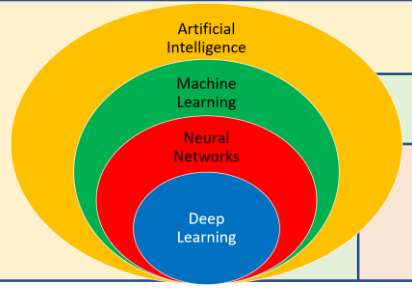


Artificial Intelligence (AI)

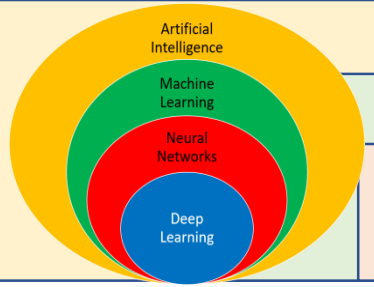
Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

Advanced Artificial Intelligence

Dr. Rastgoo



Artificial Intelligence (AI)

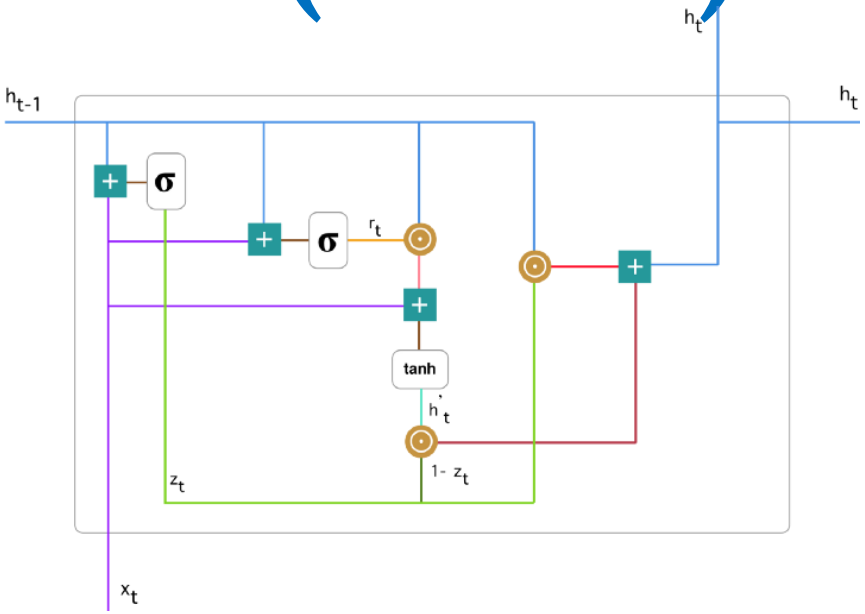
Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)

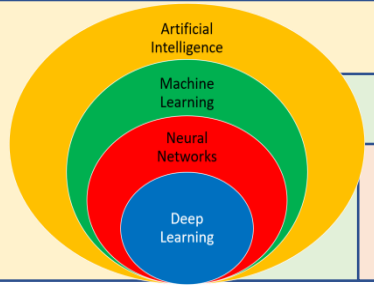
Sequential Models

Part 3: Gated Recurrent Unit (GRU)

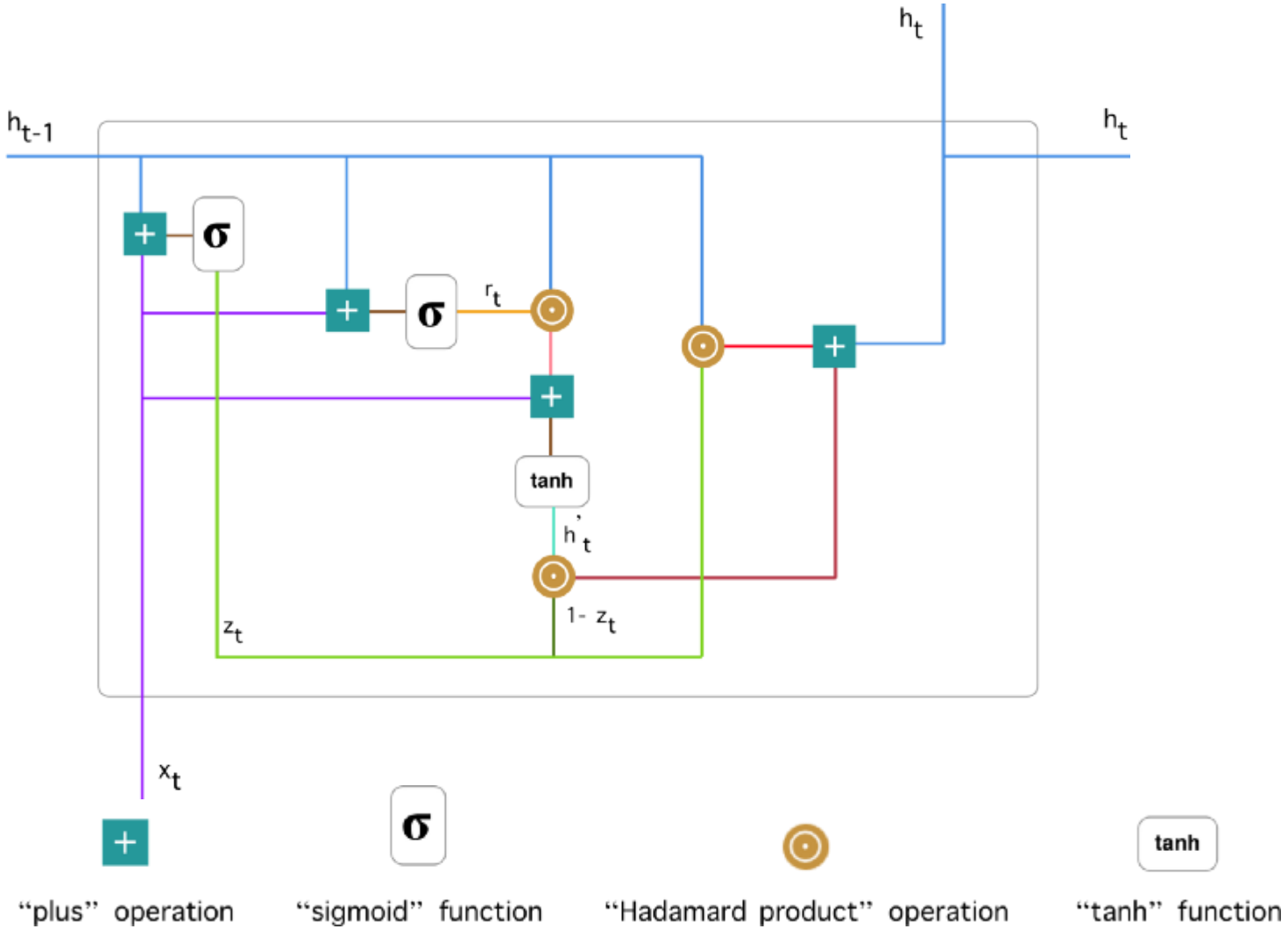


How do GRUs work?

- GRUs are improved version of standard RNN.
- To solve the vanishing gradient problem of a standard RNN, GRU uses, so-called, update gate and reset gate.
- Basically, these are two vectors which decide what information should be passed to the output.
- The special thing about them is that they can be trained to keep information from long ago, without washing it through time or remove information which is irrelevant to the prediction.

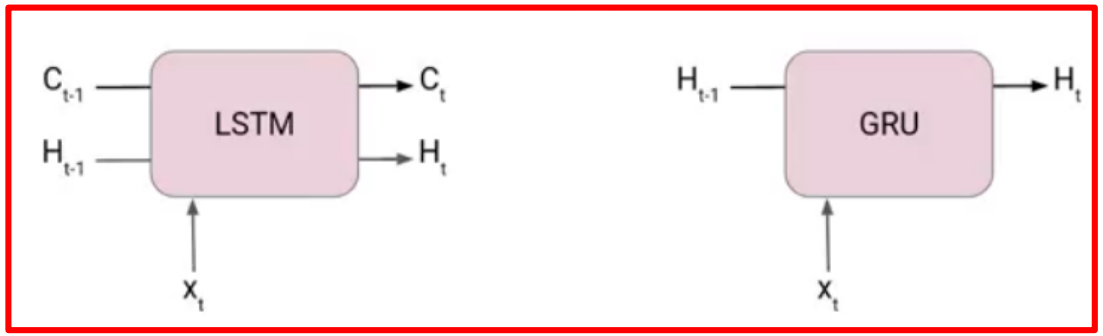


How do GRUs work?



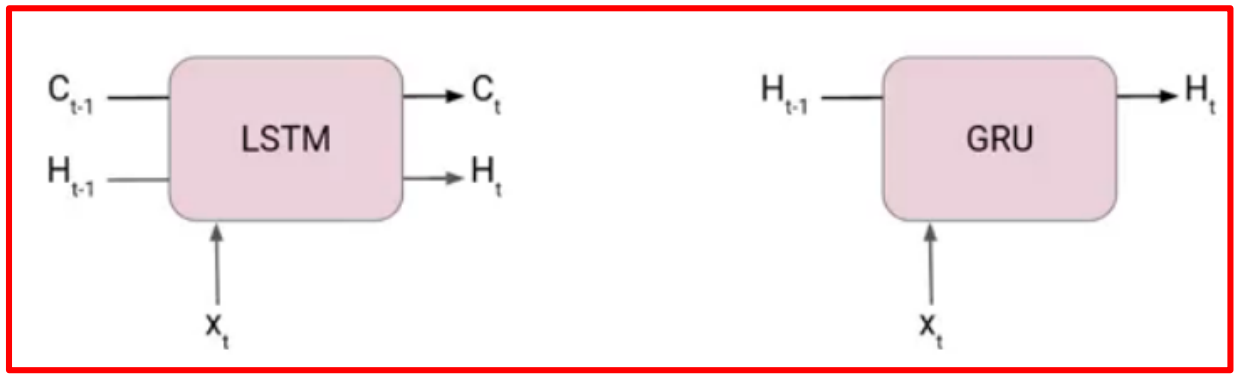
How do GRUs work?

- GRUs are very similar to Long Short Term Memory(LSTM).
- Just like LSTM, GRU uses gates to control the flow of information.
- They are relatively new as compared to LSTM.
- This is the reason they offer some improvement over LSTM and have simpler architecture.



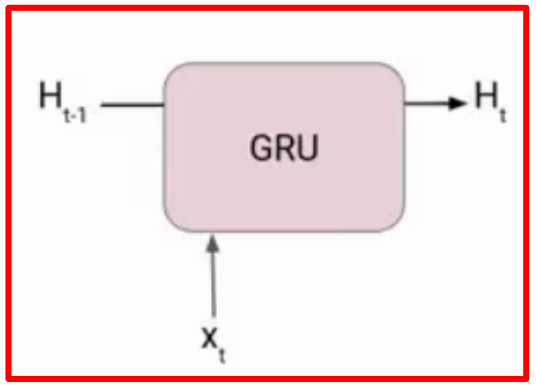
How do GRUs work?

- 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.



The architecture of Gated Recurrent Unit

- Here we have a GRU cell which more or less similar to an LSTM cell or RNN cell.
- At each timestamp t, it takes an input Xt and the hidden state Ht-1 from the previous timestamp t-1. Later it outputs a new hidden state Ht which again passed to the next timestamp.
- Now there are primarily two gates in a GRU as opposed to three gates in an LSTM cell. The first gate is the Reset gate and the other one is the update gate.



Reset Gate (Short term memory)

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

$$r_t = \sigma\left(x_t * U_t + H_{t-1} * W_t\right)$$

- 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.

Update Gate (Long Term memory)

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

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

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

Candidate Hidden State

- To find the Hidden state Ht in GRU, it follows a two-step process. The first step is to generate what is known as the candidate hidden state. As shown below

$$\widehat{H}_t = \tanh\left(x_t * U_g + \left(r_t * H_{t-1}\right) * W_g\right)$$

Candidate Hidden State

Candidate Hidden State

- The most important part of this equation is how we are using the value of the reset gate to control how much influence the previous hidden state can have on the candidate state.
- If the value of rt is equal to 1 then it means the entire information from the previous hidden state Ht-1 is being considered.
- Likewise, if the value of rt is 0 then that means the information from the previous hidden state is completely ignored.

$$\hat{H}_t = \tanh\left(x_t * U_g + \left(r_t \circ H_{t-1}\right) * W_g\right)$$

Hidden state

- Once we have the candidate state, it is used to generate the current hidden state Ht.
- It is where the Update gate comes into the picture.
- Now, this is a very interesting equation, instead of using a separate gate like in LSTM in GRU we use a single update gate to control both the historical information which is Ht-1 as well as the new information which comes from the candidate state.

$$H_t = u_t \circ H_{t-1} + (1 - u_t) \circ \widehat{H}_t$$

Hidden state

- Now assume the value of ut is around 0 then the first term in the equation will vanish which means the new hidden state will not have much information from the previous hidden state. On the other hand, the second part becomes almost one that essentially means the hidden state at the current timestamp will consist of the information from the candidate state only.

$$H_t = u_t \circ H_{t-1} + (1 - u_t) \circ \hat{H}_t$$

Hidden state

- Similarly, if the value of ut is one the second term will become entirely 0 and the current hidden state will entirely depend on the first term that is the information from the hidden state at the previous timestamp t-1.

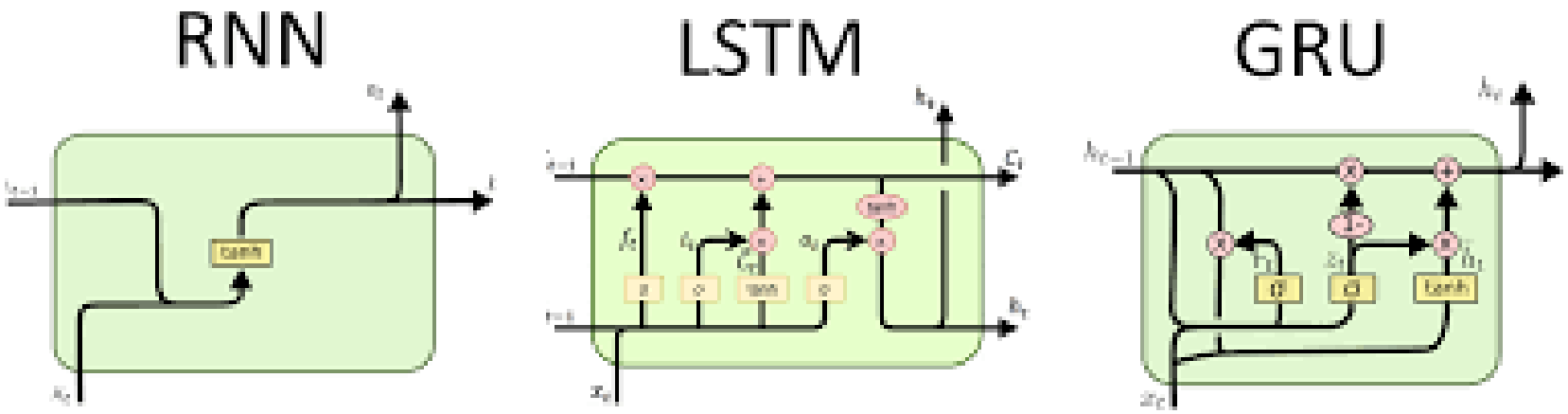
$$H_t = u_t \circ H_{t-1} + (1 - u_t) \circ \widehat{H}_t$$

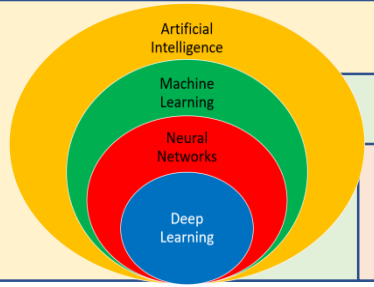
- Hence we can conclude that the value of ut is very critical in this equation and it can range from 0 to 1.

End Notes

- LSTM has three gates on the other hand GRU has only two gates.
- In LSTM they are the Input gate, Forget gate, and Output gate. Whereas in GRU we have a Reset gate and Update gate.
- In LSTM we have two states Cell state or Long term memory and Hidden state also known as Short term memory.
- In the case of GRU, there is only one state that is Hidden state (Ht).

End Notes





Artificial Intelligence (AI)

Machine Learning (ML)

Neural Networks (NNs)

Deep Learning (DL)