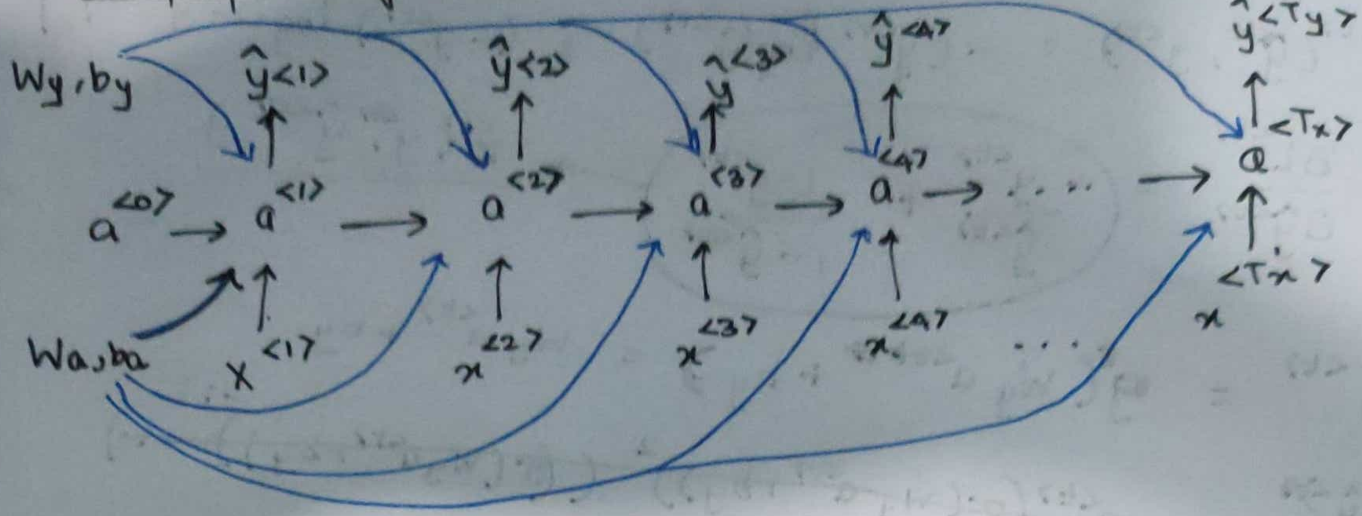
Backprop through time



in this case $T_x = T_y$

$$a^{<0>} = 0$$

$$a^{<t>} = g(\overset{\overset{V}{\uparrow}}{W_{aa}} a^{<t-1>} + \overset{\overset{u}{\uparrow}}{W_{ax}} x^{<t>} + b_a)$$

$$\hat{y}^{<t>} = g^*(W_y a^{<t>} + b_y)$$

$$a^{<t>} = g(W_a [a^{<t-1>}, x^{<t>}] + b_a)$$

[W_a is stacked W_{aa}, W_{ax} $[a^{<t-1>}, x^{<t>}]$ is stacked matrix $a^{<t-1>}, x^{<t>}$]

$$\hat{y}^{<t>} = g^*(W_y a^{<t>} + b_y)$$

$$L^{<t>}(\hat{y}^{<t>}, y^{<t>}) = -\left(y^{<t>} \log(\hat{y}^{<t>}) + (1 - y^{<t>}) \log(1 - \hat{y}^{<t>})\right)$$

$$L(\hat{y}, y) = \sum_{t=1}^{T_y} L^{<t>}(\hat{y}^{<t>}, y^{<t>})$$

$$a^{<t>} = g(V a^{<t-1>} + u x^{<t>} + b_a)$$

need to find

$$\frac{\partial L}{\partial W_y}, \frac{\partial L}{\partial V}, \frac{\partial L}{\partial u}$$

Let

$$g = g^* = I$$

identity function

$$L^{<t>}(\hat{y}^{<t>}, y^{<t>}) = -\left(y^{<t>}\log(\hat{y}^{<t>}) + (1-y^{<t>})\log(1-\hat{y}^{<t>})\right)$$

$$\frac{\partial L^{<t>}}{\partial \hat{y}^{<t>}} = \left(-\frac{y^{<t>}}{\hat{y}^{<t>}} + \frac{1-y^{<t>}}{1-\hat{y}^{<t>}}\right) \rightarrow \underline{z^{<t>}(y^{<t>}, \hat{y}^{<t>})}$$

$$\hat{y}^{<t>} = g^{*}(W_y a^{<t>} + b_y) = W_y a^{<t>} + b_y$$

$$\cancel{\frac{\partial \hat{y}^{<t>}}{\partial W_y} = a^{<t>}(\sigma(W_y a^{<t>} + b_y))^2 \left((\sigma(W_y a^{<t>} + b_y))^{-1} - 1\right)}$$

$$\cancel{= a^{<t>}\left((\hat{y}^{<t>})^{-1} - 1\right)(\hat{y}^{<t>})^2}$$

$$\frac{\partial \hat{y}^{<t>}}{\partial W_y} = a^{<t>}$$

$$\frac{\partial L^{<t>}}{\partial W_y} = \frac{\partial L^{<t>}}{\partial \hat{y}^{<t>}} \frac{\partial \hat{y}^{<t>}}{\partial W_y} = \boxed{z^{<t>} a^{<t>} = \frac{\partial L^{<t>}}{\partial W_y}}$$

1

$$\frac{\partial \hat{y}^{<t>}}{\partial b_y} = 1$$

2

$$\boxed{\frac{\partial L^{<t>}}{\partial b_y} = z^{<t>}}$$

$$a^{<t>} = v a^{<t-1>} + u x^{<t>} + b_a$$

$$\frac{\partial a^{<t>}}{\partial v} = a^{<t-1>} + v\frac{\partial a^{<t-1>}}{\partial v}, \quad \frac{\partial a^{<t>}}{\partial u} = x^{<t>} + v\frac{\partial a^{<t-1>}}{\partial u}, \quad \frac{\partial a^{<t>}}{\partial b_a} = 1$$

$$\frac{\partial \hat{y}^{<t>}}{\partial a^{<t>}} = W_y$$

$$\frac{\partial \hat{y}^{<t>}}{\partial a^{<t>}} \frac{\partial a^{<t>}}{\partial v} = \frac{\partial \hat{y}^{<t>}}{\partial v} = W_y\, a^{<t-1>} + W_y v \frac{\partial a^{<t-1>}}{\partial v}$$

$$\frac{\partial \hat{y}^{<t>}}{\partial u} = W_y x^{<t>} + W_y v \frac{\partial a}{\partial u}$$

$$\frac{\partial \hat{y}^{<t>}}{\partial b_a} = W_y$$

$$\frac{\partial a^{<t-1>}}{\partial v} = a^{t-2} + v \frac{\partial a^{<t-2>}}{\partial v}$$

$$\frac{\partial a^{<t-1>}}{\partial u} = x^{<t-1>} + v \frac{\partial a^{<t-2>}}{\partial u}$$

$$\vdots$$

$$\Rightarrow \frac{\partial \hat{y}^{<t>}}{\partial v} = W_y\, a^{<t-1>} + W_y v\, a^{<t-2>} + W_y v^2 a^{<t-3>} + \dots + W_y v^{t-1} a^{<0>}$$

$$\frac{\partial \hat{y}^{<t>}}{\partial u} = W_y\, x^{<t>} + W_y v\, x^{<t-1>} + W_y v^2 x^{<t-2>} + \dots + W_y v^{t-1} x^{<1>}$$

$$\frac{\partial \hat{y}^{<t>}}{\partial b_a} = W_y$$

$$\Rightarrow \boxed{\frac{\partial L^{<t>}}{\partial v} = \frac{\partial L^{<t>}}{\partial \hat{y}^{<t>}} \frac{\partial \hat{y}^{<t>}}{\partial v} = z^{<t>} W_y \left(\sum_{i=1}^{t} v^{t-i} a^{<i-1>} \right)} \quad (3)$$

$$\boxed{\frac{\partial L^{<t>}}{\partial u} = z^{<t>} W_y \left(\sum_{i=1}^{t} v^{(t-i)} x^{<i>} \right)} \quad (4)$$

$$\boxed{\frac{\partial L^{<t>}}{\partial b_a} = z^{<t>} W_y} \quad (5)$$

$$L = \sum_{t=1}^{T_x} L^{<t>}$$

can be gotten from

$$\frac{\partial L}{\partial u} = \sum_{t=1}^{T_x} \frac{\partial L^{<t>}}{\partial u}$$ 4

$$\frac{\partial L}{\partial v} = \sum_{t=1}^{T_x} \frac{\partial L^{<t>}}{\partial v}$$ 3

$$\frac{\partial L}{\partial b_a} = \sum_{t=1}^{T_x} \frac{\partial L^{<t>}}{\partial b_a}$$ 5

$$\frac{\partial L}{\partial W_y} = \sum_{t=1}^{T_x} \frac{\partial L^{<t>}}{\partial W_y}$$ 1

$$\frac{\partial L}{\partial b_y} = \sum_{t=1}^{T_x} \frac{\partial L^{<t>}}{\partial b_y}$$ 2