1. Explain the basic architecture of RNN cell.

RNN has a concept of “memory” which remembers all information about what has been calculated till time step t. RNNs are called recurrent because they perform the same task for every element of a sequence, with the output being depended on the previous computations.

2. Explain Backpropagation through time (BPTT)

Backpropagation through time (BPTT) is a gradient-based technique for training certain types of recurrent neural networks. It can be used to train Elman networks. The algorithm was independently derived by numerous researchers.

3. Explain Vanishing and exploding gradients

Exploding is the opposite of Vanishing and is when the gradient continues to get larger which causes a large weight update and results in the Gradient Descent to diverge. Exploding gradients occur due to the weights in the Neural Network, not the activation function.

4. Explain Long short-term memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behavior required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning

5. Explain Gated recurrent unit (GRU)

The Gated Recurrent Unit (GRU) is a type of Recurrent Neural Network (RNN) that, in certain cases, has advantages over long short term memory (LSTM). GRU uses less memory and is faster than LSTM, however, LSTM is more accurate when using datasets with longer sequences.

6. Explain Peephole LSTM

One popular LSTM variant, introduced by Gers & Schmidhuber (2000), is adding “peephole connections.” This means that we let the gate layers look at the cell state. In this peephole connection we can see that all the gates are having an input along with the cell state.

7. Bidirectional RNNs

To enable straight (past) and reverse traversal of input (future), Bidirectional RNNs, or BRNNs, are used. A BRNN is a combination of two RNNs - one RNN moves forward, beginning from the start of the data sequence, and the other, moves backward, beginning from the end of the data sequence.

8. Explain the gates of LSTM with equations.

LSTM uses three gates: input gate, forget gate, and output gate for processing.

9. Explain BiLSTM

Bidirectional Long Short-Term Memory (BiLSTM) [14] is a type of recurrent neural networks. It processes data in two directions, since it works with two hidden layers. This is the main point of divergence with LSTM. BiLSTM has proven good results in natural language processing

10. Explain BiGRU

A Bidirectional GRU, or BiGRU, is a sequence processing model that consists of two GRUs. one taking the input in a forward direction, and the other in a backwards direction. It is a bidirectional recurrent neural network with only the input and forget gates.