**Assignment 5**

**Azhan Saleem**

**11635219**

**PART I: RNN: LSTM: Core Concepts: Cell (C) State (50 Points)**

**PART I: RNN: LSTM: Core Concepts: Cell (C) State**

1. **A Flow in a Channel Perspective:**

In the LSTM (Long Short-Term Memory) neural network architecture, the Cell (C) state can be analogized to the flow of water in a channel. This perspective helps in understanding how information is retained, updated, and propagated through the network over time.

* **Flow of Information:** The Cell state serves as a conduit for information flow within the LSTM network. Like how water flows through a channel, the Cell state carries information from one time step to the next, allowing the network to retain memory over extended sequences of inputs.
* **Channel Constraints**: Just as a channel guides and shapes the flow of water, the structure of the LSTM's gates (input, forget, and output gates) regulates the flow of information within the Cell state. These gates control how much new information is added, how much old information is retained, and how much information is passed to the output.
* **Memory Retention and Update**: The Cell state accumulates memory from past time steps and updates it based on current inputs and the network's internal dynamics. This continuous flow and update of information enable the LSTM to capture long-term dependencies and patterns in sequential data.

2. **A Snapshot of the Cell State:**

Alternatively, the Cell state can be envisioned as a snapshot of the network's internal memory at a specific time step. This perspective provides insight into the instantaneous representation of information stored within the Cell state.

* **Instantaneous Memory:** At any given time, step, the Cell state encapsulates a snapshot of the network's memory, including both short-term and long-term information. This snapshot reflects the current state of the network's internal representation of the input sequence.
* **Memory Content**: The content of the Cell state snapshot represents the cumulative effect of past inputs and interactions with the network's gates. It contains encoded information about relevant features, patterns, and context present in the input sequence up to that point.
* **Temporal Context**: While each snapshot captures the current state of memory, the LSTM's recurrent nature ensures that information from previous time steps influences subsequent snapshots. Thus, the Cell state snapshots collectively capture the temporal context and dependencies present in the input sequence.

In summary, viewing the Cell state of the LSTM neural network from the perspectives of a flow in a channel and a snapshot provides complementary insights into its role in information processing and memory retention. These perspectives highlight how the Cell state facilitates the propagation of information over time and serves as a dynamic repository of contextual knowledge within the network.

**PART II: RNN: LSTM: Core Concepts: Gates**

In an LSTM (Long Short-Term Memory) neural network, each gate can be viewed as a neural network because it performs complex computations and learns to control the flow of information based on input data and learned parameters. Here's why each gate can be considered as a neural network:

1. **Input Gate as a Neural Network**:
   * The input gate determines how much new information should be added to the cell state. It takes input from the current input data and the previous hidden state and learns to selectively update the cell state.
   * The input gate consists of a sigmoid activation function layer followed by a pointwise multiplication operation. The sigmoid layer acts as a gatekeeper, deciding which information is relevant to retain, and the pointwise multiplication operation modulates the input based on the gate's decision.
   * Through training, the parameters of the input gate's layers are adjusted to optimize the network's performance in processing input data and updating the cell state. This learning process resembles the training of a neural network, where weights and biases are tuned to minimize the loss function.
2. **Forget Gate as a Neural Network:**
   * The forget gate determines how much of the previous cell state should be retained or forgotten. It takes input from the current input data and the previous hidden state and learns to selectively erase irrelevant information from the cell state.
   * Similar to the input gate, the forget gate consists of a sigmoid activation function layer followed by a pointwise multiplication operation. The sigmoid layer decides which parts of the cell state should be retained or forgotten, and the pointwise multiplication operation modulates the cell state accordingly.
   * Through training, the forget gate's parameters are optimized to prioritize relevant information for retention while suppressing irrelevant information. This process of parameter optimization mirrors the training of a neural network, where the model learns to make decisions based on input data.
3. **Output Gate as a Neural Network:**
   * The output gate determines how much of the current cell state should be outputted to the next hidden state. It takes input from the current input data and the previous hidden state and learns to selectively pass relevant information to the output.
   * Like the input and forget gates, the output gate comprises a sigmoid activation function layer followed by a pointwise multiplication operation. The sigmoid layer controls the extent to which each component of the cell state is output, and the pointwise multiplication operation scales the output accordingly.
   * Through training, the parameters of the output gate are adjusted to optimize the network's performance in generating output. This training process resembles the training of a neural network, where the model learns to produce accurate predictions based on input data.

In summary, each gate in the LSTM neural network can be viewed as a neural network because it performs computations on input data, learns to make decisions based on learned parameters, and influences the flow of information within the network. Through training, the parameters of each gate are optimized to improve the network's performance.