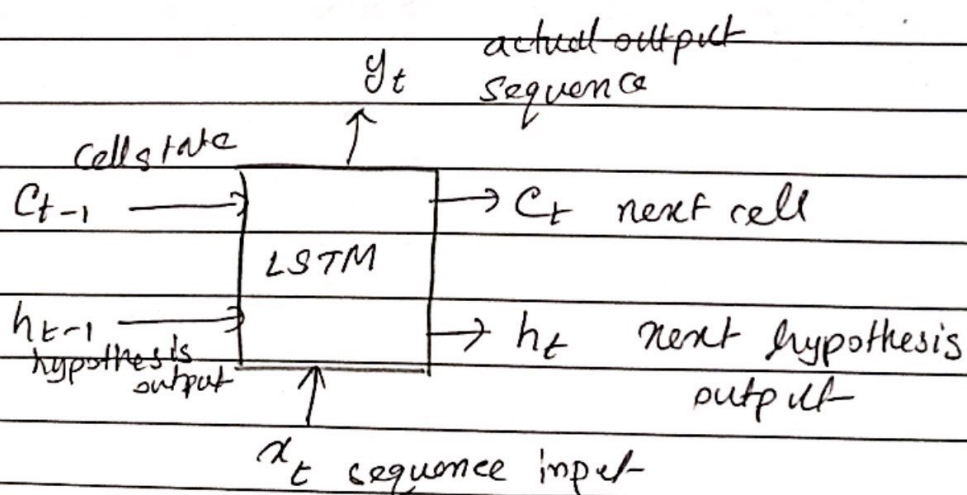


Ans 2: The difference lies in more flexible control. All RNNs have feedback loops in the recurrent layers which lets them maintain information in memory over time. LSTM is a variant of RNN that includes a memory cell that can maintain information for longer periods of time.

GRUs are similar to LSTM, but use a simplified structure that doesn't use separate memory cells and GRU uses fewer gates to control the flow of information.

GRU does not have an output gate unlike LSTM. GRU combines the input and forget gate into a single update gate. The Reset Gate of GRU is concerned with how much past information to forget or is needed to neglect.



$$y_t = \sigma(w h_t) \quad h_t = \sigma(w^h h^{t-1} + w^x x_t)$$

c changes slowly
 h changes fastly