LSTM Architecture

LSTM cells use three gates to control the information flow:

- **Forget Gate (f_t)**: Decides what information to discard from the cell state.

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

- **Input Gate (i_t)**: Determines what new information to store in the cell state.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

- **Cell State Update (C_t)**: The cell state is updated based on the forget and input gates.

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

- **Output Gate (o_t)**: Controls how much of the cell state to output as the hidden state.

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$h_t = o_t \cdot \tanh(C_t)$$

The cell state C_t is a key feature, allowing LSTMs to retain information over long sequences, effectively mitigating the vanishing gradient problem.

GRU Architecture

GRUs simplify LSTM architecture by combining the cell state and hidden state and using only two gates:

- **Update Gate (z_t)**: Controls how much of the past information needs to be retained.

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z)$$

- **Reset Gate (r_t)**: Determines how much of the previous hidden state to forget.

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

The GRU hidden state update is given by:

$$\tilde{h}_t = \tanh(W_h x_t + r_t \cdot (U_h h_{t-1}) + b_h)$$

$$h_t = z_t \cdot h_{t-1} + (1 - z_t) \cdot \tilde{h}_t$$

GRU's update mechanism combines the functions of the input and forget gates in an LSTM, resulting in a more compact and computationally efficient architecture.