NLP Assigments-3

1. The basic architecture of an RNN cell consists of a single unit that takes in a current input and the hidden state from the previous time step, and produces an output and a new hidden state. The hidden state is a summary of the past inputs and is used to preserve information over a longer period of time. The output of the RNN cell is typically used to make a prediction or to decide on an action.
2. Backpropagation through time (BPTT) is an algorithm for training RNNs that involves unrolling the RNN and treating it as a feedforward neural network, and then applying the backpropagation algorithm to it. This allows the gradients to be propagated back through time and the weights of the RNN to be updated.
3. Vanishing gradients occur when the gradients of the weights in the network become very small, which can make it difficult for the network to learn effectively. This can happen when the weights are too small or when the activation function has a small derivative. Exploding gradients occur when the gradients of the weights in the network become very large, which can make the network unstable and cause learning to diverge.
4. Long short-term memory (LSTM) is a type of RNN that is designed to handle long-term dependencies more effectively than traditional RNNs. It does this by using gates to control the flow of information in and out of the hidden state.
5. Gated recurrent unit (GRU) is a type of RNN that is similar to LSTM, but it has fewer parameters and is simpler to train. It also uses gates to control the flow of information in and out of the hidden state.
6. Peephole LSTM is a variant of LSTM that uses additional connections to the gates that allow the gates to access the hidden state at the previous time step. This can improve the ability of the LSTM to retain long-term dependencies.
7. Bidirectional RNNs (BiRNNs) are RNNs that process the input sequence in two directions: forwards and backwards. The output of the BiRNN is the concatenation of the outputs from the forward and backward RNNs. BiRNNs can be useful for tasks that involve understanding the context of the input, such as language translation or named entity recognition.
8. The gates of an LSTM are controlled by sigmoid functions that decide what information should be passed through and what should be discarded. The input gate controls the flow of information into the cell, the forget gate controls the flow of information out of the cell, and the output gate controls the flow of information from the cell to the output. The equations for these gates are as follows:

Input gate: i = sigmoid(W\_i \* x + U\_i \* h + b\_i) Forget gate: f = sigmoid(W\_f \* x + U\_f \* h + b\_f) Output gate: o = sigmoid(W\_o \* x + U\_o \* h + b\_o)

1. BiLSTM is a bidirectional version of LSTM. It processes the input sequence in both the forward and backward directions, and the output is the concatenation of the outputs from the two LSTMs. BiLSTM can be useful for tasks that involve understanding the context of the input, such as language translation or named entity recognition.
2. BiGRU is a bidirectional version of GRU. It processes the input sequence in both the forward and backward directions,