Ans 2:

RNN
- (i) no cell state
- (ii) tan hyperbolic activation function
- (iii) has vanishing gradient problem

Peephole LSTM
- (i) Cell state / memory
- (ii) LSTM resolves the vanishing gradient problem
- (iii) 3 gates : input gate, forget gate, output gate
- (iv) concatenation of previous memory and peeping into it to access maximum information.

GRU
- (i) can resolve vanishing gradient problem
- (ii) no cell state
- (iii) 2 gates : reset gate, update gate

All RNNs have feedback loops in the layer which allows it to maintain information in 'memory' over time. But RNN is time-dependent and so has the drawback of vanishing gradient problem ~~as~~ because the gradient of the loss function decays exponentially with time.

LSTM units include a 'memory cell' that can maintain information for long periods of time. This architecture lets them learn longer-term dependencies and the 3 gates are used to control memory flow. GRUs are simpler and similar to LSTM. They use 2 gates to control the flow of information but they don't use separate memory cells.

GRU gives more compression. The reset gate denotes how much past information needs to be forget. That is, the reset gate is used from the model to decide how much of the past information is needed to neglect. It multiplies input vector and hidden state with their weights. Then, it calculates element-wise multiplication.

LSTM's forget gate decides what information should be kept or thrown away. Information from the previous hidden state and current input is passed through sigmoid function.

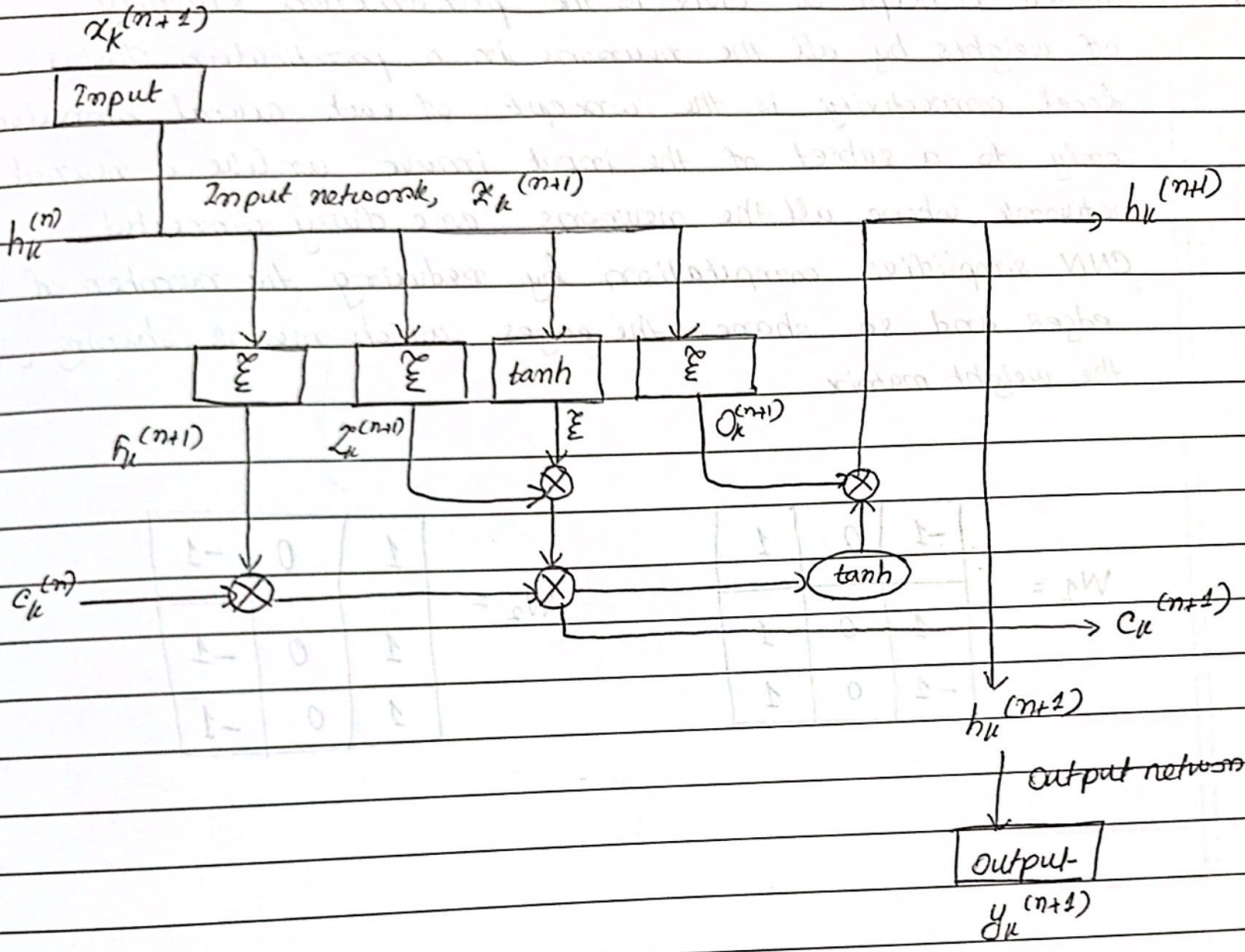

Fig: LSTM matrix workflow