Backpropagation in RNN

01 December 2022 16:43

RNN intro → RNN → practical → types of RNN

→ RNN → Backprop → $\boxed{BPTT}$

Many to One RNN

Sentiment Analysis

text → 1/0

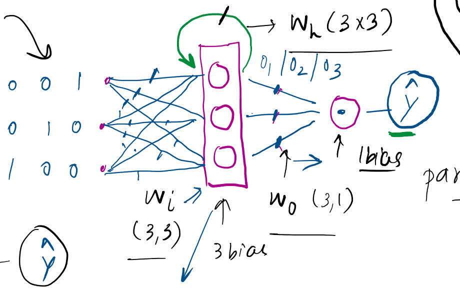
$\hat{y} \to L$

text	
cat mat rat	1
rat rat mat	1
mat mat cat	0

	o_1	o_2	o_3	$\hat{y}$	y
x_1	[1 0 0]	[0 1 0]	[0 0 1]		1
x_2	[0 0 1]	[0 0 1]	[0 1 0]		1
x_3	[0 1 0]	[0 1 0]	[1 0 0]		0

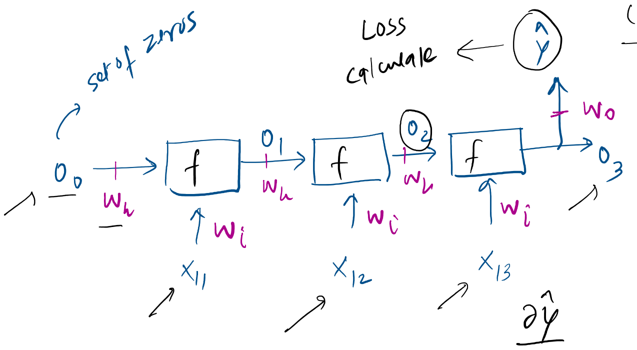
L → w_i, w_h, w_o

vocab → Cat Mat Rat
[1 0 0] [0 1 0] [0 0 1]



parameters

forward prop



$o_1 = f(x_{11} w_i + o_0 w_h)$

$o_2 = f(x_{12} w_i + o_1 w_h)$

$o_3 = f(x_{13} w_i + o_2 w_h)$

$\frac{\partial \hat{y}}{\partial w_o}$

$\hat{y} = \sigma(o_3 w_o)$

loss calculate → minimize

y

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

L min gradient descent

$\frac{\partial L}{\partial \hat{y}}$

w_i w_h w_o

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

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

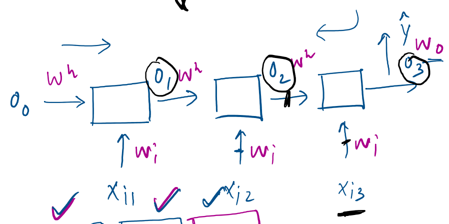
$w_o = w_o - \eta \boxed{\frac{\partial L}{\partial w_o}}$

Backprop

unfold → 3 times

$\frac{\partial L}{\partial w_o}$

$L \to \hat{y} \to o_3,\ w_o$



$$\frac{\partial L}{\partial w_o}, \quad \frac{\partial L}{\partial w_i}, \quad \frac{\partial L}{\partial w_h}$$

$$o_1 = f(x_{i1} w_i + o_0 w_h)$$

$$o_2 = f(x_{i2} w_i + o_1 w_h)$$

$$o_3 = f(x_{i3} w_i + o_2 w_h)$$

$$\hat{y} = \sigma(o_3 w_o)$$

10 words

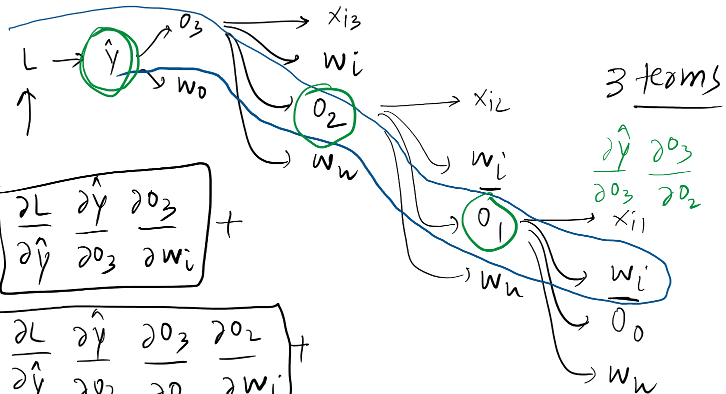
10 terms

$t=1$

$\frac{\partial L}{\partial w_o}$ $L \to \hat{y} \to o_3, w_o$

$$\boxed{\frac{\partial L}{\partial w_o} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial w_o}}$$

$\frac{\partial L}{\partial w_i}$



3 terms

$$\frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial o_2}$$

$$\frac{\partial L}{\partial w_i} = \boxed{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial w_i}} + \boxed{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial o_2} \frac{\partial o_2}{\partial w_i}} + \boxed{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial o_2} \frac{\partial o_2}{\partial o_1} \frac{\partial o_1}{\partial w_i}}$$

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

$$\frac{\partial L}{\partial w_i} = \sum_{j=1}^{3} \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_j} \frac{\partial o_j}{\partial w_i}$$

$$\boxed{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial w_i}}$$

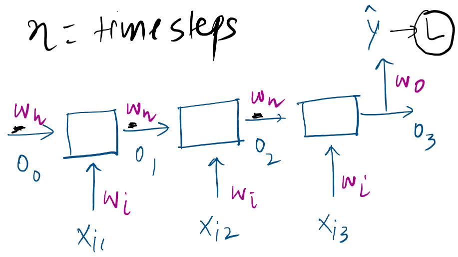
$$= \frac{\partial L}{\partial \hat{y}} \boxed{\frac{\partial \hat{y}}{\partial o_1}} \frac{\partial o_1}{\partial w_i} = \boxed{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial o_2} \frac{\partial o_2}{\partial o_1} \frac{\partial o_1}{\partial w_i}}$$

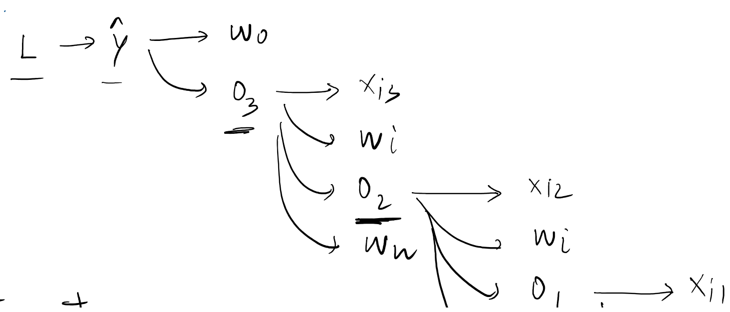
$$\frac{\partial L}{\partial \hat{y}} \boxed{\frac{\partial \hat{y}}{\partial o_2}} \frac{\partial o_2}{\partial w_i} = \boxed{\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial o_2} \frac{\partial o_2}{\partial w_i}}$$

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

n = timesteps

$$\boxed{\frac{\partial L}{\partial w_h}}$$





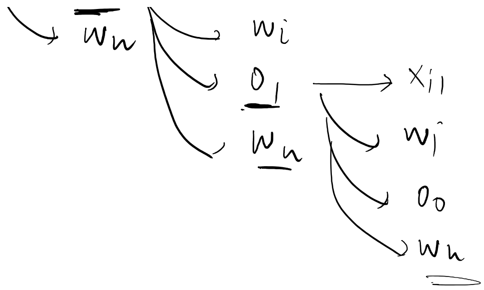
$$\frac{\partial L}{\partial w_h} = \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial w_h} +$$

$\partial L \quad \partial \hat{y} \quad \partial o_2 \quad \partial o_2 \quad +$

$$\partial w_h \quad \partial \hat{y} \quad \partial o_3 \quad \partial w_h$$

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

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



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

n = timesteps

for $j=3$

$$\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial w_h} \longrightarrow \frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_3} \frac{\partial o_3}{\partial o_1} \frac{\partial o_1}{\partial w_h}$$

for $j=10$

$$\frac{\partial L}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial o_{10}} \frac{\partial o_{10}}{\partial o_1} \frac{\partial o_1}{\partial w_h} \quad \boxed{\prod_{t=2}^{j} \frac{\partial o_t}{\partial o_{t-1}}}$$

$$\prod_{t=2}^{t=3} \frac{\partial o_t}{\partial o_{t-1}} = \frac{\partial o_2}{\partial o_1} \frac{\partial o_3}{\partial o_2}$$

$$o_t = f(x_{it} w_{inp} + o_{t-1} w_h)$$

$$\frac{\partial o_t}{\partial o_{t-1}} = \prod_{t=2}^{j} f'(x_{it} w_{inp} + o_{t-1} w_h)\, w_h$$

$$\downarrow$$

$$[0-1]$$