# Introduction to Transformers

D. Gueorguiev 6/24/22

## Preliminary Notes

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

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

### 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 unit 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

Continuous vs Epochwise mode

Backpropagation through time (BPTT)

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