Understanding LSTMs

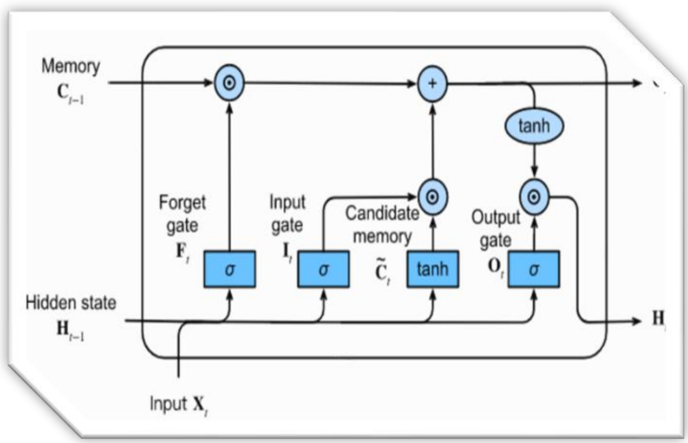
- the gated controls of data

Long Short-Term Memory (LSTM) networks are a powerful type of recurrent neural network (RNN) architecture that address a major limitation of standard RNNs: the vanishing gradient problem. This problem makes it difficult for RNNs to learn long-term dependencies in sequential data.

LSTMs tackle this by introducing a concept called a **memory cell**. This cell, along with three special gates (forget gate, input gate, and output gate), allows LSTMs to selectively remember and forget information over long sequences.

Components

- **Cell State (C):** This represents the actual memory of the LSTM cell. It can hold information for long periods of time.
- **Hidden State (H):** This is the output of the LSTM cell at a given point in time. It is calculated based on the cell state and the current input.
- **Forget Gate (F):** This gate decides what information to forget from the cell state. It uses a sigmoid activation function which outputs a value between 0 and 1. A value closer to 1 means forget more, and a value closer to 0 means forget less.



- **Input Gate (I):** This gate decides which information from the current input (X) to add to the cell state. It also uses a sigmoid activation function.
- **Candidate Memory (C~):** This is the new information that is created based on the current input and the previous hidden state. It is calculated using a hyperbolic tangent (tanh) activation function, which outputs a value between -1 and 1.
- **Output Gate (O):** This gate decides what information from the cell state to output as the hidden state. It uses a sigmoid activation function.

LSTM networks leverage gated memory cells to effectively manage information flow.

The forget gate acts as a selective filter, determining which past information (contained in the previous cell state) remains relevant. It analyzes both the current input and the preceding hidden state, outputting weights between 0 and 1 for each element in the cell state. These weights indicate the proportion of information to forget from the cell.

The input gate, using the same inputs, generates weights between 0 and 1 to control how much of the newly generated candidate memory (created using a tanh function) is incorporated into the cell state.

The cell state update essentially combines these influences, selectively forgetting the past and integrating relevant new information.

Finally, the output gate determines the impact of the updated cell state on the current hidden state, controlling how much of the cell's information is propagated forward in the network. This interplay between gates allows LSTMs to learn long-term dependencies while mitigating the vanishing gradient problem that hinders traditional RNNs.

Within LSTM networks, the critical gating mechanisms rely on specific activation functions to achieve optimal information flow.

The forget gate, input gate, and output gate all utilize sigmoid activation due to their role in selectively processing information. Since these gates determine how much information to forget, add, or output, the sigmoid function's inherent characteristic of producing values between 0 and 1 aligns perfectly with this task.

In contrast, the candidate memory leverages a tanh activation function. This choice allows for a broader range of values, encompassing both positive and negative numbers. This extended range proves beneficial for capturing the nuances of new information that needs to be integrated into the cell state. It's important to clarify that the concept of a "memory gate" isn't a distinct component within LSTMs. Instead, it represents the combined influence of the forget gate and input gate acting on the cell state, effectively regulating the balance between retaining past information and incorporating new elements.

LSTMs are widely used in various tasks, including machine translation, speech recognition, handwriting recognition, time series forecasting, and anomaly detection.

All in all:

LSTMs, or Long Short-Term Memory networks, are a type of neural network that excels at handling sequential data like text or speech. Unlike traditional RNNs, LSTMs can learn long-term dependencies in data thanks to their special memory cell and gated architecture. This makes them a powerful tool for tasks like machine translation and speech recognition.

Reference: https://www.geeksforgeeks.org/understanding-of-lstm-networks/, https://medium.com/@anishnama20/understanding-lstm-architecture-pros-and-cons-and-implementation-3e0cca194094, https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/, https://colah.github.io/posts/2015-08-Understanding-LSTMs/.