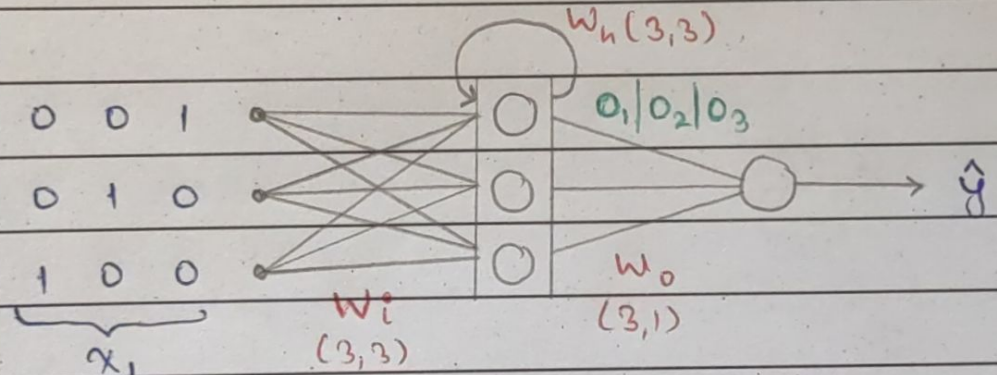
RAISONI GROUP
— a vision beyond —

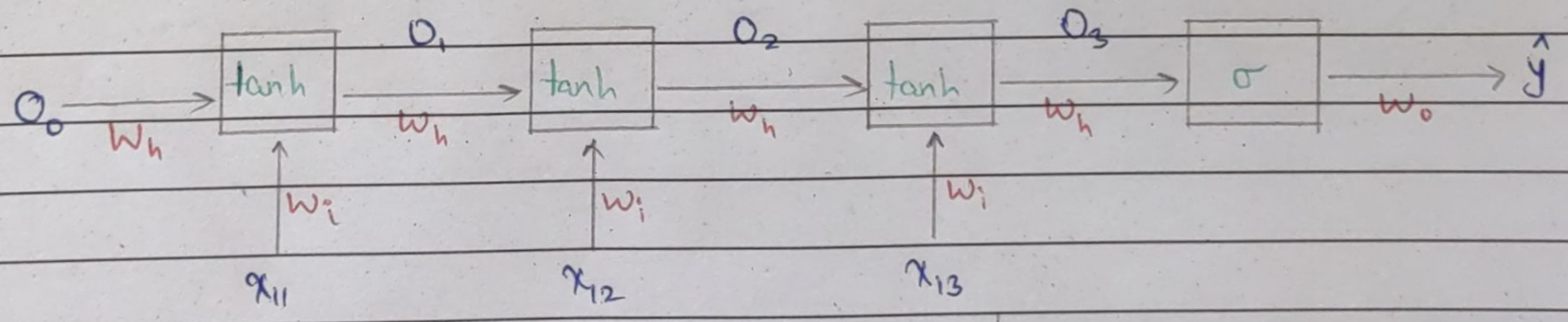
BACKPROPAGATION IN RNN

Backpropagation Through Time

Text			X	Y
cat mat rat	1		x_1 [100] [010] [001]	1
rat rat mat	1	$\Rightarrow$	x_2 [001] [001] [010]	1
mad mat cat	0		x_3 [010] [010] [100]	0

Vocab → cat mat rat
[100] [010] [001]





$O_1 = \tanh(x_{11} w_i + O_0 w_h)$

$O_2 = \tanh(x_{12} w_i + O_1 w_h)$

$O_3 = \tanh(x_{13} w_i + O_2 w_h)$

$\hat{y} = \sigma(O_3 w_o)$

Loss $L = -y_i \log(\hat{y}_i) - (1 - y_i) \log(1 - \hat{y}_i)$

Minimize loss through gradient descent.

$$w_o = w_o - \eta \frac{\partial L}{\partial w_o}$$

$$w_i = w_i - \eta \frac{\partial L}{\partial w_i}$$

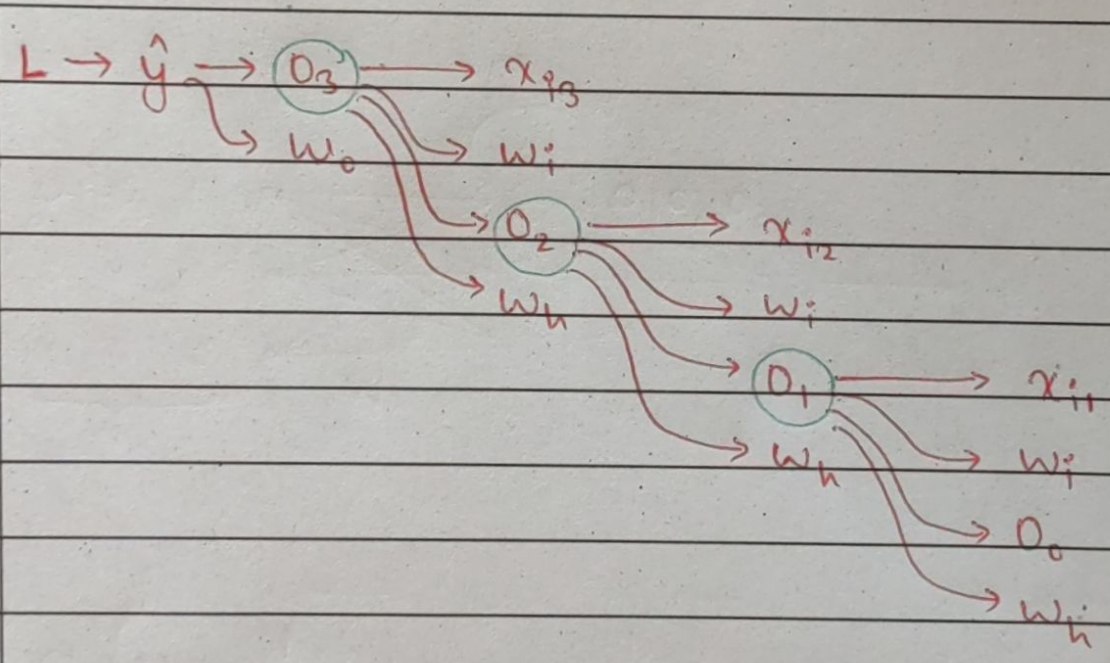
$$w_h = w_h - \eta \frac{\partial L}{\partial w_h}$$

$$\frac{\partial L}{\partial w_0} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial w_0}$$

$$\frac{\partial \hat{y}}{\partial w_0} = \sigma(1-\sigma)$$

$$\frac{\partial L}{\partial \hat{y}} = \frac{-y}{\hat{y}} + \frac{(1-y)}{(1-\hat{y})} = \frac{\hat{y}\hat{y} - y - y\hat{y} + \hat{y}}{\hat{y}(1-\hat{y})} = \frac{\hat{y}-y}{\hat{y}(1-\hat{y})}$$

$$\Rightarrow \frac{\partial L}{\partial w_0} = \frac{\hat{y}-y}{\hat{y}(1-\hat{y})} \times \sigma(1-\sigma)$$



$$\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial O_3} \times \frac{\partial O_3}{\partial O_2} \times \frac{\partial O_2}{\partial O_1} \times \frac{\partial O_1}{\partial w_i}$$

$$+ \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial O_3} \times \frac{\partial O_3}{\partial O_2} \times \frac{\partial O_2}{\partial w_i}$$

$$+ \frac{\partial L}{\partial \hat{y}} \times \frac{\partial \hat{y}}{\partial O_3} \times \frac{\partial O_3}{\partial w_i}$$

n = time steps

$$\frac{\partial L}{\partial w_i} = \sum_{j=1}^{h} \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial O_j} \frac{\partial O_j}{\partial w_i}$$

$$\Rightarrow \frac{\partial \hat{y}}{\partial O_1} \times \frac{\partial O_1}{\partial w_i} = \frac{\partial \hat{y}}{\partial O_3} \times \frac{\partial O_3}{\partial O_2} \times \frac{\partial O_2}{\partial O_1} \times \frac{\partial O_1}{\partial w_i}$$

$$\Rightarrow \frac{\partial \hat{y}}{\partial O_2} \times \frac{\partial O_2}{\partial w_i} = \frac{\partial \hat{y}}{\partial O_3} \times \frac{\partial O_3}{\partial O_2} \times \frac{\partial O_2}{\partial w_i}$$

$$\Rightarrow \frac{\partial \hat{y}}{\partial O_1} \times \frac{\partial O_1}{\partial w_1}$$

$$\frac{\partial L}{\partial w_h} = \frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial o_2} * \frac{\partial o_2}{\partial o_1} * \frac{\partial o_1}{\partial w_h}$$

$$+ \frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial o_2} * \frac{\partial o_2}{\partial w_h}$$

$$+ \frac{\partial L}{\partial \hat{y}} * \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial w_h}$$

$$\frac{\partial L}{\partial w_h} = \frac{\partial L}{\partial \hat{y}} \sum_{j=1}^{n} \frac{\partial \hat{y}}{\partial o_j} \frac{\partial o_j}{\partial w_h}$$

$$\Rightarrow \frac{\partial \hat{y}}{\partial o_1} * \frac{\partial o_1}{\partial w_h} = \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial o_2} * \frac{\partial o_2}{\partial o_1} * \frac{\partial o_1}{\partial w_h}$$

$$\Rightarrow \frac{\partial \hat{y}}{\partial o_2} * \frac{\partial o_2}{\partial w_h} = \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial o_2} * \frac{\partial o_2}{\partial w_h}$$

$$\Rightarrow \frac{\partial \hat{y}}{\partial o_3} * \frac{\partial o_3}{\partial w_h}$$