Introduction to Transformers

D. Gueorguiev 6/24/22

Table of Contents

[Preliminaries 1](#_Toc107049124)

[Language Modeling 1](#_Toc107049125)

[Sequential Modeling 1](#_Toc107049126)

[Notes on RNN (Recurrent Neural Networks) 1](#_Toc107049127)

[Notes on LSTM (Long Short Term Memory) 2](#_Toc107049128)

[Notes on Gated Recurrent Neural Nets 2](#_Toc107049129)

[Latent Semantic Analysis 2](#_Toc107049130)

# Preliminaries

*Note 0*: Knowledge of feed-forward neural networks and the back-propagation algorithm is assumed.

## Language Modeling

Tools for Language Modeling:

* Sequential Models
* Latent Semantic Analysis (LSA)

## Sequential Modeling

In general, we build ML models where we have data points which are usually uncorrelated, and no order relation can be imposed on them.

*Sequential models*: In many cases such as language, voice, and time-series data a data point is dependent on a set of other data points which have already been processed. We call such stream of data points a stream of *sequential data*. Machine learning models which accept input or create an output sequence of data points are known as *sequential models*.

Previous tools for sequence modeling:

* Recurrent Neural Networks
* Long Short Term Memory
* Gated Recurrent Neural Nets

### Notes on RNN (Recurrent Neural Networks)

What is RNN?

Neural network with at least one cycle. If the network contains cycle the computation is not uniquely defined by the interconnection pattern and the *temporal dimension* must be considered. When the output of one node is fed back to the same unit we are dealing with recursive computation. We must define what we expected from the network – is the fixed point of the recursive evaluation the desired result or one of the intermediate computations? We can assume that every computation takes a fixed amount of time and can be expressed as a certain number of time units.

Continuous vs Epochwise mode

Backpropagation through time (BPTT)

### Notes on LSTM (Long Short Term Memory)

### Notes on Gated Recurrent Neural Nets

## Latent Semantic Analysis (LSA)

LSA belongs to the family of *Probabilistic Methods* for analyzing and modeling the structure of the language.