LSTM

There are many posts out there about LSTMs, here are a few of my favorites:

- Chris Olah's LSTM post
- Edwin Chen's LSTM post
- Andrej Karpathy's lecture on RNNs and LSTMs from CS231n

The output of the *Learn Gate* is $N_t i_t$ where: (i_t is the ignore factor)

$$N_t = \tanh(W_n[STM_{t-1}, E_t] + b_n)$$

$$i_t = \sigma(W_i[STM_{t-1}, E_t] + b_i)$$

The output of the *Forget Gate* is $LTM_{t-1}f_t$ where: (f_t is the forget factor)

$$f_t = \sigma(W_f[STM_{t-1}, E_t] + b_f)$$

The output of the *Remember Gate* is:

$$LTM_{t-1}f_t + N_t i_t$$

The output of the *Use Gate* is $U_t V_t$ where:

$$U_t = \tanh(W_u LTM_{t-1}f_t + b_u)$$

$$V_t = \sigma(W_v[STM_{t-1}, E_t] + b_v)$$

If you would like to deepen your knowledge even more, go over the following tutorial. Focus on the overview titled: **Long Short-Term Memory Units (LSTMs)**.

Sequence batching in RNNs: https://www.youtube.com/watch?v=pdSr5F-9qE0

Something to note: here batch size means the size of each batch, NOT the number of batches. So if each batch has 2 sequences and 3 words. Next batch will have the next 3 words of the 2 sequences.

Information about **GRUs** can be found in the following links:

- Michael Guerzhoy's post

LSTM with Peephole connections: adding LTM to every connection: