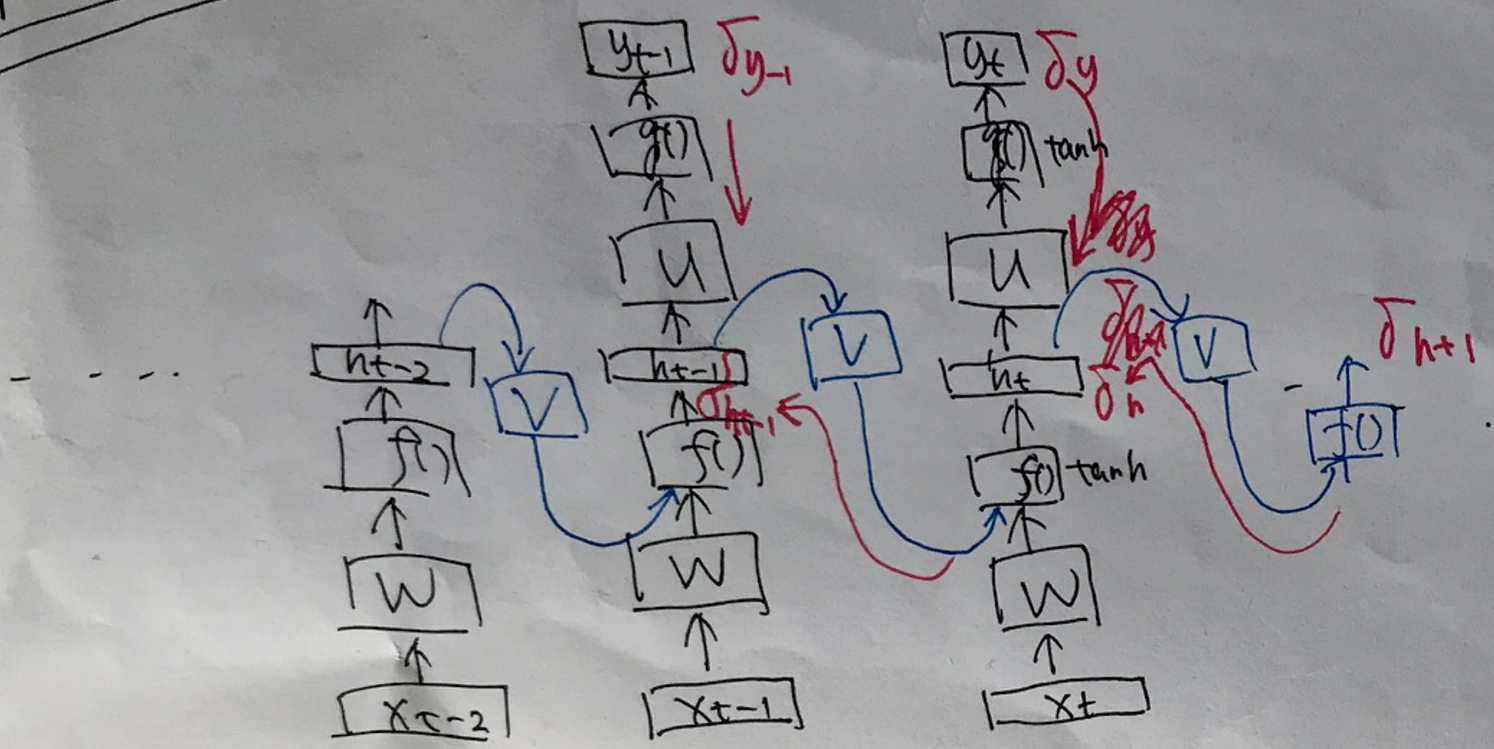
RNN

$$h_t = f(Wx_t + Vh_{t-1} + b_h)$$

$$y_t = g(Uh_t + b_y)$$



$$\frac{\partial L}{\partial U} = \underbrace{\frac{\partial L}{\partial a_y}\cdot\frac{\partial a_y}{\partial z_y}}_{\text{error term}} \cdot \frac{\partial z_y}{\partial U} \quad (= h_t) = \underbrace{L'g'}_{\delta_y} h_t$$

$$z_y = Uh_t + b_y$$
$$a_y = g(z)$$
(same as FFNN)

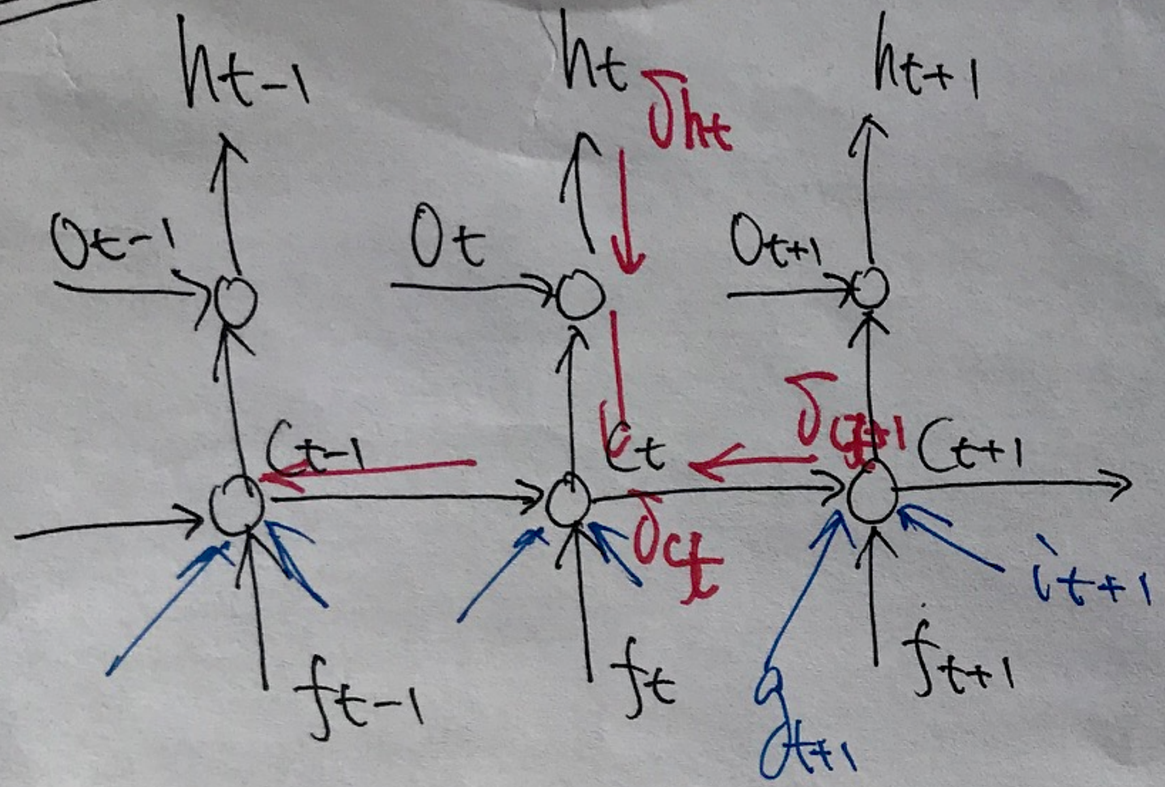
$$\frac{\partial L}{\partial W} = \underbrace{\frac{\partial L}{\partial a_h}\cdot\frac{\partial a_h}{\partial z_h}}_{\text{error term}} \cdot \frac{\partial z_h}{\partial W} \quad (= x_t) = \delta_h x_t$$

$$z_h = Wx_t + Vh_{t-1} + b_h$$
$$a_h = f(z)$$

$$\frac{\partial L}{\partial V} = \underbrace{\frac{\partial L}{\partial a_h}\cdot\frac{\partial a_h}{\partial z_h}}_{\text{error term}} \cdot \frac{\partial z_h}{\partial V} \quad (= h_{t-1}) = \delta_h h_{t-1}$$

$$\delta_h = g'U\delta_y + f'V\delta_{h+1}$$

LSTM



$$\delta C_{t+1} = \delta h_t \odot \tanh'(C_t) \odot O_t + \delta C_{t+1} \odot f_{t+1}$$

The gating mechaisms allow for gradient related to the meonory cell C_t to stay high across long time ranges.