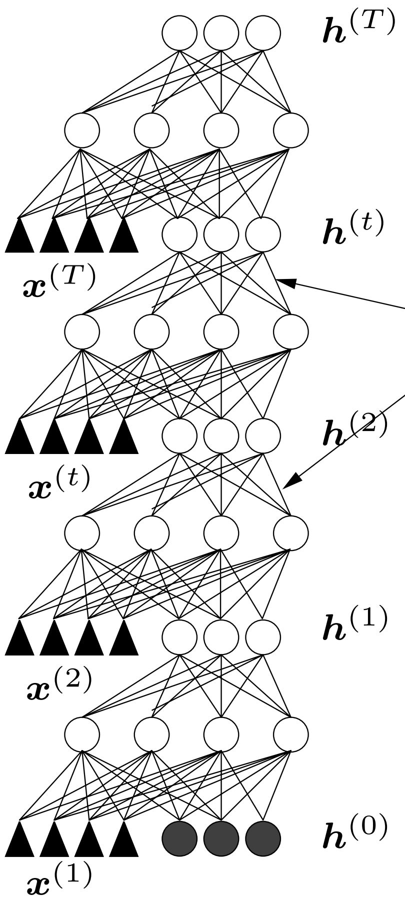
Back-Propagation Though Time

How to Learn: Time Unfolding



The last state is an adaptive encoding of the whole sequence

Weights are shared (replicated)

The unfolded network has a feedforward structure: e.g., gradient can be computed by Backpropagation (through time)

Example of Back-Propagation Through Time

Learning task

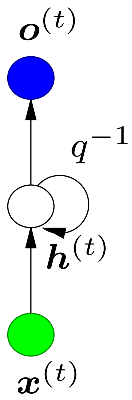
target sequence

$$\underset{t=1}{1} \leftarrow \underset{t=2}{-1} \leftarrow \underset{t=3}{1}$$

input sequence

$$\underset{t=1}{\begin{bmatrix}1\\0\end{bmatrix}} \leftarrow \underset{t=2}{\begin{bmatrix}0.5\\-0.2\end{bmatrix}} \leftarrow \underset{t=3}{\begin{bmatrix}0\\1\end{bmatrix}}$$

Recursive network



Equations

$$\begin{aligned}
\boldsymbol{h}^{(0)} &= \mathbf{0}, \\
\boldsymbol{a_h}^{(t)} &= \boldsymbol{U}\boldsymbol{x}^{(t)} + \boldsymbol{W}\boldsymbol{h}^{(t-1)} + \boldsymbol{b}, \\
\boldsymbol{h}^{(t)} &= tanh(\boldsymbol{a_h}^{(t)}), \\
a_o^{(t)} &= \boldsymbol{V}\boldsymbol{h}^{(t)} + c, \\
o^{(t)} &= tanh(a_o^{(t)}), \\
L &= \textstyle\sum_t e^{(t)} = \sum_t (y^{(t)} - o^{(t)})^2
\end{aligned}$$

2 input units, 2 hidden units, 1 output unit →

$$\boldsymbol{U} = \begin{bmatrix}u_{11} & u_{12}\\ u_{21} & u_{22}\end{bmatrix}, \quad \boldsymbol{W} = \begin{bmatrix}w_{11} & w_{12}\\ w_{21} & w_{22}\end{bmatrix}, \quad \boldsymbol{V} = \begin{bmatrix}v_1 & v_2\end{bmatrix}$$

$$\boldsymbol{h}^{(t)} = \begin{bmatrix}h_1^t\\ h_2^t\end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix}b_1\\ b_2\end{bmatrix}, \quad c$$

Example of Back-Propagation Through Time

$$
\begin{array}{rcccccc}
time & & t=1 & & t=2 & & t=3 \\
target & & 1 & & -1 & & 1 \\
 & & \downarrow & & \downarrow & & \downarrow \\
error & & e^{(1)} & & e^{(2)} & & e^{(3)} \\
 & & \uparrow & & \uparrow & & \uparrow \\
output & & o^{(1)} & & o^{(2)} & & o^{(3)} \\
 & & \uparrow \boldsymbol{V} & & \uparrow \boldsymbol{V} & & \uparrow \boldsymbol{V} \\
\begin{bmatrix}0\\0\end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix}h_1^{(1)}\\h_2^{(1)}\end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix}h_1^{(2)}\\h_2^{(2)}\end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix}h_1^{(3)}\\h_2^{(3)}\end{bmatrix} \\
 & & \uparrow \boldsymbol{U} & & \uparrow \boldsymbol{U} & & \uparrow \boldsymbol{U} \\
input & & \begin{bmatrix}1\\0\end{bmatrix} & & \begin{bmatrix}0.5\\-0.2\end{bmatrix} & & \begin{bmatrix}0\\1\end{bmatrix}
\end{array}
$$

$$
\begin{array}{rccl}
time & & t=1 & \\
target & & 1 & \\
 & & \downarrow & \\
error & & e^{(1)} & = (1 - o^{(1)})^2 \\
 & & \uparrow & \\
output & & o^{(1)} & = tanh\left(\begin{bmatrix}v_1 & v_2\end{bmatrix}\begin{bmatrix}h_1^{(1)}\\h_2^{(1)}\end{bmatrix} + c\right) \\
 & & \uparrow \boldsymbol{V} & \\
 & \begin{bmatrix}0\\0\end{bmatrix} \xrightarrow{\boldsymbol{W}} & \begin{bmatrix}h_1^{(1)}\\h_2^{(1)}\end{bmatrix} & = tanh\left(\begin{bmatrix}u_{11} & u_{12}\\u_{21} & u_{22}\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}b_1\\b_2\end{bmatrix}\right) \\
 & & \uparrow \boldsymbol{U} & \\
input & & \begin{bmatrix}1\\0\end{bmatrix} &
\end{array}
$$

Example of Back-Propagation Through Time

$$
\begin{array}{ccccccc}
\color{red}{time} & & \color{red}{t=1} & & \color{red}{t=2} & & \color{red}{t=3} \\
\color{blue}{target} & & 1 & & -1 & & 1 \\
& & \downarrow & & \downarrow & & \downarrow \\
\color{blue}{error} & & e^{(1)} & & e^{(2)} & & e^{(3)} \\
& & \uparrow & & \uparrow & & \uparrow \\
\color{blue}{output} & & o^{(1)} & & o^{(2)} & & o^{(3)} \\
& & \uparrow \boldsymbol{V} & & \uparrow \boldsymbol{V} & & \uparrow \boldsymbol{V} \\
\begin{bmatrix} 0 \\ 0 \end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix} h_1^{(3)} \\ h_2^{(3)} \end{bmatrix} \\
& & \uparrow \boldsymbol{U} & & \uparrow \boldsymbol{U} & & \uparrow \boldsymbol{U} \\
\color{blue}{input} & & \begin{bmatrix} 1 \\ 0 \end{bmatrix} & & \begin{bmatrix} 0.5 \\ -0.2 \end{bmatrix} & & \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\end{array}
$$

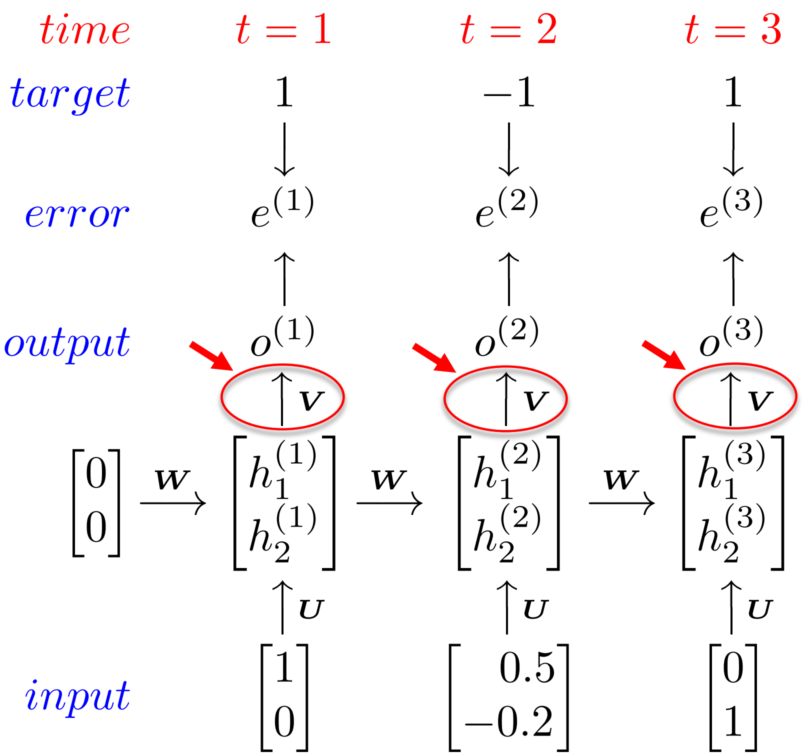
$$
\begin{array}{cccl}
\color{red}{time} & & \color{red}{t=2} & \\
\color{blue}{target} & & -1 & \\
& & \downarrow & \\
\color{blue}{error} & & e^{(2)} & = (-1 - o^{(2)})^2 \\
& & \uparrow & \\
\color{blue}{output} & & o^{(2)} & = tanh\left(\begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \end{bmatrix} + c \right) \\
& & \uparrow \boldsymbol{V} & \\
\begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \end{bmatrix} & = tanh\left(\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} 0.5 \\ -0.2 \end{bmatrix} + \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right) \\
& & \uparrow \boldsymbol{U} & \\
\color{blue}{input} & & \begin{bmatrix} 0.5 \\ -0.2 \end{bmatrix} &
\end{array}
$$

Example of Back-Propagation Through Time

$$\begin{array}{ccccccc}
\text{time} & & t=1 & & t=2 & & t=3 \\
\text{target} & & 1 & & -1 & & 1 \\
 & & \downarrow & & \downarrow & & \downarrow \\
\text{error} & & e^{(1)} & & e^{(2)} & & e^{(3)} \\
 & & \uparrow & & \uparrow & & \uparrow \\
\text{output} & & o^{(1)} & & o^{(2)} & & o^{(3)} \\
 & & \uparrow \boldsymbol{V} & & \uparrow \boldsymbol{V} & & \uparrow \boldsymbol{V} \\
\begin{bmatrix} 0 \\ 0 \end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix} h_1^{(1)} \\ h_2^{(1)} \end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix} h_1^{(3)} \\ h_2^{(3)} \end{bmatrix} \\
 & & \uparrow \boldsymbol{U} & & \uparrow \boldsymbol{U} & & \uparrow \boldsymbol{U} \\
\text{input} & & \begin{bmatrix} 1 \\ 0 \end{bmatrix} & & \begin{bmatrix} 0.5 \\ -0.2 \end{bmatrix} & & \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\end{array}$$

$$\begin{array}{llll}
\text{time} & & t=3 & \\
\text{target} & & 1 & \\
 & & \downarrow & \\
\text{error} & & e^{(3)} & = (1 - o^{(3)})^2 \\
 & & \uparrow & \\
\text{output} & & o^{(3)} & = tanh\left(\begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} h_1^{(3)} \\ h_2^{(3)} \end{bmatrix} + c \right) \\
 & & \uparrow \boldsymbol{V} & \\
\begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \end{bmatrix} & \xrightarrow{\boldsymbol{W}} & \begin{bmatrix} h_1^{(3)} \\ h_2^{(3)} \end{bmatrix} & = tanh\left(\begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} \begin{bmatrix} h_1^{(2)} \\ h_2^{(2)} \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right) \\
 & & \uparrow \boldsymbol{U} & \\
\text{input} & & \begin{bmatrix} 0 \\ 1 \end{bmatrix} &
\end{array}$$

Example of Back-Propagation Through Time



$$L = \sum_t e^{(t)} = \sum_t (y^{(t)} - o^{(t)})^2$$

$$\frac{\partial L}{\partial \boldsymbol{V}} = \frac{\partial \sum_t e^{(t)}}{\partial \boldsymbol{V}} = \sum_t \frac{\partial (y^{(t)} - o^{(t)})^2}{\partial \boldsymbol{V}} = \sum_t -2(y^{(t)} - o^{(t)}) \frac{\partial o^{(t)}}{\partial \boldsymbol{V}}$$

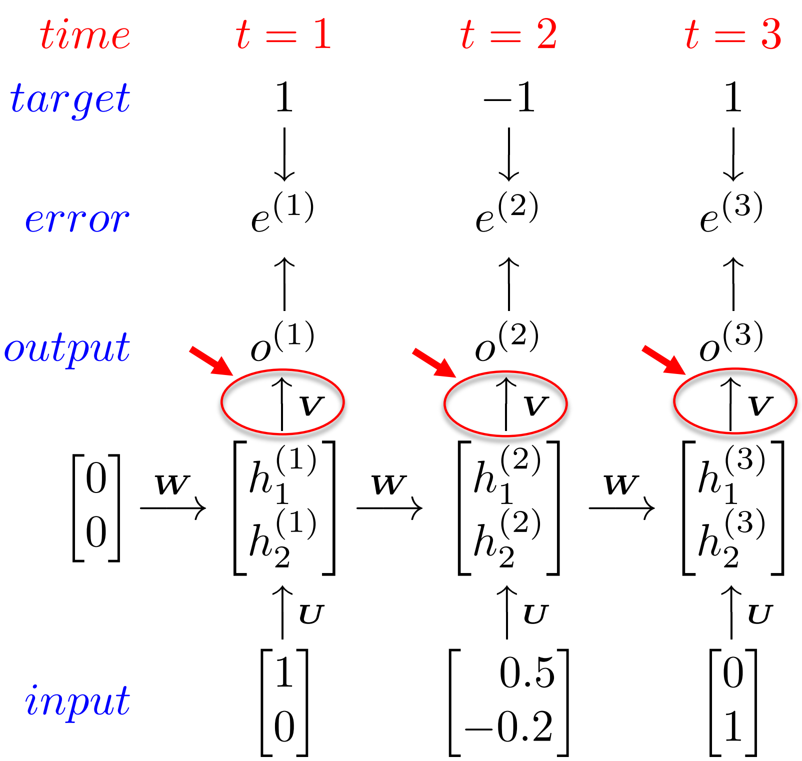
$$\frac{\partial L}{\partial \boldsymbol{V}} = -2(1 - o^{(1)}) \frac{\partial o^{(1)}}{\partial \boldsymbol{V}} - 2(-1 - o^{(2)}) \frac{\partial o^{(2)}}{\partial \boldsymbol{V}} - 2(1 - o^{(3)}) \frac{\partial o^{(3)}}{\partial \boldsymbol{V}}$$

$$\frac{\partial L}{\partial \boldsymbol{V}} = -2[(1 - o^{(1)}) \frac{\partial o^{(1)}}{\partial a_o^{(1)}} \frac{\partial a_o^{(1)}}{\partial \boldsymbol{V}} + (-1 - o^{(2)}) \frac{\partial o^{(2)}}{\partial a_o^{(2)}} \frac{\partial a_o^{(2)}}{\partial \boldsymbol{V}} + (1 - o^{(3)}) \frac{\partial o^{(3)}}{\partial a_o^{(3)}} \frac{\partial a_o^{(3)}}{\partial \boldsymbol{V}}]$$

recalling that $o^{(t)} = tanh(a_o^{(t)}) \Rightarrow \frac{\partial o^{(t)}}{\partial a_o{}^{(t)}} = 1 - (o^{(t)})^2$

moreover, since $a_o^{(t)} = [v_1 \; v_2] \begin{bmatrix} h_1^{(t)} \\ h_2^{(t)} \end{bmatrix} + c \Rightarrow \frac{\partial a_o^{(t)}}{\partial \boldsymbol{V}} = [h_1^{(t)} \; h_2^{(t)}]$

Example of Back-Propagation Through Time



putting all together

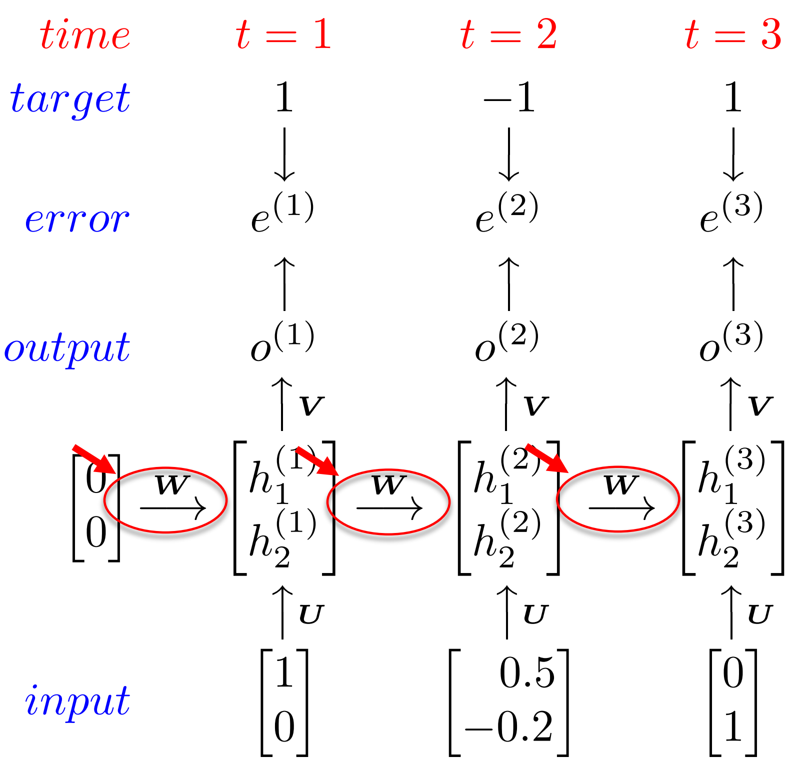
$$\frac{\partial o^{(1)}}{\partial \boldsymbol{V}} = \frac{\partial o^{(1)}}{\partial a_o{}^{(1)}} \left[\frac{\partial a_o{}^{(1)}}{\partial v_1} \;\; \frac{\partial a_o{}^{(1)}}{\partial v_2} \right] = (1 - (o^{(1)})^2) \left[h_1^{(1)} \;\; h_2^{(1)} \right]$$

$$\frac{\partial o^{(2)}}{\partial \boldsymbol{V}} = \frac{\partial o^{(2)}}{\partial a_o{}^{(2)}} \left[\frac{\partial a_o{}^{(2)}}{\partial v_1} \;\; \frac{\partial a_o{}^{(2)}}{\partial v_2} \right] = (1 - (o^{(2)})^2) \left[h_1^{(2)} \;\; h_2^{(2)} \right]$$

$$\frac{\partial o^{(3)}}{\partial \boldsymbol{V}} = \frac{\partial o^{(3)}}{\partial a_o{}^{(3)}} \left[\frac{\partial a_o{}^{(3)}}{\partial v_1} \;\; \frac{\partial a_o{}^{(3)}}{\partial v_2} \right] = (1 - (o^{(3)})^2) \left[h_1^{(3)} \;\; h_2^{(3)} \right]$$

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{V}} \;\; = \;\; -2\Big[& (1 - o^{(1)})(1 - (o^{(1)})^2) \left[h_1^{(1)} \;\; h_2^{(1)} \right] + \\ & (-1 - o^{(2)})(1 - (o^{(2)})^2) \left[h_1^{(2)} \;\; h_2^{(2)} \right] + \\ & (1 - o^{(3)})(1 - (o^{(3)})^2) \left[h_1^{(3)} \;\; h_2^{(3)} \right] \Big] \end{aligned}$$

Example of Back-Propagation Through Time



$$\boxed{\frac{\partial L}{\partial \boldsymbol{V}} = -2[(1-o^{(1)})\frac{\partial o^{(1)}}{\partial a_o^{(1)}}\frac{\partial a_o^{(1)}}{\partial \boldsymbol{V}} + (-1-o^{(2)})\frac{\partial o^{(2)}}{\partial a_o^{(2)}}\frac{\partial a_o^{(2)}}{\partial \boldsymbol{V}} + (1-o^{(3)})\frac{\partial o^{(3)}}{\partial a_o^{(3)}}\frac{\partial a_o^{(3)}}{\partial \boldsymbol{V}}]}$$

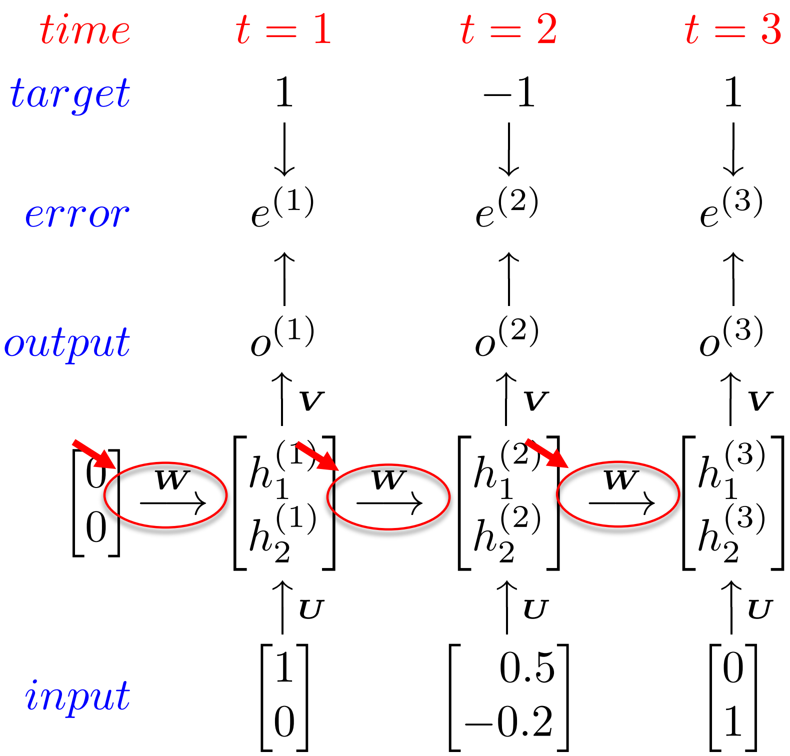
$$\frac{\partial L}{\partial \boldsymbol{W}} = -2[(1-o^{(1)})\frac{\partial o^{(1)}}{\partial a_o^{(1)}}\frac{\partial a_o^{(1)}}{\partial \boldsymbol{W}} + (-1-o^{(2)})\frac{\partial o^{(2)}}{\partial a_o^{(2)}}\frac{\partial a_o^{(2)}}{\partial \boldsymbol{W}} + (1-o^{(3)})\frac{\partial o^{(3)}}{\partial a_o^{(3)}}\frac{\partial a_o^{(3)}}{\partial \boldsymbol{W}}]$$

recalling that

$$a_o^{(t)} = [v_1\ v_2]\begin{bmatrix} h_1^{(t)} \\ h_2^{(t)} \end{bmatrix} + c \Rightarrow \frac{\partial a_o^{(t)}}{\partial \boldsymbol{W}} = v_1\frac{\partial h_1^{(t)}}{\partial \boldsymbol{W}} + v_2\frac{\partial h_2^{(t)}}{\partial \boldsymbol{W}}$$

$$\begin{bmatrix} h_1^{(t)} \\ h_2^{(t)} \end{bmatrix} = \begin{bmatrix} tanh(a_{h_1}^{(t)}) \\ tanh(a_{h_2}^{(t)}) \end{bmatrix} \Rightarrow \frac{\partial}{\partial \boldsymbol{W}}\begin{bmatrix} h_1^{(t)} \\ h_2^{(t)} \end{bmatrix} = \begin{bmatrix} (1-(h_1^{(t)})^2)\frac{\partial a_{h_1}^{(t)}}{\partial \boldsymbol{W}} \\ (1-(h_2^{(t)})^2)\frac{\partial a_{h_2}^{(t)}}{\partial \boldsymbol{W}} \end{bmatrix}$$

Example of Back-Propagation Through Time



recalling that $\boldsymbol{a}_h{}^{(t)} = \boldsymbol{U}\boldsymbol{x}^{(t)} + \boldsymbol{W}\boldsymbol{h}^{(t-1)} + \boldsymbol{b}$

$$\begin{bmatrix}\frac{\partial a_{h_1}^{(t)}}{\partial \boldsymbol{W}}\\ \frac{\partial a_{h_2}^{(t)}}{\partial \boldsymbol{W}}\end{bmatrix} = \begin{bmatrix}\frac{\partial(w_{11}h_1^{(t-1)}+w_{12}h_2^{(t-1)})}{\partial \boldsymbol{W}}\\ \frac{\partial(w_{21}h_1^{(t-1)}+w_{22}h_2^{(t-1)})}{\partial \boldsymbol{W}}\end{bmatrix} = \begin{matrix}\begin{bmatrix}h_1^{(t-1)} & h_2^{(t-1)}\\ 0 & 0\end{bmatrix} + w_{11}\frac{\partial h_1^{(t-1)}}{\partial \boldsymbol{W}} + w_{12}\frac{\partial h_2^{(t-1)}}{\partial \boldsymbol{W}}\\ \begin{bmatrix}0 & 0\\ h_1^{(t-1)} & h_2^{(t-1)}\end{bmatrix} + w_{21}\frac{\partial h_1^{(t-1)}}{\partial \boldsymbol{W}} + w_{22}\frac{\partial h_2^{(t-1)}}{\partial \boldsymbol{W}}\end{matrix}$$

t=1 $\Rightarrow \begin{bmatrix}\frac{\partial h_1^{(0)}}{\partial \boldsymbol{W}}\\ \frac{\partial h_2^{(0)}}{\partial \boldsymbol{W}}\end{bmatrix} = \begin{bmatrix}\frac{\partial h_1^{(1)}}{\partial \boldsymbol{W}}\\ \frac{\partial h_2^{(1)}}{\partial \boldsymbol{W}}\end{bmatrix} =$ zero $2 \times (2 \times 2)$ tensor $\Rightarrow \frac{\partial a_o^{(1)}}{\partial \boldsymbol{W}} = \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$

t=2 $\Rightarrow \begin{bmatrix}\frac{\partial h_1^{(2)}}{\partial \boldsymbol{W}}\\ \frac{\partial h_2^{(2)}}{\partial \boldsymbol{W}}\end{bmatrix} = \begin{bmatrix}(1-(h_1^{(2)})^2)\frac{\partial a_{h_1}^{(2)}}{\partial \boldsymbol{W}}\\ (1-(h_2^{(2)})^2)\frac{\partial a_{h_2}^{(2)}}{\partial \boldsymbol{W}}\end{bmatrix} = \begin{matrix}(1-(h_1^{(2)})^2)\begin{bmatrix}h_1^{(1)} & h_2^{(1)}\\ 0 & 0\end{bmatrix}\\ (1-(h_2^{(2)})^2)\begin{bmatrix}0 & 0\\ h_1^{(1)} & h_2^{(1)}\end{bmatrix}\end{matrix}$

$$\frac{\partial a_o^{(2)}}{\partial \boldsymbol{W}} = v_1(1-(h_1^{(2)})^2)\begin{bmatrix}h_1^{(1)} & h_2^{(1)}\\ 0 & 0\end{bmatrix} + v_2(1-(h_2^{(2)})^2)\begin{bmatrix}0 & 0\\ h_1^{(1)} & h_2^{(1)}\end{bmatrix}$$

Example of Back-Propagation Through Time

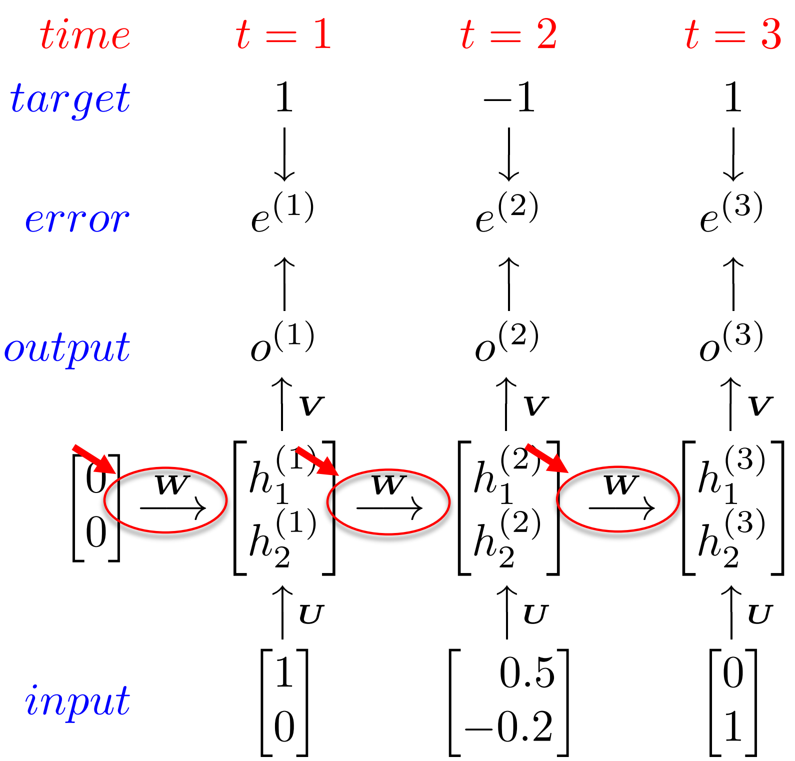
time	$t=1$	$t=2$	$t=3$
target	1	-1	1
	$\downarrow$	$\downarrow$	$\downarrow$
error	$e^{(1)}$	$e^{(2)}$	$e^{(3)}$
	$\uparrow$	$\uparrow$	$\uparrow$
output	$o^{(1)}$	$o^{(2)}$	$o^{(3)}$
	$\uparrow \boldsymbol{V}$	$\uparrow \boldsymbol{V}$	$\uparrow \boldsymbol{V}$
$\begin{bmatrix}0\\0\end{bmatrix} \xrightarrow{\boldsymbol{W}}$	$\begin{bmatrix}h_1^{(1)}\\h_2^{(1)}\end{bmatrix} \xrightarrow{\boldsymbol{W}}$	$\begin{bmatrix}h_1^{(2)}\\h_2^{(2)}\end{bmatrix} \xrightarrow{\boldsymbol{W}}$	$\begin{bmatrix}h_1^{(3)}\\h_2^{(3)}\end{bmatrix}$
	$\uparrow \boldsymbol{U}$	$\uparrow \boldsymbol{U}$	$\uparrow \boldsymbol{U}$
input	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0.5\\-0.2\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$

t=3 $\Rightarrow$

$$\frac{\partial h_1^{(3)}}{\partial \boldsymbol{W}} = (1-(h_1^{(3)})^2)\Bigg(\begin{bmatrix}h_1^{(2)} & h_2^{(2)}\\ 0 & 0\end{bmatrix} + w_{11}(1-(h_1^{(2)})^2)\begin{bmatrix}h_1^{(1)} & h_2^{(1)}\\ 0 & 0\end{bmatrix} + w_{12}(1-(h_2^{(2)})^2)\begin{bmatrix}0 & 0\\ h_1^{(1)} & h_2^{(1)}\end{bmatrix}\Bigg)$$

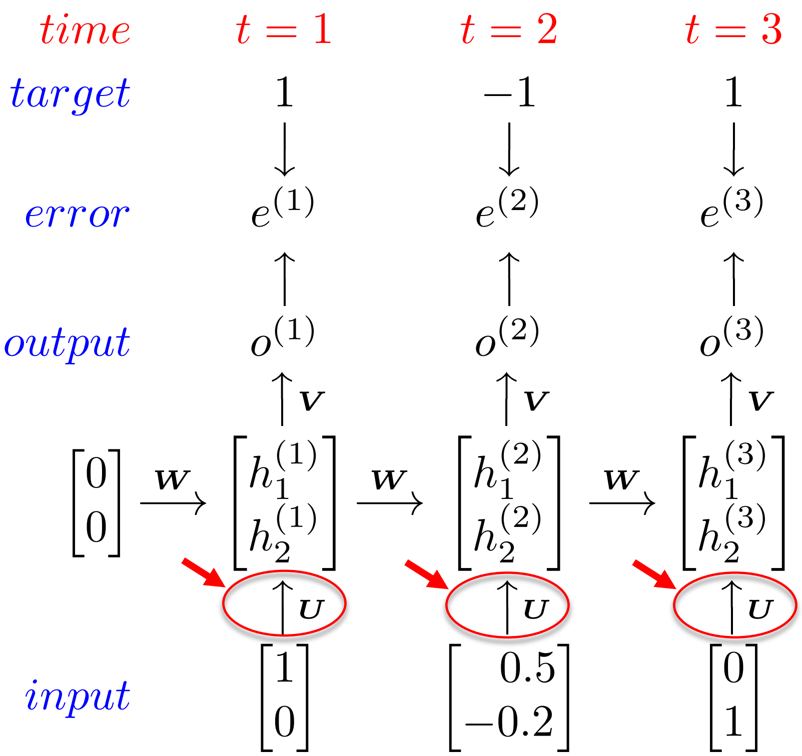
$$\frac{\partial h_2^{(3)}}{\partial \boldsymbol{W}} = (1-(h_2^{(3)})^2)\Bigg(\begin{bmatrix}0 & 0\\ h_1^{(2)} & h_2^{(2)}\end{bmatrix} + w_{21}(1-(h_1^{(2)})^2)\begin{bmatrix}h_1^{(1)} & h_2^{(1)}\\ 0 & 0\end{bmatrix} + w_{22}(1-(h_2^{(2)})^2)\begin{bmatrix}0 & 0\\ h_1^{(1)} & h_2^{(1)}\end{bmatrix}\Bigg)$$

Example of Back-Propagation Through Time



$$\begin{bmatrix}\frac{\partial h_1^{(t)}}{\partial \boldsymbol{W}}\\[4pt] \frac{\partial h_2^{(t)}}{\partial \boldsymbol{W}}\end{bmatrix} = \begin{matrix} (1-(h_1^{(t)})^2)\left(\begin{bmatrix} h_1^{(t-1)} & h_2^{(t-1)}\\ 0 & 0\end{bmatrix} + w_{11}\frac{\partial h_1^{(t-1)}}{\partial \boldsymbol{W}} + w_{12}\frac{\partial h_2^{(t-1)}}{\partial \boldsymbol{W}}\right) \\ (1-(h_2^{(t)})^2)\left(\begin{bmatrix} 0 & 0\\ h_1^{(t-1)} & h_2^{(t-1)}\end{bmatrix} + w_{21}\frac{\partial h_1^{(t-1)}}{\partial \boldsymbol{W}} + w_{22}\frac{\partial h_2^{(t-1)}}{\partial \boldsymbol{W}}\right)\end{matrix}$$

Example of Back-Propagation Through Time



$$\frac{\partial L}{\partial \boldsymbol{W}} = -2[(1-o^{(1)})\frac{\partial o^{(1)}}{\partial a_o^{(1)}}\frac{\partial a_o^{(1)}}{\partial \boldsymbol{U}} + (-1-o^{(2)})\frac{\partial o^{(2)}}{\partial a_o^{(2)}}\frac{\partial a_o^{(2)}}{\partial \boldsymbol{U}} + (1-o^{(3)})\frac{\partial o^{(3)}}{\partial a_o^{(3)}}\frac{\partial a_o^{(3)}}{\partial \boldsymbol{U}}]$$

?