# Introduction to Transformers

D. Gueorguiev 6/24/22

## Preliminary Notes

In general, 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:

* Latent Semantic Analysis
* 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.

Continuous vs Epochwise mode

Backpropagation through time (BPTT)

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