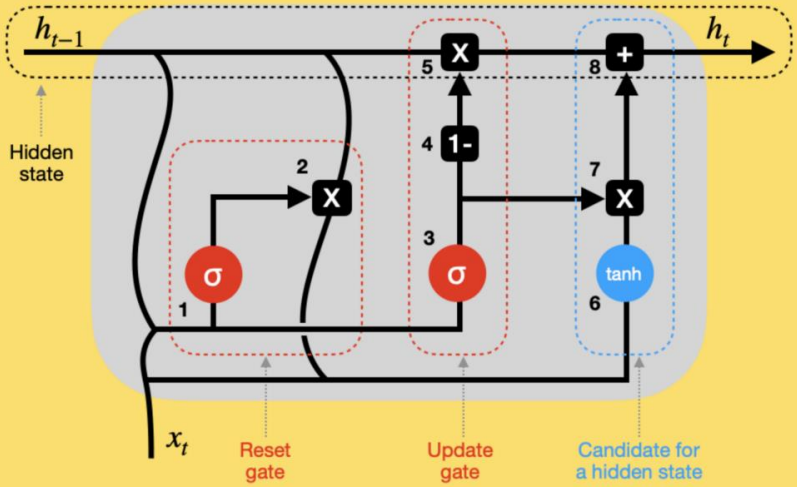
GATED RECURRENT UNIT (GRU)

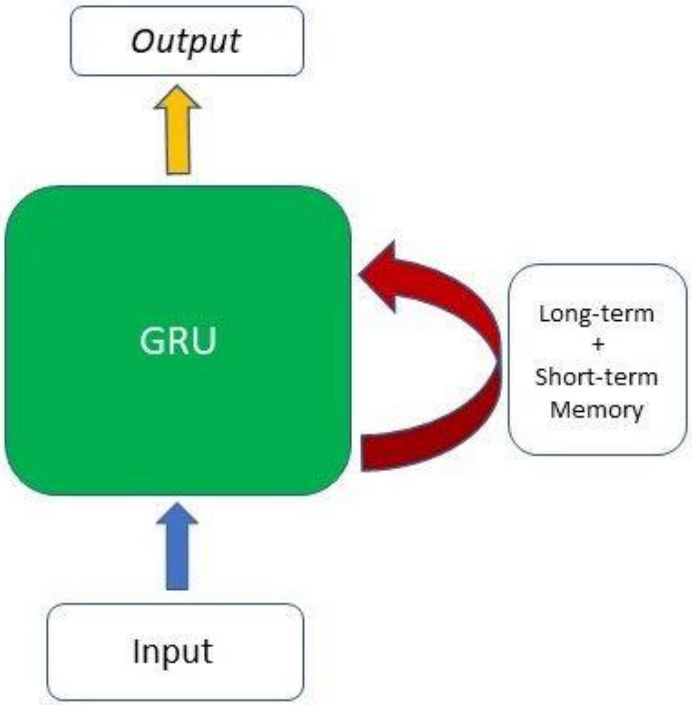


Dr. Amit Sinhal

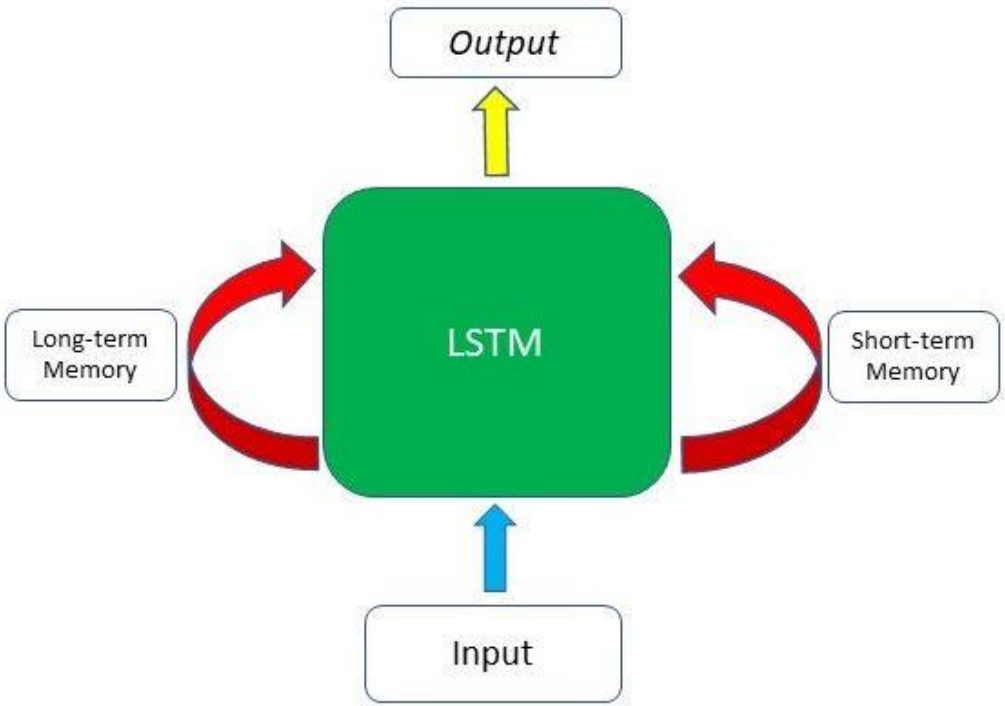
Professor

Department of Computer Science & Engineering

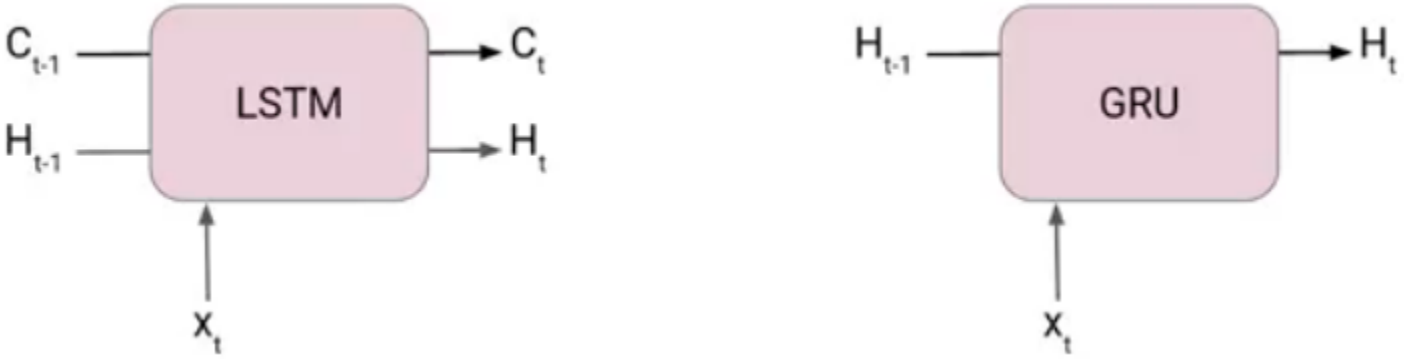
GRU Vs LSTM



VS



GRU Vs LSTM

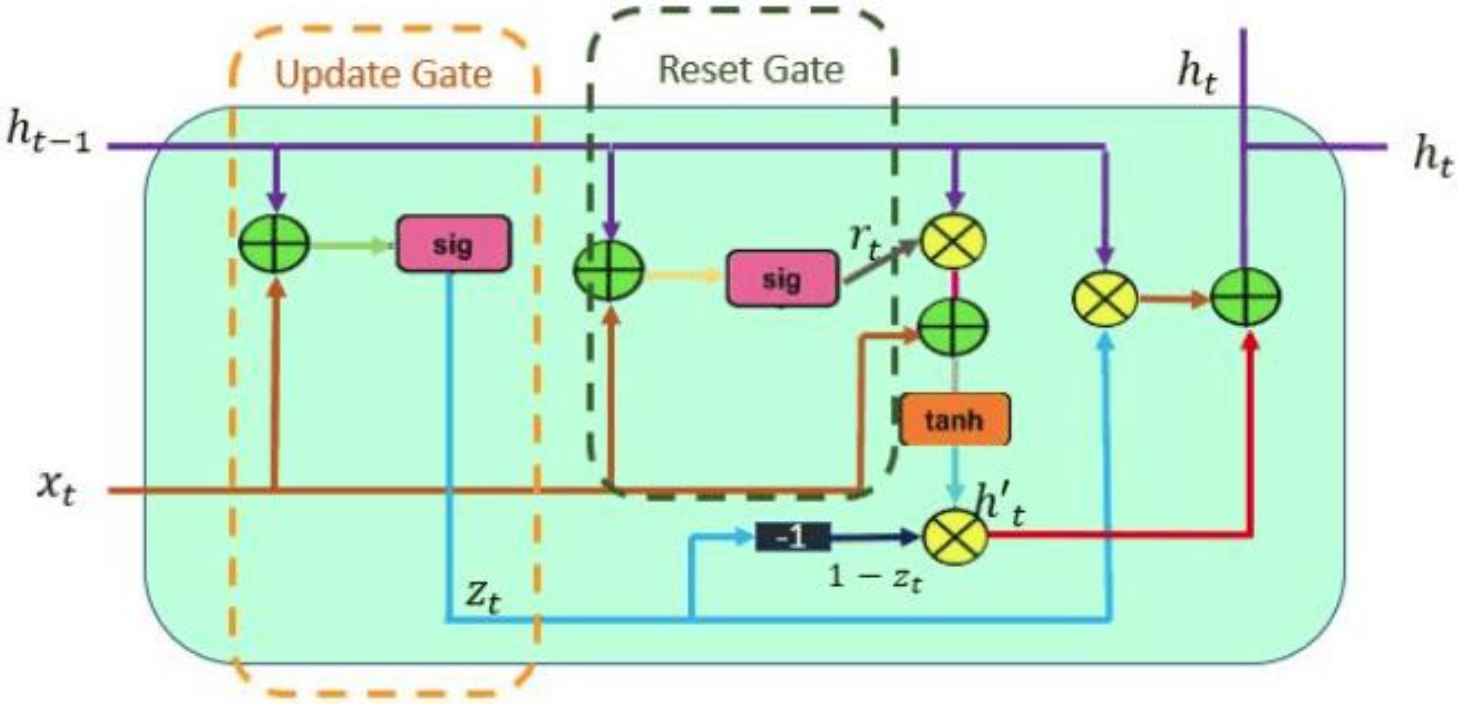


GRUs are very similar to Long Short Term Memory(LSTM). Just like LSTM, GRU uses gates to control the flow of information.

Another Interesting thing about GRU is that, unlike LSTM, it does not have a separate cell state (Ct). It only has a hidden state(Ht). Due to the simpler architecture, GRUs are faster to train.

What is Gated Recurrent Unit or GRU?

The workflow of the Gated Recurrent Unit, in short GRU, is the same as the RNN but the difference is in the operation and gates associated with each GRU unit. To solve the problem faced by standard RNN, GRU incorporates the two gate operating mechanisms called Update gate and Reset gate.



Update Gate

The update gate is responsible for determining the amount of previous information that needs to pass along the next state. This is really powerful because the model can decide to copy all the information from the past and eliminate the risk of vanishing gradient.

Similarly, we have an Update gate for long-term memory and the equation of the gate is shown below.

$$u_t = \sigma(x_t * U_u + H_{t-1} * W_u)$$

The only difference is of weight metrics i.e Uu and Wu.

Reset Gate

The reset gate is used from the model to decide how much of the past information is needed to neglect; in short, it decides whether the previous cell state is important or not.

The Reset Gate is responsible for the short-term memory of the network i.e the hidden state (Ht). Here is the equation of the Reset gate.

$$r_t = \sigma(x_t * U_r + H_{t-1} * W_r)$$

If you remember from the LSTM gate equation it is very similar to that. The value of **rt** will range from 0 to 1 because of the sigmoid function. Here Ur and Wr are weight matrices for the reset gate.

First, the reset gate comes into action it stores relevant information from the past time step into new memory content. Then it multiplies the input vector and hidden state with their weights. Next, it calculates element-wise multiplication between the reset gate and previously hidden state multiple. After summing up the above steps the non-linear activation function is applied and the next sequence is generated.

What is the difference between GRU & LSTM?

The few differencing points are as follows:

- The GRU has two gates, LSTM has three gates
- GRU does not possess any internal memory, they don't have an output gate that is present in LSTM
- In LSTM the input gate and target gate are coupled by an update gate and in GRU reset gate is applied directly to the previous hidden state.
- In LSTM the responsibility of reset gate is taken by the two gates i.e., input and target.