**NLP Assignment 3 Solution**

**1. Explain the basic architecture of RNN cell.**

**Ans:** RNN cells are a type of neural network architecture commonly used in natural language processing (NLP) tasks.

RNN cell is composed of three main parts:

1. Input Gate: Responsible for receiving input, processing it and deciding whether to update the hidden state or not.
2. Output Gate: Responsible for generating output from the hidden state.
3. Hidden State: This is a vector which stores information from the previous time step. It is updated by the Input Gate with new information from the current time step. This hidden state is then used to generate output at the current time step.

**2. Explain Backpropagation through time (BPTT)**

**Ans:** Backpropagation Through Time (BPTT) is a type of supervised learning algorithm used to train recurrent neural networks (RNNs). It is an extension of the standard backpropagation algorithm, which is used to train feedforward neural networks. BPTT works by unrolling the RNN over time and then performing backpropagation over the unfolded network. This allows the algorithm to calculate the gradients of the cost function with respect to the weights of the RNN. These gradients can then be used to update the weights in order to minimize the cost function. BPTT is an effective algorithm for training RNNs, but it has the disadvantage of being computationally expensive and requiring a large amount of data.

**3. Explain Vanishing and exploding gradients**

**Ans:** Vanishing gradients refer to a phenomenon in neural networks where the gradient during backpropagation becomes increasingly smaller and smaller. This usually occurs when the model has many layers and can lead to the model’s weights and biases not updating as expected. This can cause the model to not learn as quickly and can even lead to the model not learning at all.

Exploding gradients refer to the opposite phenomenon in neural networks, where the gradient during backpropagation becomes increasingly larger and larger. This can happen when the model has many layers and can lead to the model’s weights and biases updating too quickly. This can cause the model to overfit the training data and can lead to the model not generalizing well.

**4. Explain Long short-term memory (LSTM)**

**Ans:** Long short-term memory (LSTM) is a type of recurrent neural network (RNN) that is capable of learning long-term dependencies. It is well-suited to tasks such as natural language processing and speech recognition, where long-term context is important. The LSTM architecture consists of memory cells, input gates, output gates, and forget gates, which help the network remember information over a long period of time. The memory cells are responsible for storing the information, while the gates control the flow of information into and out of the cells. LSTM networks are trained using backpropagation through time and have been used to achieve state-of-the-art performance on a variety of tasks.

**5. Explain Gated recurrent unit (GRU)**

**Ans:** Gated Recurrent Unit (GRU) is a type of Recurrent Neural Network (RNN) architecture that is similar to Long Short-Term Memory (LSTM) networks, but uses fewer parameters and has fewer layers. Like LSTM, GRU has a gating mechanism that controls the flow of data within the network. The GRU architecture consists of two gates: a reset gate and an update gate. The reset gate determines how much of the past information to forget, while the update gate decides how much of the new data to use in the current state. The GRU architecture is designed to learn long-term dependencies, allowing it to capture and store more contextual information from the input data, enabling it to make more accurate predictions.

**6. Explain Peephole LSTM**

**Ans:** Peephole LSTM is a type of Long Short-Term Memory (LSTM) network. It is an extension of the traditional LSTM network that allows for the direct connection between the current cell state and the current input gate. This direct connection allows the network to make use of more contextual information when making decisions about how to process incoming input. Additionally, peephole LSTM networks are able to better capture long-term dependencies in the data. This is particularly useful for tasks such as speech recognition and natural language processing.

**7. Bidirectional RNNs**

**Ans:** Bidirectional RNNs are recurrent neural networks that process data in both forward and backward directions. This allows them to capture and process long-term dependencies in both directions, which is beneficial for tasks such as language modeling and machine translation. Bidirectional RNNs are constructed by having two separate RNNs, one that processes the data in the forward direction and one that processes the data in the backward direction. The outputs of the two RNNs are then merged in some way, typically by concatenating them together. Bidirectional RNNs can significantly improve the performance of machine learning models on tasks such as language modeling and machine translation.

**8. Explain the gates of LSTM with equations.**

**Ans:** The gates of LSTM are used to control the flow of information into and out of the memory cell.

The gates are defined by the following equations:

Input Gate: i\_t = σ(W\_i · [h\_t-1, x\_t] + b\_i)

Forget Gate: f\_t = σ(W\_f · [h\_t-1, x\_t] + b\_f)

Output Gate: o\_t = σ(W\_o · [h\_t-1, x\_t] + b\_o)

Cell State: c\_t = f\_t \* c\_t-1 + i\_t \* tanh(W\_c · [h\_t-1, x\_t] + b\_c)

Hidden State: h\_t = o\_t \* tanh(c\_t)\*\*

**9. Explain BiLSTM**

**Ans:** BiLSTM (Bidirectional Long Short-Term Memory) is a recurrent neural network (RNN) architecture that processes input sequences in both directions with two separate hidden layers. It is a type of RNN that can remember information for long periods of time and can process both forward and backward sequences of data. It is a combination of two LSTMs, one processing the input sequence in forward direction and the other processing the sequence in backward direction. The output of both LSTMs are then combined and used as an input to a fully connected layer. BiLSTM is often used in natural language processing tasks such as sentiment analysis and language modeling.

**10. Explain BiGRU**

**Ans:** BiGRU (Bidirectional Gated Recurrent Units) is a type of recurrent neural network (RNN) architecture that processes input data in both directions, allowing the network to learn the context of a sequence of data more effectively. A BiGRU consists of two separate recurrent neural networks, one processing the input data in the forward direction and the other in the backward direction. The two networks then combine their output, allowing the network to learn the context of the entire sequence of data. This helps the network to better understand the data and is especially useful for tasks such as natural language processing and time series prediction.