Ans to the Ques no: 3

The update and reset gate of GRU are used to solve
the vanishing gradient problem of RNN.

~~Upat~~

Update gate determines the amount of information needs
to be passed from previous ~~step to~~ layer to next layer.
It is similar to output gate in LSTM

Reset gate is used to decide how much of previous
information needs to forget from the model. It is
similar to forget and input gate of LSTM

Work flow of GRU

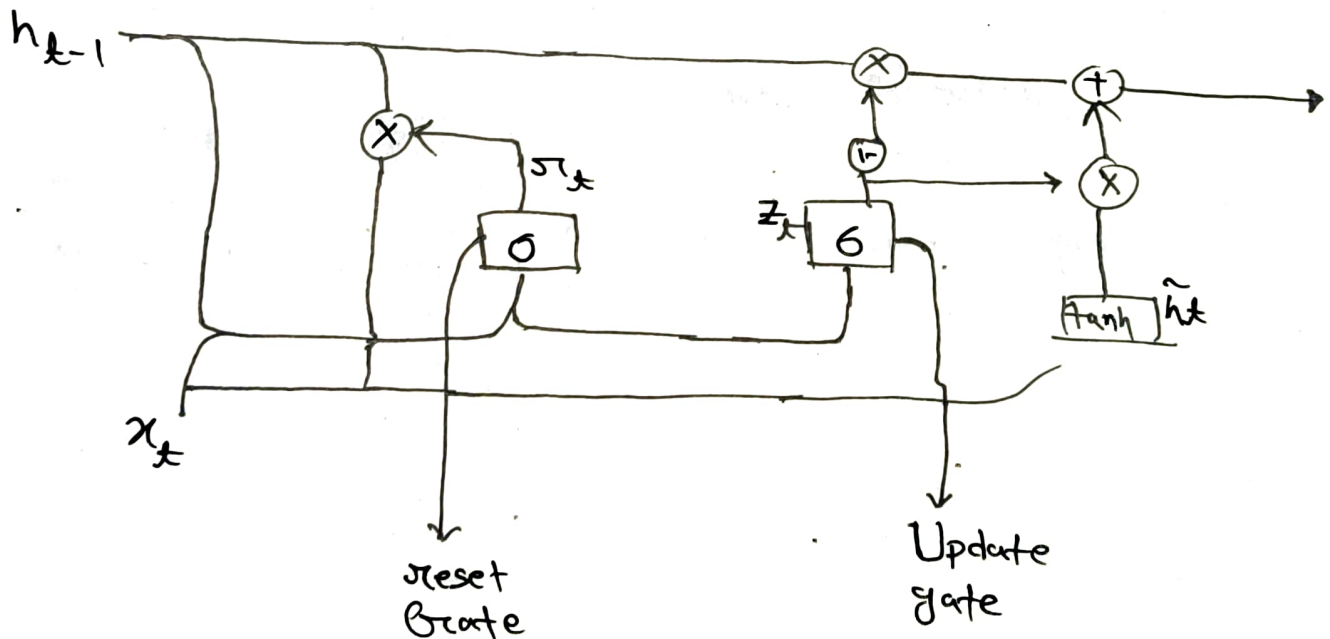$$r_t = \sigma\left(w^{(r)}x_t + U^{(r)} \cdot h_{t-1}\right)$$

$$z_t = \sigma\left(w^{(z)}x_t + U^{(z)} h_{t-1}\right)$$

$$\tilde{h}_t = \cancel{z_t \odot h_{t-1}} \tanh\left(w x_t + r_t \odot U h_{t-1}\right)$$

$$n_t = z_t \odot h_{t-1} + (1-z_t) \odot \tilde{h}_t$$

In GRU, we combine the input and forget get to use as single update gate

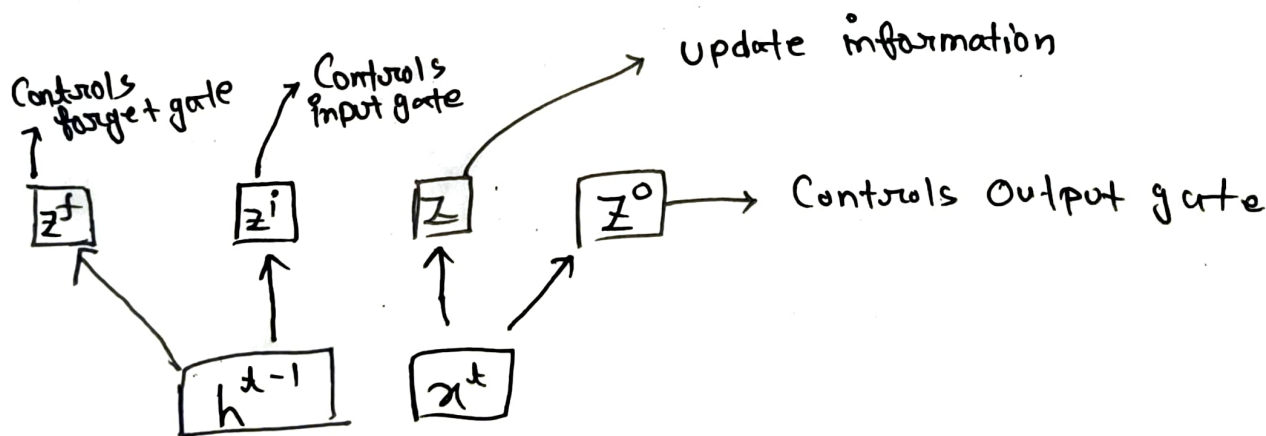GRU does not need cell layer. to pass values along the next Layer

The

Work Flow of LSTM

$$i_t = \sigma(w_i^T \cdot [x_t \cdot h_{t-1}])$$

$$\tilde{c}_t = \tanh[wc[h_{t-1}, x_t] + b_e]$$
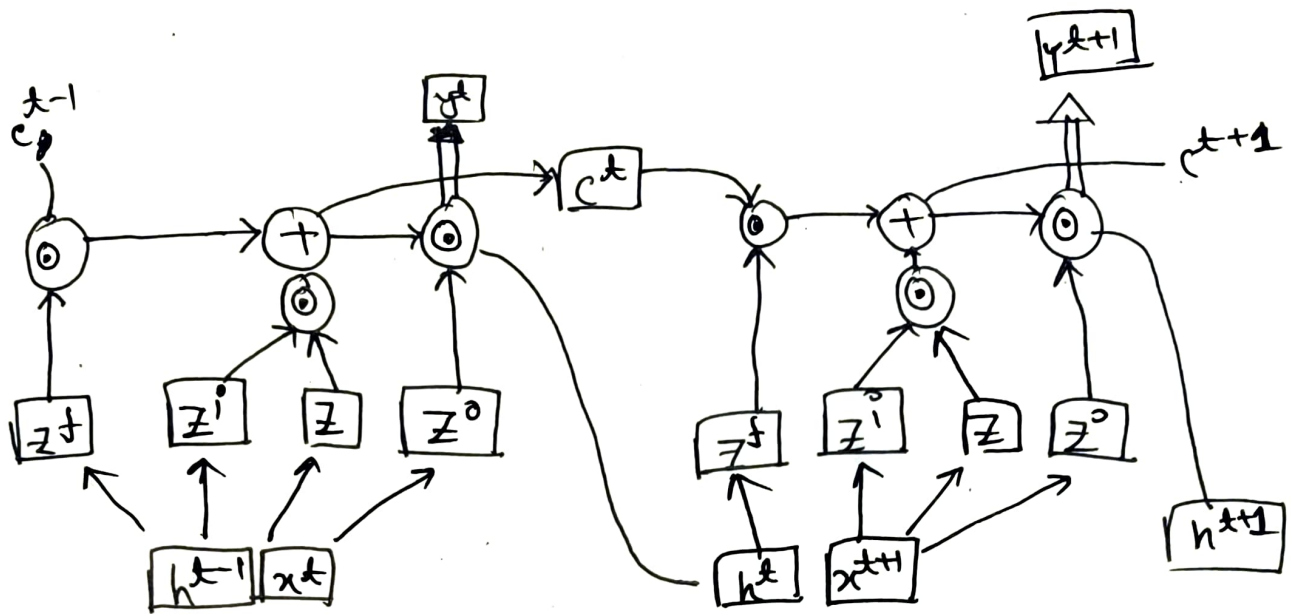
$$C_t = f_t * C_{t-1} + i_t * \tilde{c}_t$$

forget   forget   input
gate   Gate   Gate

Controller

$$f_t = \sigma(w_f^T [x_t, h_{t-1}])$$

Controls   Controls   update information
forget gate   input gate

Controls
z^f z^i z z^o $\longrightarrow$ Controls Output gate

h^{t-1} x^t

These ~~matrix~~ co

The computation of these matrices should be done concurrently

LSTM needs cell layer to pass values along with next layer.

Reasons behind the faster feature of GRU in compare to RNN:

1. GRU uses less training parameters which enables it to work in less memory. The concept of less memory usage makes it faster than RNN.

2. GRU is simpler which makes it easier to modify. On the contrary, RNN have feedback loops in reccourrent layers for which it becomes difficult to modify.