The core difference of RNN, Peephole-LSTM and GRU :

RNN are designed to work with sequential data. RNN uses the previous information in the sequence to produce the current output. In RNN weights and bias for all the nodes in the layer are same. The workflow of GRU is same as RNN but the difference is in the operations inside the GRU unit. In the GRU there are two unit. gates 1) Reset gate, 2) update gate. Each gates has it's own weights and biases. For all nodes in one layer are same. GRU has only one activation function. LSTM are pretty much similar to GRU's . But there are diffrences they are intended to solve the vanishing problem gradient problem.

LSTM has 4 gates 1) Reset gate 2) update gate 3) forget gate 4) output gate. It has two activation function.
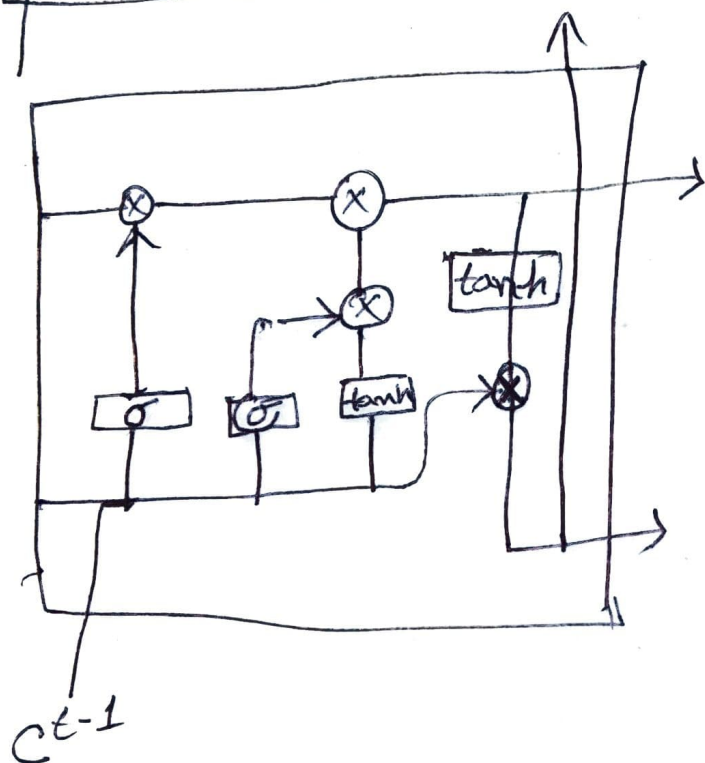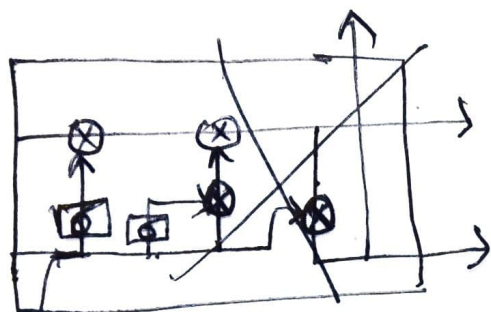
The Basic difference of LSTM forget gate and GRU's reset gate :

GRU Reset gate : The reset gate is used to decide whether the previous cell state is important or not. Sometimes the reset gate is not used in simple GRU.

LSTM Forget gate : It controls what is kept vs forgotten from previous cell state. In laymen terms it will decide how much information from the previous state should be kept and forget forget remaining.

P. T. O

Matrix workflow of the LSTM:

$$Z$$

$$\boxed{W}\,\boxed{\begin{array}{c} x^t \\ h^{t-1} \end{array}}$$

$$\boxed{z^i} = \sigma\left(\boxed{W^i}\,\boxed{\begin{array}{c} x^t \\ h^{t-1} \end{array}}\right)$$

$$\boxed{z^f} = \sigma\left(\boxed{W^f}\,\boxed{\begin{array}{c} x^t \\ h^{t-1} \end{array}}\right)$$

$$\boxed{z^o} = \sigma\left(\boxed{W^o}\,\boxed{\begin{array}{c} x^t \\ h^{t-1} \end{array}}\right)$$

c^{t-1}

Controls forget gate

Controls input gate

updating information

Control output gate

$$\boxed{z^f} \quad \boxed{z^i} \quad \boxed{z} \quad \boxed{z^o}$$

$$\boxed{h^{t-1}} \quad \boxed{x^t}$$