SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of AI &DS

Course Name – 19AD602 DEEP LEARNING

III Year / VI Semester

UNIT-4 OPTIMIZATION AND GENERALIZATION

Topic: LSTM

CASE STUDY:

Leveraging LSTMs for Stock Price Prediction

A financial firm implemented Long Short-Term Memory (LSTM) networks to predict stock prices based on historical market data. By capturing long-term dependencies in time series data, the LSTM model outperformed traditional models in forecasting trends, reducing prediction errors by 15%. This approach helped traders make more informed investment decisions.

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN) designed to better capture long-term dependencies in sequential data. They address the limitations of traditional RNNs, particularly the vanishing gradient problem, which can hinder learning when dealing with long sequences.

LSTM

An LSTM unit consists of the following components:

1. **Cell State (Ct)**: Acts as the memory of the network, carrying information across time steps with minimal modification.
2. **Forget Gate (ft)**: Decides what information to discard from the cell state.
3. **Input Gate (it)**: Determines what new information to add to the cell state.
4. **Output Gate (ot)**: Decides what part of the cell state to output.
5. **Candidate Cell State (C~)**: Proposes new information to be added to the cell state.

Chill and Grow

FORGET GATE

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

Chill and Gro

- f_t is the forget gate's output.
- σ is the sigmoid activation function that outputs values between 0 and 1.
- h_{t-1} is the previous hidden state (short-term memory).
- x_t is the current input.
- W_f and b_f are the weights and biases for the forget gate.

INPUT GATE AND CANDIDATE CELL STATE

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

Chill and Grow

- i_t is the input gate's output.
- $\tilde{C}_t$ is the candidate memory (new information to be added).
- W_i, W_C, b_i, and B_c are weights and biases for the input gate and candidate memory.

CELL STATE UPDATE

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

Chill and

- C_t is the new cell state.
- $f_t * C_{t-1}$ forgets part of the previous cell state.
- $i_t * \tilde{C}_t$ adds new information to the cell state.

OUTPUT STATE

$$o_t=\sigma(W_o\cdot[h_{t-1},x_t]+b_o)$$

HIDDEN STATE UPDATE

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

Chill and Grow

- o_t is the output gate's activation.
- h_t is the new hidden state (short-term memory).

FORGET GATE

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

- f_t is the forget gate's output.
- σ is the sigmoid activation function that outputs values between 0 and 1.
- h_{t-1} is the previous hidden state (short-term memory).
- x_t is the current input.
- W_f and b_f are the weights and biases for the forget gate.

LSTM

1. Gradient Through the Cell State

$$\partial L/\partial C_t = (\partial L/\partial h_t) \cdot o_t \cdot (1 - \tanh^2(C_t))$$

2. Updating the Cell State

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

3. Backpropagating the Gradient Through Time

Chill and Grow

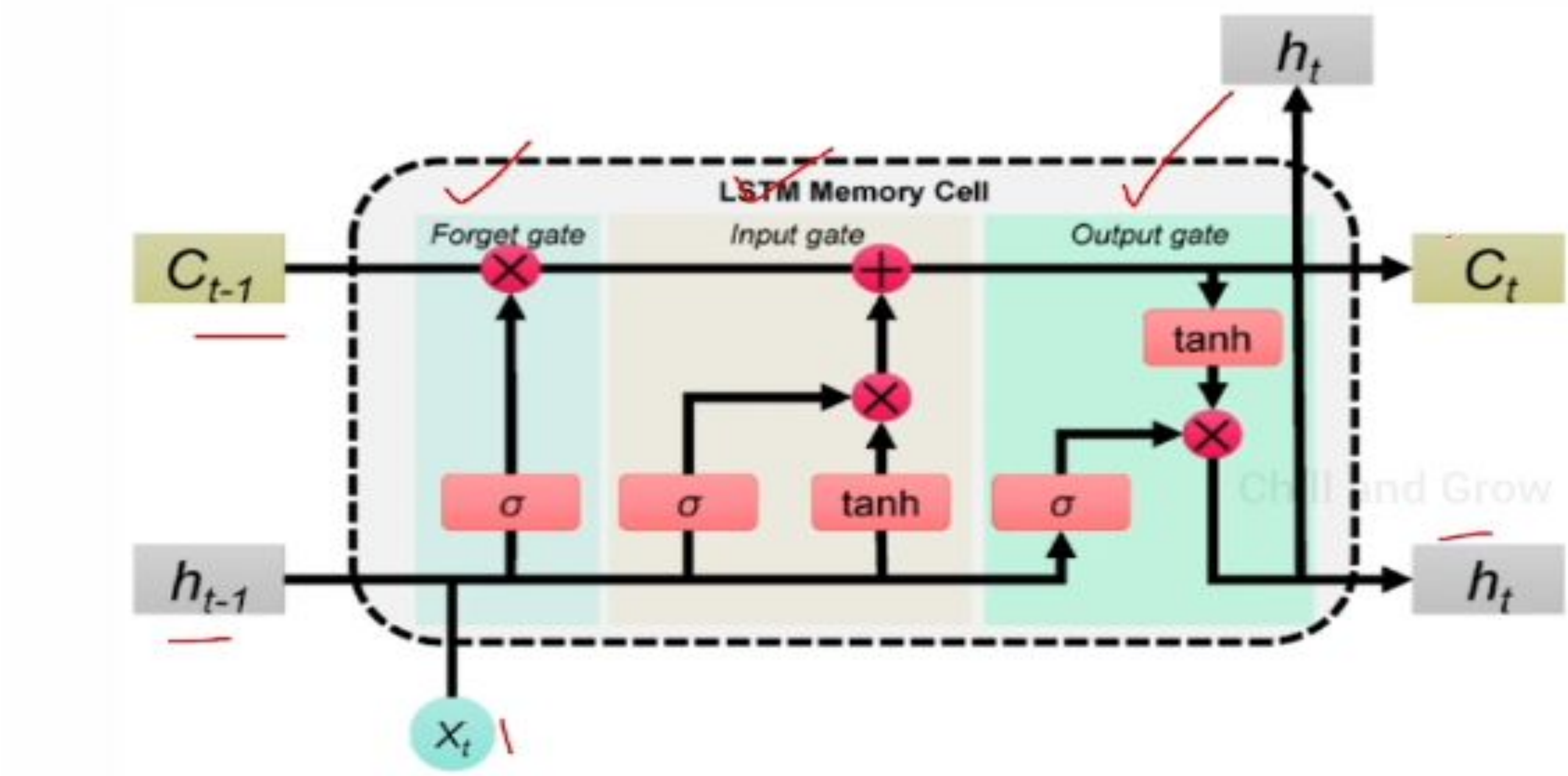
$$\partial L/\partial C_{t-1} = (\partial L/\partial C_t) \cdot f_t$$

- **∂Ct**: This is the gradient of the loss with respect to the current cell state Ct.
- **ft**: The forget gate at time t controls how much of the gradient flows back to the previous cell state Ct−1.

LSTM

If ft is close to 1 (which often happens when the LSTM decides the information is important to retain), the gradient ∂L∂Ct−1 is almost the same as ∂L∂Ct. This means the gradient does not diminish as it propagates backward through time, helping the network learn long-term dependencies effectively.

LSTM



LSTM Memory Cell
Forget gate
Input gate
Output gate
C_{t-1}
C_t
h_{t-1}
h_t
h_t
X_t
σ
σ
tanh
σ
tanh

THANK YOU