LSTM (Long Short-Term Memory) Model

Gaurang
M23MA1007

Under the Supervision of

Dr.Vivek Vijay and Dr. Sandeep Kumar Yadav

॥ त्वं ज्ञानमयो विज्ञानमयोऽसि ॥

Department of Mathematics
IIT Jodhpur

Nov 27, 2023

Contents

Current Section

Introduction

- LSTM stands for Long Short-Term Memory, a type of recurrent neural network (RNN).
- Designed to capture long-term dependencies in sequence data.
- Effective in addressing the vanishing gradient problem.
  It address the vanishing gradient problem by using a special architecture that includes memory cells and gating mechanisms (input, output, and forget gates). These components help maintain and regulate the flow of information and gradients throughout the network, allowing it to learn long-term dependencies more effectively without the gradients diminishing excessively.

Introduction

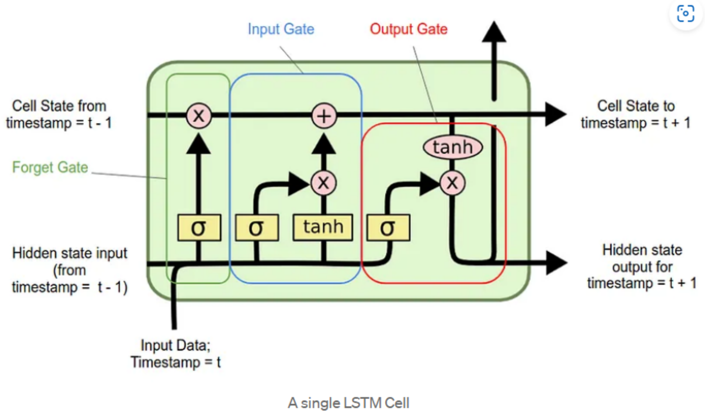

Figure 1: Model Training

Current Section

LSTM Cell Structure

- Each LSTM cell contains three gates: forget gate, input gate, and output gate.
- These gates regulate the flow of information.

Current Section

Forget Gate

- Determines which information to discard from the cell state.
- **Mathematical Concept:**

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

where σ is the sigmoid function, W_f is the weight matrix, h_{t-1} is the previous hidden state, x_t is the current input, and b_f is the bias.

Current Section

Input Gate

- Decides which values to update in the cell state.
- **Mathematical Concept:**

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

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

  where i_t is the input gate activation, $\tilde{C}_t$ is the candidate cell state, and W_i, W_C are weight matrices.

Current Section

Cell State Update

- Updates the cell state based on the forget and input gates.
- **Mathematical Concept:**

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

where C_t is the new cell state, C_{t-1} is the previous cell state.

Current Section

Output Gate

- Determines which part of the cell state to output.
- **Mathematical Concept:**

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

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

where o_t is the output gate activation, h_t is the new hidden state, and W_o is the weight matrix.

Current Section

Hidden State Propagation

- The hidden state (h_t) and cell state (C_t) are passed to the next LSTM cell in the sequence.

Current Section

Training LSTM

- LSTM networks are trained using backpropagation through time (BPTT).
- Adjusts the weights W_f, W_i, W_C, W_o based on error gradients.

Current Section

Applications

- Used in time series forecasting, language modeling, and speech recognition.
- Effective in tasks involving sequential data.

Current Section

Conclusion

- LSTMs are powerful tools for sequence data analysis.
- Address the limitations of traditional RNNs by capturing long-term dependencies.

Current Section

Future Works

- Explore advanced variations of LSTM, such as Bidirectional LSTMs and GRUs.
- Apply LSTMs to more complex and diverse datasets.

Current Section

Introduction

- The goal is to predict the share price of Facebook stocks using an LSTM model.
- Data source: Yahoo Finance (MSFT.csv for demonstration).

Data Preprocessing

- Load and preprocess the data.
- Convert date strings to datetime objects.
- Plot the stock closing prices.

Data Preprocessing

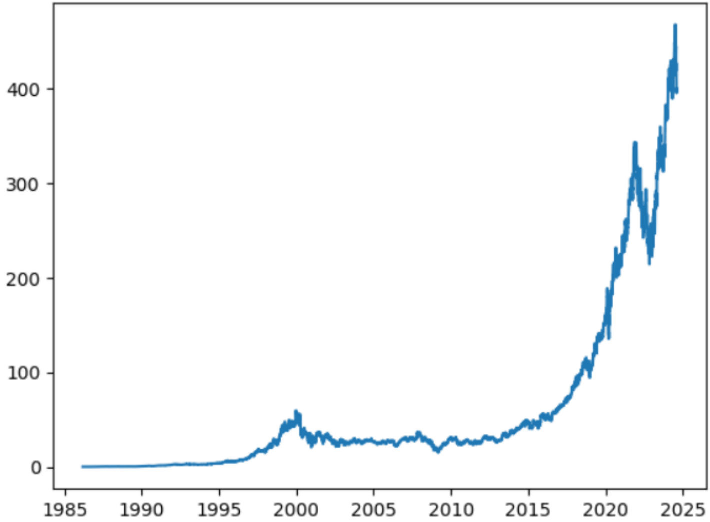

Figure 2: Stock Closing Prices

Data Normalization

- Normalize the data for better performance.
- Use MinMaxScaler to scale the data between 0 and 1.

Data Normalization

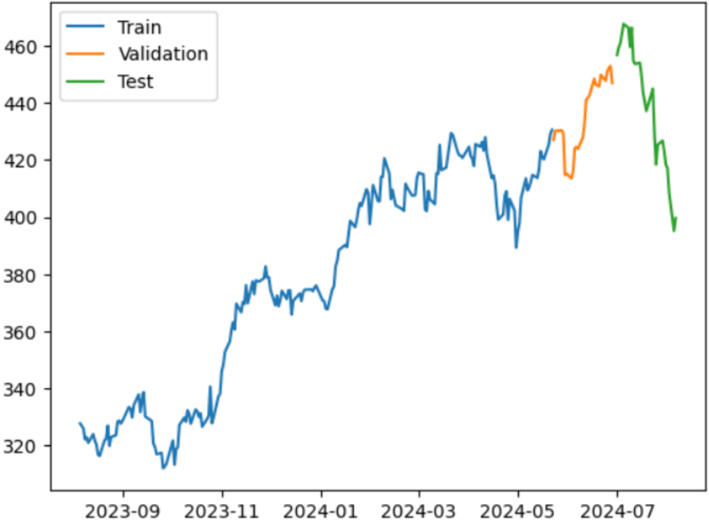

Figure 3: Normalized Data

Model Training

- Split the data into training and testing sets.
- Define and compile the LSTM model.
- Train the model on the training data.

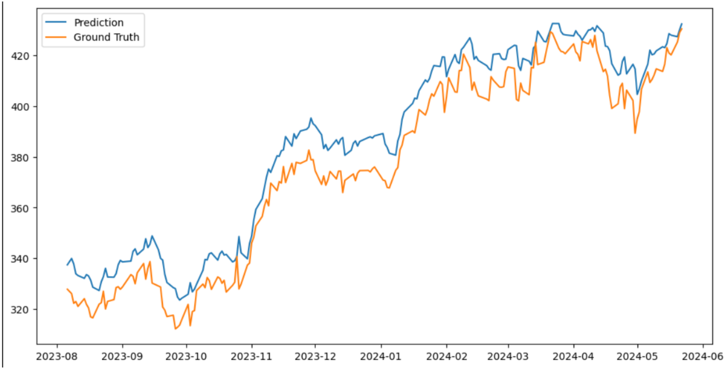

Figure 4: Model Training

Model Training

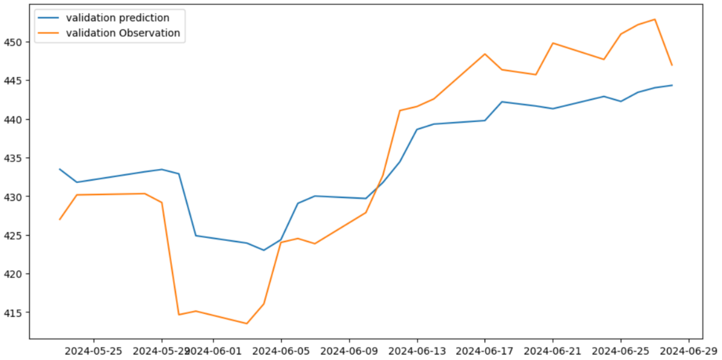

Figure 5: Validation part

Predictions and Results

- Make predictions on the test data.
- Plot the predicted vs. actual stock prices.

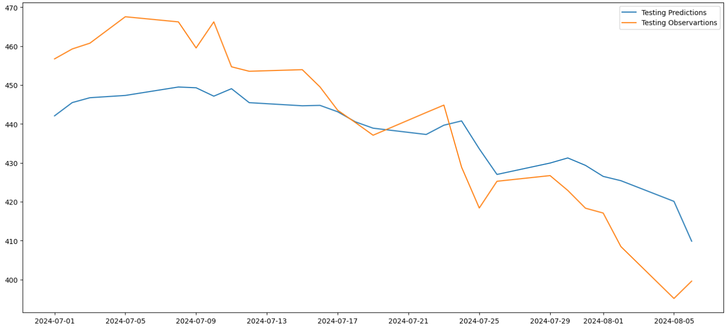

Figure 6: Predicted vs Actual Stock Prices

Conclusion

- LSTM models are effective for stock price prediction.
- The model captures the trends and patterns in the stock price data.

Future Works

- Explore more features and parameters to improve the model accuracy.
- Apply the model to other stocks and financial instruments.

Current Section

Acknowledgement

- Special thanks to my supervisors, Dr.Vivek Vijay and Dr. Sandeep Kumar Yadav.
- Appreciation to the Department of Mathematics, IIT Jodhpur.

Current Section

Thank You

Thank you for your attention!
Questions?