GRU (Gated Recurrent Unit)

Simplified LSTM with fewer parameters

Two-Gate Mechanism

Update Gate
Combines forget and input gates from LSTM

Reset Gate
Controls past information usage

Faster Training
Than LSTM

Similar Performance
On many tasks

Easier to Implement
And tune

LSTM vs GRU: Key Differences

LSTM

3 gates : Forget, Input, Output

GRU

2 gates : Update, Reset

2 states : Cell state (C) + Hidden state (h)

More parameters to learn

Better for complex, long sequences

Slower training time

More hyperparameters to tune

1 state : Hidden state (h) only

Fewer parameters (~25-30% less)

Better for smaller datasets

Faster training time

Simpler to implement and tune

LSTM Calculation (3D Vector Example)

Given Input:

```
x_t = [0.5, -0.3, 0.8] (current input)
h_{t-1} = [0.2, 0.4, -0.1] (previous hidden state)
C_{t-1} = [0.6, -0.2, 0.3] (previous cell state)
```

Step 1: Forget Gate (f_t)

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Decides what to forget from cell state

```
Result: f_t = [0.65, 0.42, 0.78]
```

Step 2: Input Gate (i_t) + Candidate ($\tilde{C}_t$)

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

Decides what new information to add

Result: $i_t = [0.53, 0.71, 0.48]$, $\tilde{C}_t = [0.32, -0.54, 0.67]$

Step 3: Update Cell State (C_t)

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

Result: $C_t = [0.56, -0.47, 0.55]$

Step 4: Output Gate (o_t) + Hidden State (h_t)

GRU Calculation (3D Vector Example)

Given Input:

$x_t = [0.5, -0.3, 0.8]$ (current input)

$h_{t-1} = [0.2, 0.4, -0.1]$ (previous hidden state)

Note: GRU has NO separate cell state - only hidden state!

Step 1: Reset Gate (r_t)

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r)$$

Controls how much past info to use for candidate

```
Result: r_t = [0.58, 0.73, 0.45]
```

Step 2: Update Gate (z_t)

```
z_t = σ(W_z · [h_{t-1}, x_t] + b_z)
```


Combines forget & input: decides mix of old vs new

Practice Materials

Interactive RNN Visualization Tool

RNN Explainer

Interactive simulation tool to visually understand how LSTM and GRU work

Go to Practice Site →

Gate Operation Visualization

Interactive Parameter Control

Real-time Results