDB set are of different lengths, so if we want to teach in batching mode, we need to pad the vectors to equal length with zeros.

Please use the IMDB review collection, available on torchtext and keras:

<https://keras.io/api/datasets/imdb/>

This time you can check:

- Type of selected recursive layer (RNN vs. LSTM)
- The dimension of the recursive layer
- The impact of trimming sequences to partial length (you should determine the maximum length when batching, and compare the result obtained in this way with the results obtained by trimming all texts to a predetermined value, e.g. 10, 20 words)

**To improve performance, you can work on a subset smaller than the full train.**

The exercise is graded on a scale of 0-10 points and you have until the end of the semester to complete it.