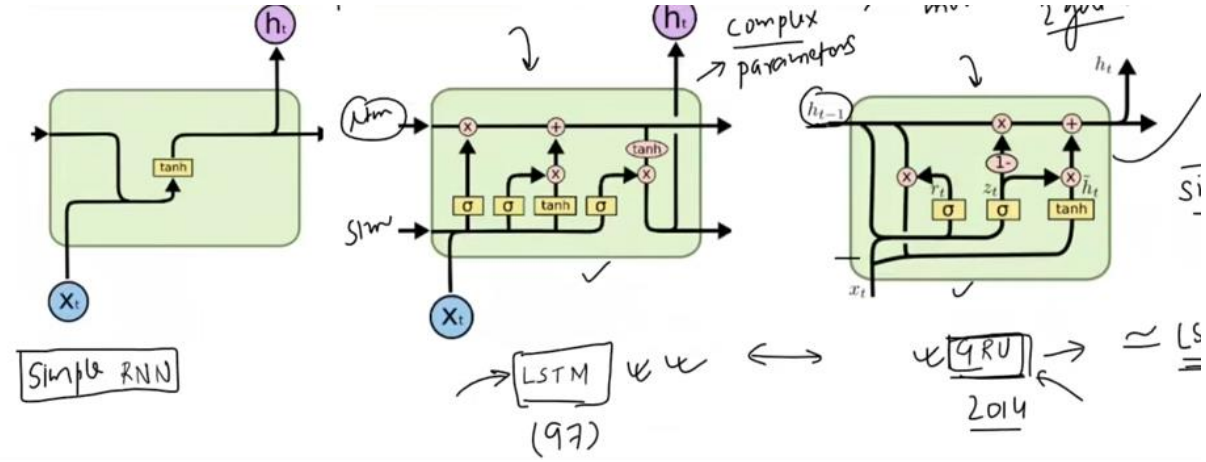
Gated Reccurent Unit (GRU)



Why GRU exisit ?

Simple RNN-> VGD Problem arise hota hai so We use LSTM,

What is the Problem of LSTM ? becoz of which GRU exist

LSTM came in 1997 & GRU Came in 2k14 very recent right ?

As LSTM has lots of parameter so it takes a lot of time in training.

This Problem is solved by GRU

It gives simpler Architecture

There are Certain places where GRU performs better than LSTM but not everywhere,

GRU architecture is simple & straight & takes less time to train & has less parameter.

Long Short-Term Memory (LSTM) networks, while effective in addressing the vanishing gradient problem and capturing long-term dependencies in sequential data, have a more complex architecture compared to Gated Recurrent Units (GRUs).

GRUs were designed to simplify the gating mechanisms while maintaining or improving performance. Here are some aspects where GRUs offer simplicity and potential advantages over LSTMs:

1. **Simpler architecture:** GRUs have a more straightforward structure with two gates (update and reset gates) compared to the three gates in LSTMs (input, forget, and output gates). The reduced number of gates in GRUs can make them computationally less expensive to train and may require fewer parameters.
2. **Fewer parameters:** The simpler architecture of GRUs typically results in fewer parameters compared to LSTMs. This can be advantageous, especially in scenarios where computational resources are limited, as training a model with fewer parameters may require less time and memory.
3. **Ease of training:** GRUs are often reported to be easier to train than LSTMs. The reduced complexity of the architecture may contribute to faster convergence during training, making GRUs more attractive for certain applications, particularly when training resources are limited.
4. **Similar performance:** While LSTMs and GRUs have different architectures, they have been found to perform similarly in many tasks. In practice, the choice between LSTM and GRU often depends on the specific requirements of the task and available computational resources.

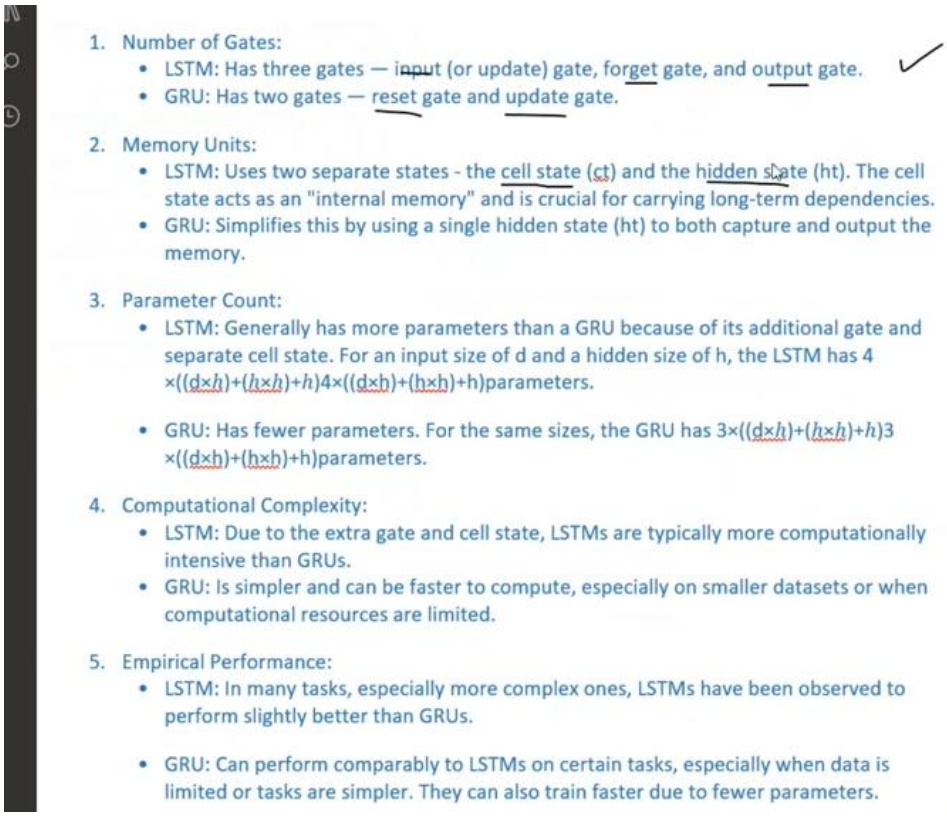
It's essential to note that the performance of LSTM and GRU can vary depending on the specific dataset and task at hand. While GRUs provide a simpler alternative to LSTMs, LSTMs still excel in certain scenarios and remain widely used in the field of deep learning. The choice between LSTM and GRU often involves experimentation to determine which architecture performs better for a particular use case.

Big Idea Behind GRU

In LSTM , Short Term context & Long Term Context ko alag rakhte theah

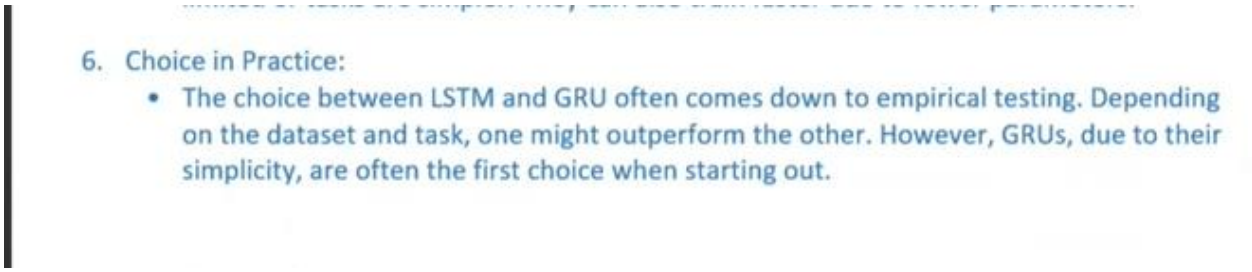
cell state & hidden state are different,

LSTM vs GRU



1. Number of Gates:
   - LSTM: Has three gates — input (or update) gate, forget gate, and output gate.
   - GRU: Has two gates — reset gate and update gate.
2. Memory Units:
   - LSTM: Uses two separate states - the cell state (ct) and the hidden state (ht). The cell state acts as an "internal memory" and is crucial for carrying long-term dependencies.
   - GRU: Simplifies this by using a single hidden state (ht) to both capture and output the memory.
3. Parameter Count:
   - LSTM: Generally has more parameters than a GRU because of its additional gate and separate cell state. For an input size of d and a hidden size of h, the LSTM has 4 ×((d×h)+(h×h)+h)4×((d×h)+(h×h)+h)parameters.
   - GRU: Has fewer parameters. For the same sizes, the GRU has 3×((d×h)+(h×h)+h)3 ×((d×h)+(h×h)+h)parameters.
4. Computational Complexity:
   - LSTM: Due to the extra gate and cell state, LSTMs are typically more computationally intensive than GRUs.
   - GRU: Is simpler and can be faster to compute, especially on smaller datasets or when computational resources are limited.
5. Empirical Performance:
   - LSTM: In many tasks, especially more complex ones, LSTMs have been observed to perform slightly better than GRUs.
   - GRU: Can perform comparably to LSTMs on certain tasks, especially when data is limited or tasks are simpler. They can also train faster due to fewer parameters.

Complex task mai LSTM & Simple Task mai GRU



6. Choice in Practice:
   - The choice between LSTM and GRU often comes down to empirical testing. Depending on the dataset and task, one might outperform the other. However, GRUs, due to their simplicity, are often the first choice when starting out.

GRU is first to startout then LSTM