LSTM Numerical Example

If we have a simple input sequence, and we need to predict the next value:

$$X = [1,2,3]$$

LSTM Parameters

- Input dimension = 1
- Hidden state dimension = 1
- Forget gate, input gate, output gate, and candidate values are calculated using predefined weights.

Step 1: Initialize Parameters

Let's assume the following arbitrary values for the LSTM parameters:

Weights

$$\begin{aligned} W_f &= 0.5, \quad W_{hf} = 0.1, \quad b_f = 0 \\ W_i &= 0.6, \quad W_{hi} = 0.2, \quad b_i = 0 \\ W_c &= 0.7, \quad W_{hc} = 0.3, \quad b_c = 0 \\ W_o &= 0.8, \quad W_{ho} = 0.4, \quad b_o = 0 \end{aligned}$$

Initial States

$$h_0 = 0, \; C_0 = 0$$

Step 2: Compute LSTM Values for Each Time Step

Time Step $t = 1$, Input $x_1 = 1$

1 Forget gate:

$$f_1 = \sigma(0.5(1) + 0.1(0) + 0) = \sigma(0.5) = \frac{1}{1 + e^{-0.5}} \approx 0.622$$

2 Input gate:

$$i_1 = \sigma(0.6(1) + 0.2(0) + 0) = \sigma(0.6) = \frac{1}{1 + e^{-0.6}} \approx 0.645$$

3 Candidate cell state:

$$\tilde{C}_1 = \tanh(0.7(1) + 0.3(0) + 0) = \tanh(0.7) \approx 0.604$$

4 Cell state update:

$$C_1 = (0.622 \cdot 0) + (0.645 \cdot 0.604) = \mathbf{0.390}$$

5 Output gate:

$$o_1 = \sigma(0.8(1) + 0.4(0) + 0) = \sigma(0.8) = \frac{1}{1 + e^{-0.8}} \approx 0.690$$

6 Hidden state update:

$$h_1 = 0.690 \cdot \tanh(0.390) = 0.690 \cdot 0.372 \approx \mathbf{0.257}$$

Time Step $t = 2$, Input $x_2 = 2$

1 Forget gate:

$$f_2 = \sigma(0.5(2) + 0.1(\mathbf{0.257}) + 0) = \sigma(1.026) \approx 0.736$$

2 Input gate:

$$i_2 = \sigma(0.6(2) + 0.2(\mathbf{0.257}) + 0) = \sigma(1.252) \approx 0.777$$

3 Candidate cell state:

$$\tilde{C}_2 = \tanh(0.7(2) + 0.3(0.257) + 0) = \tanh(1.471) \approx 0.899$$

4 Cell state update:

$$C_2 = (0.736 \cdot \mathbf{0.390}) + (0.777 \cdot 0.899) = 0.287 + 0.698 = \mathbf{0.985}$$

5 Output gate:

$$o_2 = \sigma(0.8(2) + 0.4(0.257) + 0) = \sigma(1.436) \approx 0.808$$

6 Hidden state update:

$$h_2 = 0.808 \cdot \tanh(0.985) = 0.808 \cdot 0.756 \approx \mathbf{0.611}$$

Time Step $t = 3$, Input $x_3 = 3$

1 Forget gate:

$$f_3 = \sigma(0.5(3) + 0.1(\mathbf{0.611}) + 0) = \sigma(1.561) \approx 0.827$$

2 Input gate:

$$i_3 = \sigma(0.6(3) + 0.2(\mathbf{0.611}) + 0) = \sigma(1.882) \approx 0.867$$

3 Candidate cell state:

$$\tilde{C}_3 = \tanh(0.7(3) + 0.3(\mathbf{0.611}) + 0) = \tanh(2.183) \approx 0.974$$

4 Cell state update:

$$C_3 = (0.827 \cdot \mathbf{0.985}) + (0.867 \cdot 0.974) = 0.815 + 0.844 = \mathbf{1.659}$$

5 Output gate:

$$o_3 = \sigma(0.8(3) + 0.4(0.611) + 0) = \sigma(2.244) \approx 0.904$$

6 Hidden state update:

$$h_3 = 0.904 \cdot \tanh(1.659) = 0.904 \cdot 0.930 \approx \mathbf{0.841}$$

Time Step $t = 4$, Predict $x_4 = ?$

1 Forget Gate:

$$f_4 = \sigma(0.5(4) + 0.1(\mathbf{0.841}) + 0) = \sigma(2.084) \approx 0.889$$

2 Input Gate:

$$i_4 = \sigma(0.6(4) + 0.2(\mathbf{0.841}) + 0) = \sigma(2.536) \approx 0.926$$

3 Candidate Cell State:

$$\tilde{C}_4 = \tanh(0.7(4) + 0.3(\mathbf{0.841}) + 0) = \tanh(2.652) \approx 0.990$$

4 Cell State Update:

$$C_4 = (0.889 \cdot \mathbf{1.659}) + (0.926 \cdot 0.990) = 1.474 + 0.916 = 2.390$$

5 Output Gate:

$$o_4 = \sigma(0.8(4) + 0.4(0.841) + 0) = \sigma(3.336) \approx \mathbf{0.965}$$

6 Hidden State Update:

$$h_4 = 0.965 \cdot \tanh(2.390) = 0.965 \cdot 0.983 \approx \mathbf{0.949}$$

Step 2: Predict the Next Value

Since we are predicting a numerical sequence like 1,2,3, the output of the LSTM must be mapped to a real number.
We assume a linear transformation from h_t to the predicted value:

$$\hat{y} = W_y h_t + b_y$$

Let's assume:

$$W_y = 4,\ b_y = 0$$
$$\hat{y} = 4 \times 0.949 = 3.796$$

Final Prediction

The LSTM predicts 3.8 , which is close to 4 , showing that the model has learned the pattern.