**Long Short-Term Memory (LSTM)**

LSTMs (Long Short-Term Memory networks) are a type of recurrent neural network (RNN) capable of processing and analyzing sequential data such as time series, text, and speech. They use a memory cell and gates to control the flow of information, allowing them to selectively retain or discard information as needed. This mechanism helps them avoid the vanishing gradient problem that plagues traditional RNNs.

There are three main stages in an LSTM: the forget stage, the input stage, and the output stage.

1. **Forget Gate Stage**:
   * **Input:** Long-term memory (cell state) and short-term memory (hidden state).
   * **Process:** The short-term memory and the current input are multiplied by their respective weights and summed. This sum is then passed through a sigmoid function, producing values between 0 and 1.
   * **Output:** This output is multiplied with the long-term memory to determine how much of it should be retained or discarded. This process essentially "forgets" parts of the long-term memory.
2. **Input Gate Stage**:
   * **Input:** Long-term memory and current input.
   * **Process:** The short-term memory and current input are again multiplied by their respective weights and passed through a sigmoid function. Simultaneously, the same values are passed through a tanh function to generate candidate values for the long-term memory.
   * **Output:** The output from the sigmoid function determines how much of the candidate values (from the tanh function) should be added to the long-term memory. This updates the long-term memory with new information.
3. **Output Gate Stage**:
   * **Input:** Updated long-term memory from the input stage and current input.
   * **Process:** The updated long-term memory is passed through a tanh function to scale the values between -1 and 1, producing potential short-term memory values. The short-term memory and current input are multiplied by their respective weights and passed through a sigmoid function to determine the proportion of the potential short-term memory that should be kept.
   * **Output:** The final short-term memory value is obtained by multiplying the potential short-term memory with the output from the sigmoid function. This updated short-term memory is the output of the model.

In summary, LSTMs use these stages to control what information is retained or discarded, allowing them to effectively manage long-term dependencies and improve performance on sequential data tasks.

A diagram of a graph

Description automatically generated