Gate

$\odot$ – pointwise multiplication

σ – sigmoid. function.

W^{x_t}

$$\begin{bmatrix} 0.7 \\ 0.2 \\ 0.9 \\ 0.7 \end{bmatrix} \odot \begin{bmatrix} 0.01 \\ 0.99 \\ 0.8 \\ 0.01 \end{bmatrix} = \begin{bmatrix} 0.7 \times 0.01 \\ 0.2 \times 0.99 \\ 0.9 \times 0.8 \\ 0.7 \times 0.01 \end{bmatrix}$$

h_{t-1} f h_t

$func(x_t) \sim tanh(Wx_t)$ ~~else~~

$func(x_t, h_{t-1}) \sim tanh(W^x x_t + W^h h_{t-1})$

Perplexity

$$-\frac{1}{4}\left(\ln\left(\tfrac{1}{4}\right) + \ln\left(\tfrac{1}{3}\right) + \ln\left(\tfrac{1}{4}\right) + \ln\left(\tfrac{1}{3}\right)\right) = 1.242$$

$$e^{1.242} = 3.464.$$

RNN
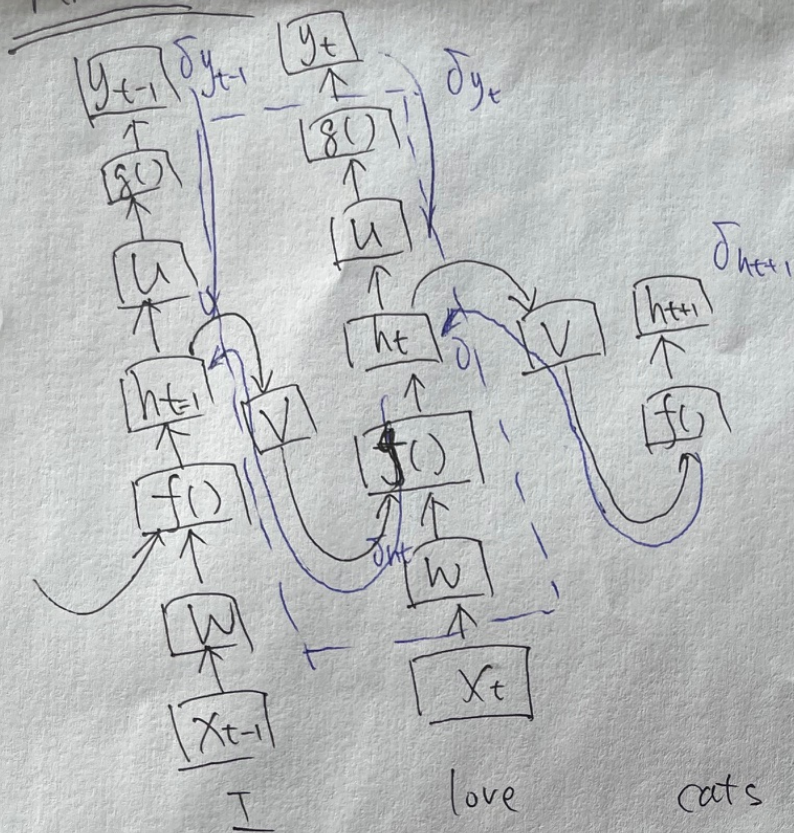
$$h_t = f(W x_t + V h_{t-1} + b_h)$$
$$y_t = g(U h_t + b_y)$$

I   love   cats

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

$$z_y = U h_t + b_y$$
$$a_y = g(z_y)$$
(same as FFNN)

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

$$z_h = W x_t + V h_{t-1} + b_h$$
$$a_h = f(z_h)$$

$$\frac{\partial L}{\partial V} = \frac{\partial L}{\partial a_h} \cdot \frac{\partial a_h}{\partial z_h} \cdot \frac{\partial z_h}{\partial V} = \delta_h h_{t-1}$$

$$\delta_{h_t} = g' U \delta_{y_t} + f' V \delta_{h_{t+1}}$$

LSTM: $\delta_{c_t} = \delta_{h_t} \odot \tanh'(c_t) \odot o_t + \delta_{c_{t-1}} \odot f_{t+1}$