Ans to Q-2

Core difference of RNN, Peehole-LSTM and GRU;

RNN are designed to work with sequential data. Sequential data can be form of text, audio, video etc. 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 work flow of GRU is same as RNN but the difference is in the operation inside the GRU unit. Inside the GRU it has two gates. One is Reset gate another update gate. Each gates has it's own weights and biases. For all node in one layer weights and biases are same. GRU is less complex than LSTM because it has less number of gates. GRU exposes the complete memory and hidden layers.

LSTM's are pretty much similar to GRU's. They are intended to solve the vanishing gradient problem. It has 4 gates. Such as 1) Reset gate 2) update gate 3) Forgot gate. As it has more gates so it is complex. It doesn't expose the ~~bid~~ hidden layers and complete memory.

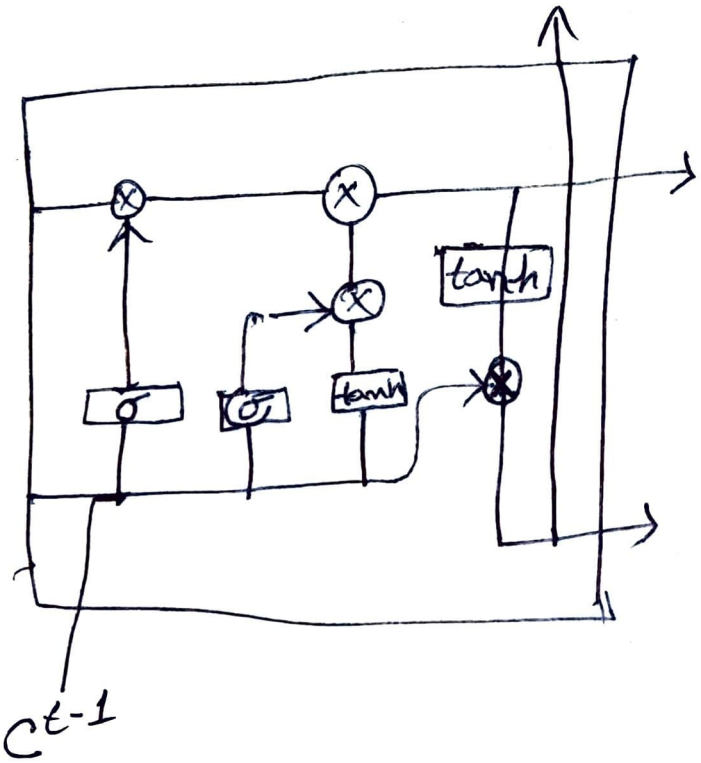
Differences of LSTM forgate Gate and GRU Reset gate:

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. It is used from the model to decide how much of the past data to forgate.

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

P.T.O

Matrix workflow of the LSTM:



$$z = W \begin{bmatrix} x^t \\ h^{t-1} \end{bmatrix}$$

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

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

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