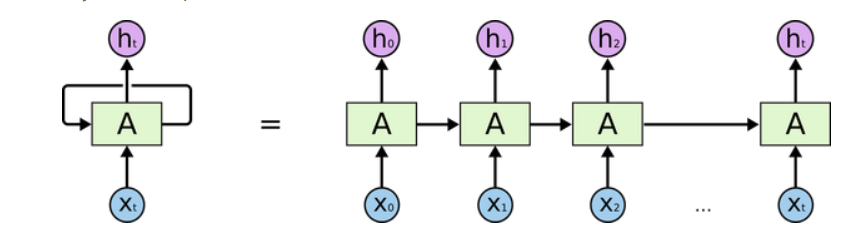
**NLP- ASSIGNMENT-3**

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



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

Backpropagation Through Time, or BPTT, is the application of the Backpropagation training algorithm to recurrent neural network applied to sequence data like a time series. A recurrent neural network is shown one input each timestep and predicts one output. Conceptually, BPTT works by unrolling all input timesteps.

1. **Explain Vanishing and exploding gradients**

Exploding gradient occurs when the derivatives or slope will get larger and larger as we go backward with every layer during backpropagation. This situation is the exact opposite of the vanishing gradients. This problem happens because of weights, not because of the activation function.

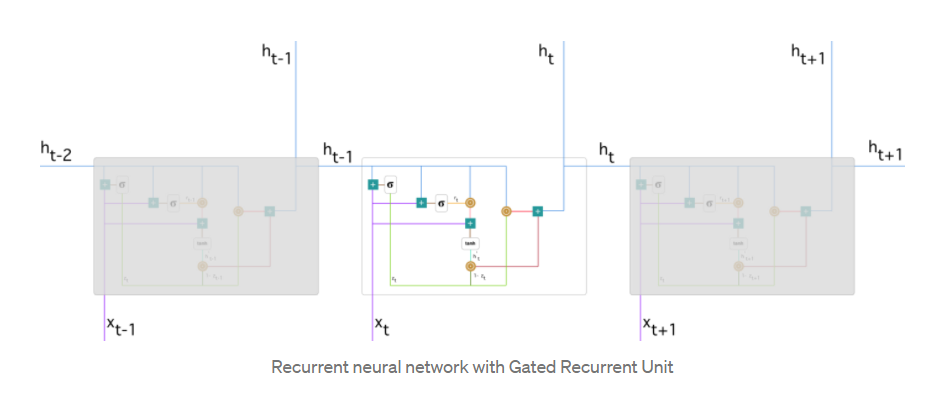
1. **Explain Long short-term memory (LSTM)**

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.

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

To solve the vanishing gradient problem of a standard RNN, GRU uses, so-called, update gate and reset gate. Basically, these are two vectors which decide what information should be passed to the output. The special thing about them is that they can be trained to keep information from long ago, without washing it through time or remove information which is irrelevant to the prediction.



1. **Explain Peephole LSTM.**

Peephole connections refer to a modification to the basic LSTM architecture. ... Surprisingly, LSTM augmented by “peephole connections” from its internal cells to its multiplicative gates can learn the fine distinction between sequences of spikes separated by either 50 or 49 discrete time steps.

1. **Bidirectional RNNs.**

Bidirectional recurrent neural networks (BRNN) connect two hidden layers of opposite directions to the same output. With this form of generative deep learning, the output layer can get information from past (backwards) and future (forward) states simultaneously. Invented in 1997 by Schuster and Paliwal,[1] BRNNs were introduced to increase the amount of input information available to the network.

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

LSTM stands for Long Short Term Memory, I myself found it difficult to directly understand LSTM without any prior knowledge of the Gates and cell state used in Long Short Term Memory neural networks so, this post is an attempt to get familier with a LSTM model which uses gates and cell state.

LSTM is made up of Gates:

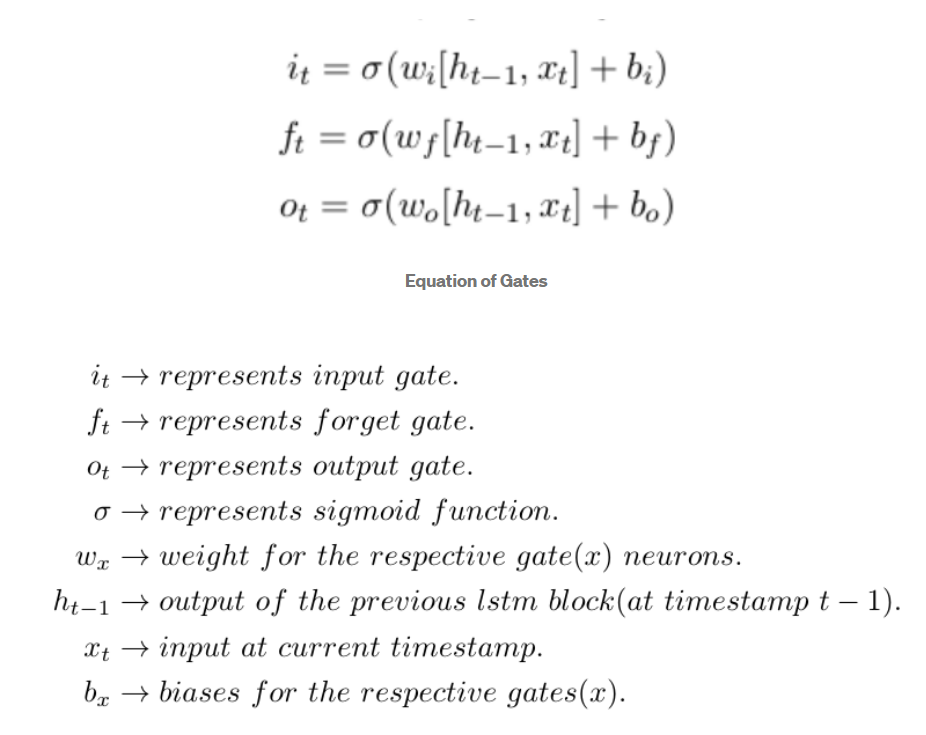
In LSTM we will have 3 gates:

1) Input Gate.

2) Forget Gate.

3) Output Gate.

The equations for the gates in LSTM are:



1. **Explain BiLSTM.**

A Bidirectional LSTM, or biLSTM, is a sequence processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backwards direction.

1. **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.