1. Explain the basic architecture of RNN cell.

The basic architecture of a Recurrent Neural Network (RNN) cell consists of a recurrent unit that processes sequential input data. The RNN cell maintains an internal state, which serves as its memory, allowing it to capture and store information about previous inputs. The architecture of an RNN cell can be represented as follows:

Input

Recurrent Unit

Internal State

Output

1. Explain Backpropagation through time (BPTT)

Backpropagation through time (BPTT) is an algorithm used to train recurrent neural networks (RNNs) by propagating gradients backwards through time. It is an extension of the backpropagation algorithm for feedforward neural networks, designed to handle the recurrent nature and temporal dependencies of RNNs

1. Explain Vanishing and exploding gradients

Vanishing and exploding gradients are issues that can occur during the training of deep neural networks, including recurrent neural networks (RNNs). These problems arise due to the repeated multiplication of gradients as they are backpropagated through the layers of a network, potentially causing the gradients to become extremely small (vanishing gradients) or excessively large (exploding gradients).

Vanishing Gradients

Exploding Gradients

1. Explain Long short-term memory (LSTM)

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem and capture long-term dependencies in sequential data. LSTM networks are widely used in various natural language processing (NLP), speech recognition, and time series analysis tasks.

The key idea behind LSTM is the incorporation of memory cells and specialized gating mechanisms that enable the network to selectively remember or forget information over long sequences. This enables LSTM to capture and maintain information from earlier time steps, mitigating the vanishing gradient problem and allowing for more effective learning.

1. Explain Gated recurrent unit (GRU)

The Gated Recurrent Unit (GRU) is a variant of recurrent neural network (RNN) architecture that addresses the vanishing gradient problem and captures long-term dependencies in sequential data, similar to LSTM (Long Short-Term Memory) networks. GRU simplifies the architecture compared to LSTM by combining the forget and input gates into a single update gate, resulting in a more compact and efficient model.

Hidden State (ht)

Update Gate (zt)

Reset Gate (rt)

Candidate State (~ht)

1. Explain Peephole LSTM

Peephole LSTM is an extension of the Long Short-Term Memory (LSTM) architecture that introduces additional connections called "peepholes" between the cell state and the gates. These peephole connections allow the gates to directly observe the cell state, enhancing the model's ability to capture and utilize long-term dependencies in sequential data.

In a standard LSTM, the input gate, forget gate, and output gate only depend on the current input and the previous hidden state. Peephole connections enable the gates to also have visibility into the current cell state, providing more information for the gating mechanisms to make decisions

1. Bidirectional RNNs

Bidirectional Recurrent Neural Networks (RNNs) are a type of RNN architecture that processes input sequences in both forward and backward directions. Unlike traditional RNNs that only consider the past context, bidirectional RNNs take into account both the past and future context when making predictions or analyzing sequential data.

1. Explain the gates of LSTM with equations.

LSTM (Long Short-Term Memory) networks utilize gating mechanisms to control the flow of information and regulate the memory state. These gates, consisting of sigmoid and element-wise multiplication operations, determine how much information should be stored, forgotten, and outputted by the LSTM cell. The gates in LSTM include the input gate, forget gate, and output gate.

1. Explain BiLSTM

BiLSTM (Bidirectional Long Short-Term Memory) is a variant of the LSTM (Long Short-Term Memory) architecture that incorporates bidirectional processing of sequential data. It combines the power of LSTM with the ability to capture information from both past and future contexts. BiLSTM processes the input sequence in both the forward and backward directions simultaneously, allowing the network to capture dependencies in both directions.

The BiLSTM architecture consists of two LSTMs: one processing the sequence in the forward direction (from the first element to the last) and the other in the backward direction (from the last element to the first). Each LSTM maintains its own hidden state and processes the input independently.

1. Explain BiGRU

BiGRU (Bidirectional Gated Recurrent Unit) is a variant of the Gated Recurrent Unit (GRU) architecture that incorporates bidirectional processing of sequential data. Similar to BiLSTM, BiGRU combines the power of GRU with the ability to capture information from both past and future contexts. BiGRU processes the input sequence in both the forward and backward directions simultaneously, allowing the network to capture dependencies in both directions.