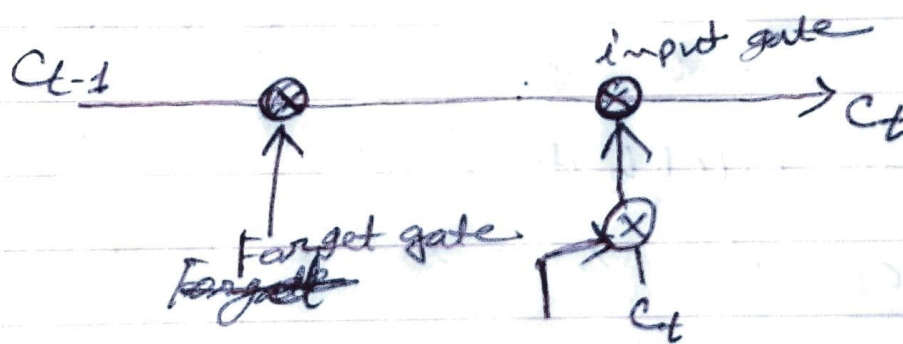@   Ans to Question No-4

vanishing gradient problem solve the shorting problem in RNN. When there is a long data set in RNN vanishing gradient work find time.

LSTM use sigmoid or tanh to solve the vanishing gradient problem.

Forget gate controls how much data to ~~same~~ remember. The sigmoid layer controls how much ~~each component~~ ~~sh~~ outputs in between 0-1 should each component get. If the number is 0 all the data will be forgotten but if it is 1 all the data will be passed.

$$c_t = f_t * c_{t-1} + i_t * \bar{c}_t$$

here, f_t can handle vanishing gradient problem.