======================================================================

**Gated Recurrent Unit**

======================================================================

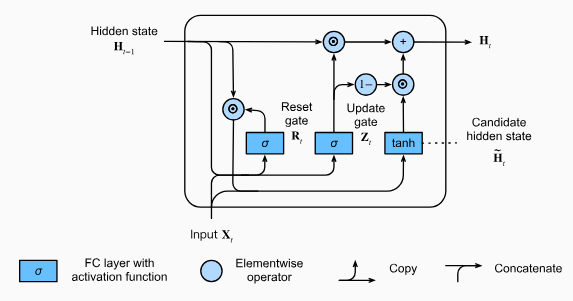
GRU (Gated Recurrent Unit) is a type of Recurrent Neural Network (RNN) designed to handle sequential data and solve the vanishing gradient problem — just like LSTM — but with a simpler architecture.

It is often preferred when faster training or fewer parameters are desired.

--------------------------------------------------------------------------------------------------------------

**GRU Architecture**

Unlike LSTM (which has 3 gates and a cell state), GRU has only 2 gates and no separate cell state — everything is managed through the hidden state hₜ.



**Key Components:**

1. Update Gate (zₜ)

* Decides how much of the past information to keep
* Acts like a combination of the forget and input gates from LSTM

zₜ = σ(Wz · [hₜ₋₁, xₜ] + bz)

2. Reset Gate (rₜ)

* Controls how much of the previous state to forget when combining with the current input.

rₜ = σ(Wr · [hₜ₋₁, xₜ] + br)

3. Candidate Hidden State (ĥₜ)

* New memory content created based on the current input and the reset-modified hidden state

ĥₜ = tanh(W · [rₜ \* hₜ₋₁, xₜ] + b)

4. Final Hidden State (hₜ)

* Combines old and new memory using the update gate

hₜ = (1 - zₜ) \* hₜ₋₁ + zₜ \* ĥₜ

-----------------------------------------------------------------------------------------------------------------

**Summary Table**

| **Component** | **Role** |
| --- | --- |
| zₜ (Update) | How much of past to keep |
| rₜ (Reset) | How much of past to forget |
| ĥₜ | New candidate memory |
| hₜ | Final hidden state |

**Benefits of GRU**

| **Feature** | **Advantage** |
| --- | --- |
| 🧠 Fewer gates | Simpler architecture than LSTM |
| ⚡ Faster training | Less computation = quicker convergence |
| 🔗 No separate memory | Only one state to manage (simpler design) |
| 🧪 Performs well | Often matches or exceeds LSTM on many tasks |
| 🧵 Good for short-to-medium sequences | Without much performance loss |

**Drawbacks of GRU**

| **Issue** | **Description** |
| --- | --- |
| 🔍 Less control | No separate cell state — slightly less expressive than LSTM |
| 📉 Not always better | Sometimes underperforms LSTM on **very long sequences** |
| 🧪 Less widely studied | Fewer pre-trained models compared to LSTM and Transformer |

**GRU vs LSTM**

| **Feature** | **GRU** | **LSTM** |
| --- | --- | --- |
| Gates | 2 (Update, Reset) | 3 (Input, Forget, Output) |
| Cell state | ❌ No separate cell state | ✅ Yes (Cₜ) |
| Speed | ✅ Faster | ⚠️ Slower |
| Memory control | Simpler | More flexible |
| Performance | Equal or slightly worse | Often slightly better |