1. Basic Architecture of RNN Cell

An RNN (Recurrent Neural Network) cell processes sequential data by maintaining a hidden state that captures information from previous time steps.

- **It takes an input `x_t` and the previous hidden state `h_t-1`.**
- **Outputs a new hidden state `h_t` by applying a function, usually `tanh` or `ReLU`.**

Key Idea: RNNs share weights across time steps, allowing memory of previous inputs.

---

2. Backpropagation Through Time (BPTT)

BPTT is the method used to train RNNs. It unrolls the RNN across time and applies backpropagation across each time step.

- **It computes gradients through the sequence.**
- **Helps the model learn dependencies across time.**

Issue: Can lead to vanishing or exploding gradients.

---

3. Vanishing and Exploding Gradients

- **Vanishing Gradients: Gradients become too small, stopping the model from learning long-range dependencies.**
- **Exploding Gradients: Gradients become too large, causing unstable training.**

Both problems arise during BPTT due to repeated multiplication of gradients.

---

4. Long Short-Term Memory (LSTM)

LSTM is a type of RNN designed to handle long-term dependencies by using gates to control the flow of information.

- **Maintains a cell state that can carry long-term information.**

- **Uses input, forget, and output gates to manage information.**

---

5. Gated Recurrent Unit (GRU)

GRU is a simplified version of LSTM that combines the forget and input gates into a single update gate.

- **Fewer parameters than LSTM.**
- **Performs comparably in many tasks, and trains faster.**

---

6. Peephole LSTM

A variation of LSTM where the gates (input, forget, output) also have access to the cell state, not just the hidden state.

- **Helps the model learn precise timing and count-based patterns.**

---

7. Bidirectional RNNs

Processes the sequence in both forward and backward directions.

- **Two RNNs run: one from start to end, the other from end to start.**
- **Outputs are usually concatenated.**

Benefit: Access to past and future context.

---

8. Gates of LSTM (with equations described in words)

LSTM has three gates and a cell state:

- **Forget gate: Decides what to discard from the cell state.**
  Takes input and previous hidden state → outputs a value between 0 and 1.

- **Input gate: Decides what new information to add.**
  Combines input and hidden state to update the cell.

- **Output gate: Decides what to output as the new hidden state.**

Cell state is updated by combining the forget gate (to remove info) and input gate (to add info). Hidden state is based on the cell state and output gate.

---

9. BiLSTM (Bidirectional LSTM)

A Bidirectional LSTM combines two LSTMs:

- **One reads the input sequence from start to end.**

- **The other reads it from end to start.**

Their outputs are concatenated or combined, giving the model both past and future context for each point in the sequence.

---

10. BiGRU (Bidirectional GRU)

Same concept as BiLSTM, but uses GRU cells instead of LSTM.

- **Lighter and faster than BiLSTM.**

- **Still captures forward and backward context effectively.**