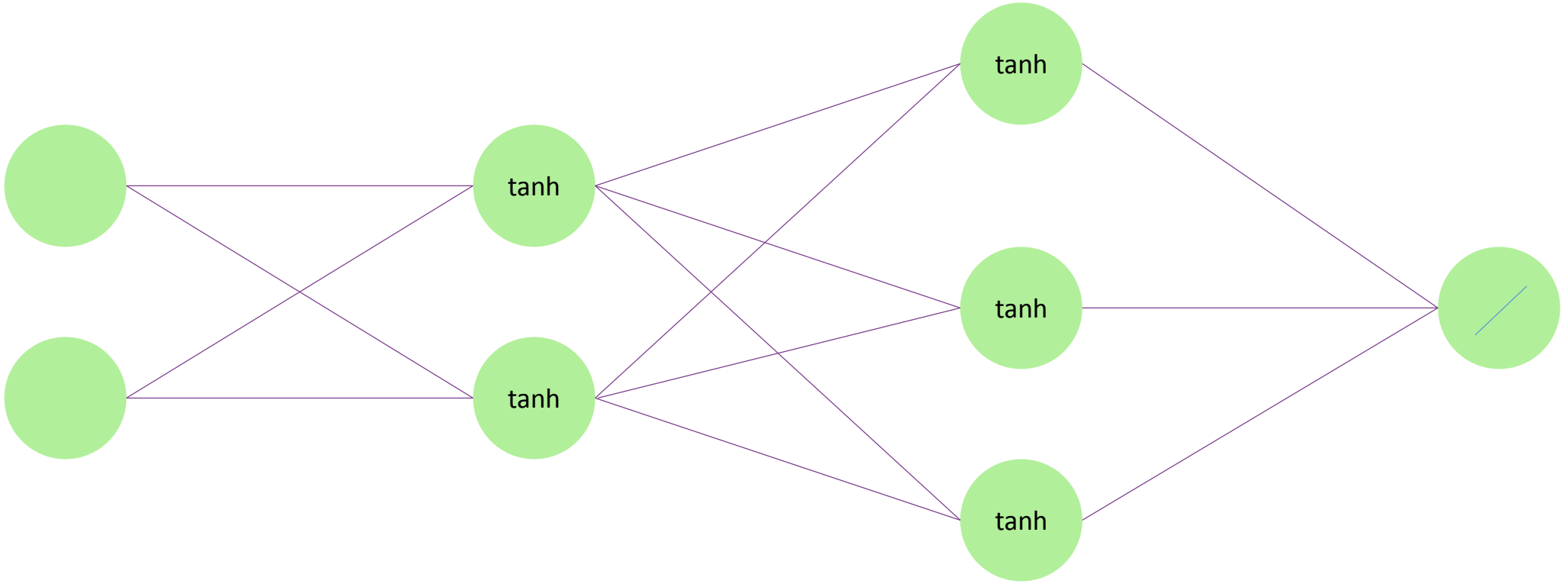
Back Propagation Algorithm

SunnerLi

Outline

1. Component
2. Derivate
   - ① error to last score
   - ② error to other score
   - ③ next input to this output
   - ④ output to input
   - ⑤ summary
3. Formula summery
4. Implementation
5. Error comparison

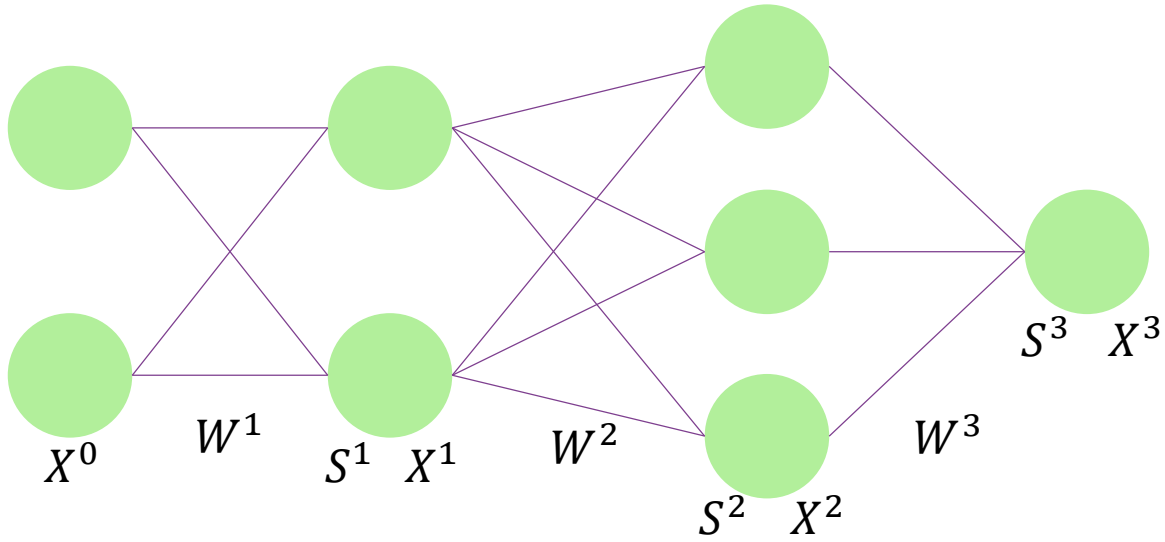
Component



$$X^0 \xrightarrow{W^1} S^1 \xrightarrow{\tanh} X^1 \xrightarrow{W^2} S^2 \xrightarrow{\tanh} X^2 \xrightarrow{W^3} S^3 \xrightarrow{\text{linear}} X^3$$

Derivate

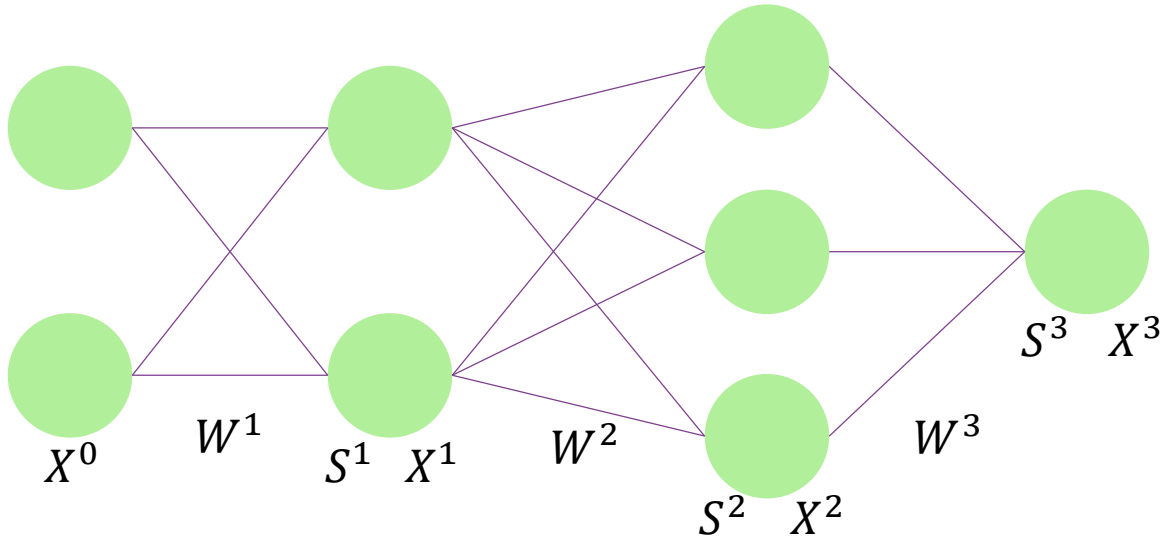
L is the index of last layer



$$s_1^{(L)} = \sum_{i=0}^{d^{L-1}} w_{i1}^{(L)} x_i^{(L-1)} \qquad \frac{\partial s_1^{(L)}}{\partial w_{i1}^{(L)}} = \frac{\partial\{\sum_{i=0}^{d^{L-1}} w_{i1}^{(L)} x_i^{(L-1)}\}}{\partial w_{i1}^{(L)}} = x_i^{(L-1)}$$

$$s_j^{(l)} = \sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)} \qquad \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \frac{\partial\{\sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)}\}}{\partial w_{ij}^{(l)}} = x_i^{(l-1)}$$

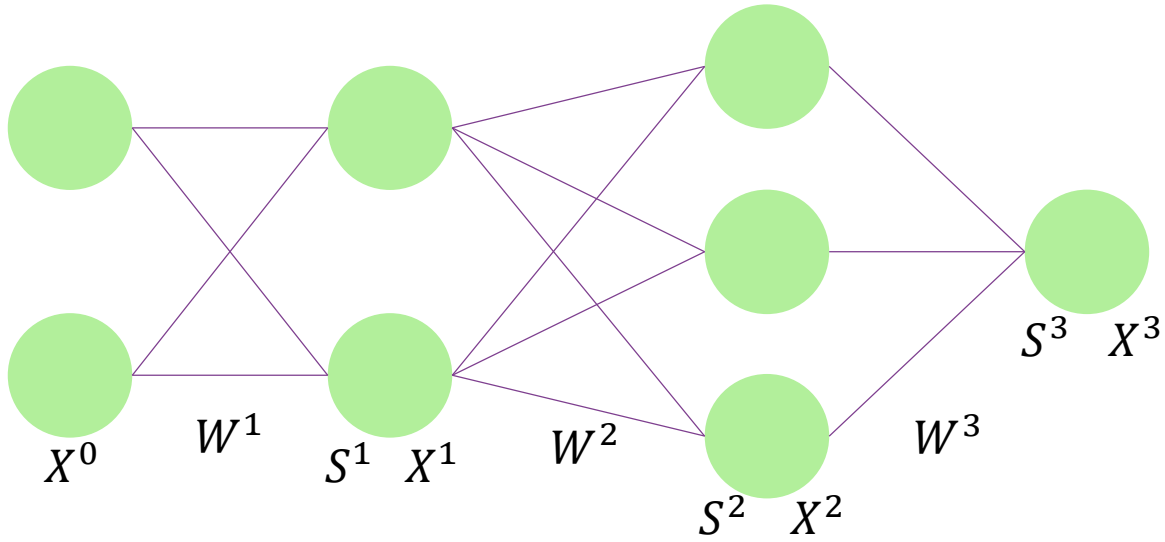
Derivate



$$s_1^{(L)} = \sum_{i=0}^{d^{L-1}} w_{i1}^{(L)} x_i^{(L-1)} \qquad \frac{\partial s_1^{(L)}}{\partial w_{i1}^{(L)}} = \frac{\partial\{\sum_{i=0}^{d^{L-1}} w_{i1}^{(L)} x_i^{(L-1)}\}}{\partial w_{i1}^{(L)}} = x_i^{(L-1)} \qquad \frac{\partial e_n}{\partial w_{i1}^{(L)}} = \frac{\partial e_n}{\partial s_1^{(L)}} * \frac{\partial s_1^{(L)}}{\partial w_{i1}^{(L)}} = \delta_1^{(L)} * x_i^{(L-1)}$$

$$s_j^{(l)} = \sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)} \qquad \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \frac{\partial\{\sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)}\}}{\partial w_{ij}^{(l)}} = x_i^{(l-1)} \qquad \frac{\partial e_n}{\partial w_{ij}^{(l)}} = \frac{\partial e_n}{\partial s_j^{(l)}} * \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} * x_i^{(l-1)}$$

Derivate – error to last score



$$e_n = \left(y_n - s_1^{(L)}\right)^2 \qquad \frac{\partial e_n}{\partial s_1^{(L)}} = -2 * (y_n - s_1^{(L)})$$

最後一層的誤差求導，直接將tag和結果相減以後，乘以-2即可。
如此簡單是因為，我們使用平方誤差法來評估誤差，他的微分即乘以-2

Derivate – error to other score

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}}$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

Derivate – error to other score

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}}$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

$$\delta_j^{(l)} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)}$$

Derivate – error to other score

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} *$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

Derivate – next input to this output

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} *$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

$$s_j^{(l)} = \sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)} \qquad \frac{\partial s_j^{(l)}}{\partial x_i^{(l-1)}} = \frac{\partial \{\sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)}\}}{\partial x_i^{(l-1)}} = w_{ij}^{(l-1)}$$

Derivate – next input to this output

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} *$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

$$s_j^{(l)} = \sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)}$$

$$s_k^{(l+1)} = \sum_{j=0}^{d^{l}} w_{jk}^{(l+1)} x_j^{(l)}$$

$$\frac{\partial s_j^{(l)}}{\partial x_i^{(l-1)}} = \frac{\partial \{\sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)}\}}{\partial x_i^{(l-1)}} = w_{ij}^{(l-1)}$$

Derivate – next input to this output

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} *$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

$$s_j^{(l)} = \sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)} \qquad \frac{\partial s_j^{(l)}}{\partial x_i^{(l-1)}} = \frac{\partial\{\sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)}\}}{\partial x_i^{(l-1)}} = w_{ij}^{(l-1)}$$

$$s_k^{(l+1)} = \sum_{j=0}^{d^{l}} w_{jk}^{(l+1)} x_j^{(l)} \qquad \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} = \frac{\partial\{\sum_{j=0}^{d^{l}} w_{jk}^{(l+1)} x_j^{(l)}\}}{\partial x_j^{(l)}} = w_{jk}^{(l)}$$

Derivate – next input to this output

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} *$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

$$s_k^{(l+1)} = \sum_{j=0}^{d^l} w_{jk}^{(l+1)} x_j^{(l)} \qquad \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} = \frac{\partial \{\sum_{j=0}^{d^l} w_{jk}^{(l+1)} x_j^{(l)}\}}{\partial x_j^{(l)}} = w_{jk}^{(l)}$$

Derivate – next input to this output

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} *$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

$$s_k^{(l+1)} = \sum_{j=0}^{d^l} w_{jk}^{(l+1)} x_j^{(l)} \qquad \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} = \frac{\partial \{\sum_{j=0}^{d^l} w_{jk}^{(l+1)} x_j^{(l)}\}}{\partial x_j^{(l)}} = w_{jk}^{(l)}$$

Derivate – next input to this output

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} *$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

Derivate – output to input

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} *$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

$$x_j^{(l)} = \tanh(s_j^{(l)})$$

$$\frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \frac{\partial \tanh(s_j^{(l)})}{\partial s_j^{(l)}} = \tanh'\left(s_j^{(l)}\right) = sech^2(s_j^{(l)})$$

Derivate – output to input

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$

誤差對於
這一層輸入
的導數

誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

$$x_j^{(l)} = \tanh(s_j^{(l)})$$

$$\frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \frac{\partial \tanh(s_j^{(l)})}{\partial s_j^{(l)}} = \tanh'\left(s_j^{(l)}\right) = sech^2(s_j^{(l)})$$

Derivate – output to input

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \frac{\partial e_n}{\partial s_k^{(l+1)}} * \frac{\partial s_k^{(l+1)}}{\partial x_j^{(l)}} * \frac{\partial x_j^{(l)}}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$

誤差對於
這一層輸入
的導數

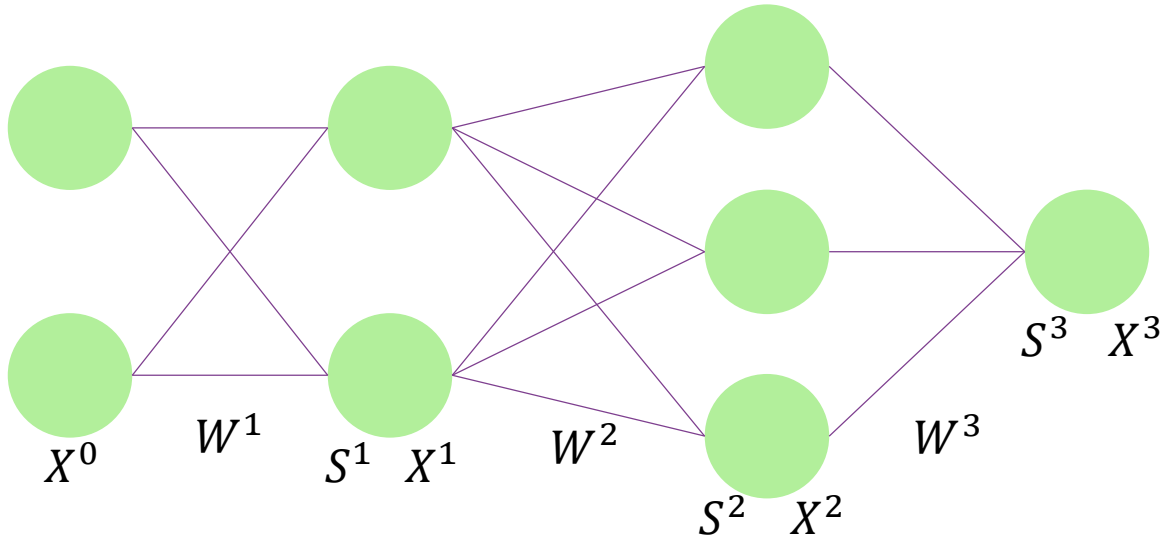
誤差對於
下一層輸入
的導數

誤差對於
這一層
輸出的導數

誤差對於
這一層
輸入的導數

Derivate – summary

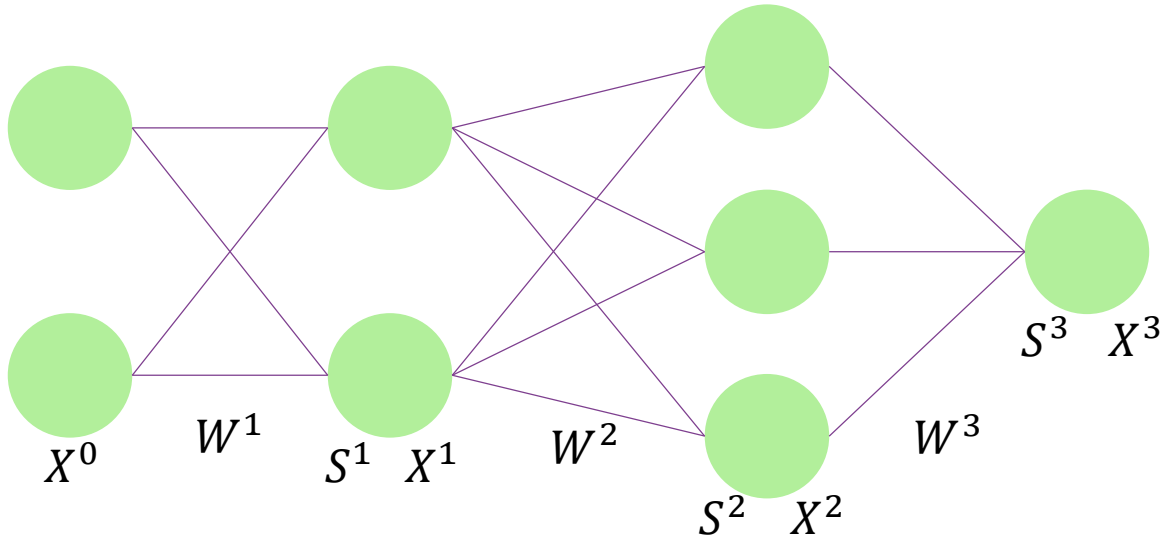
$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$



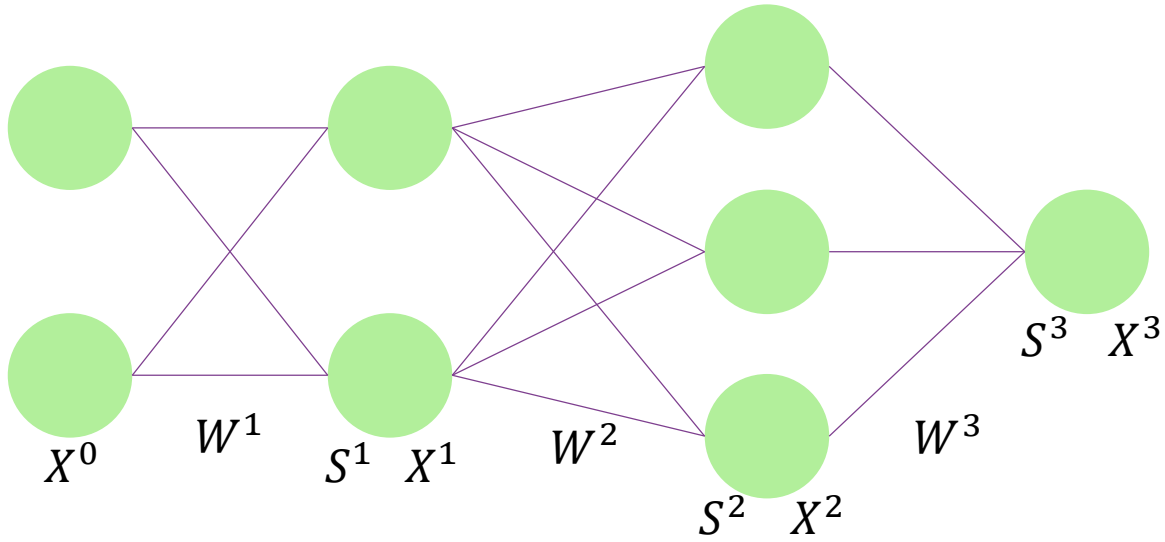
Derivate – summary

$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$

$$\frac{\partial e_n}{\partial s_1^{(L)}} = -2 * (y_n - s_1^{(L)})$$



Derivate – summary



$$\frac{\partial e_n}{\partial s_j^{(l)}} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$

$$\frac{\partial e_n}{\partial s_1^{(L)}} = -2 * (y_n - s_1^{(L)})$$

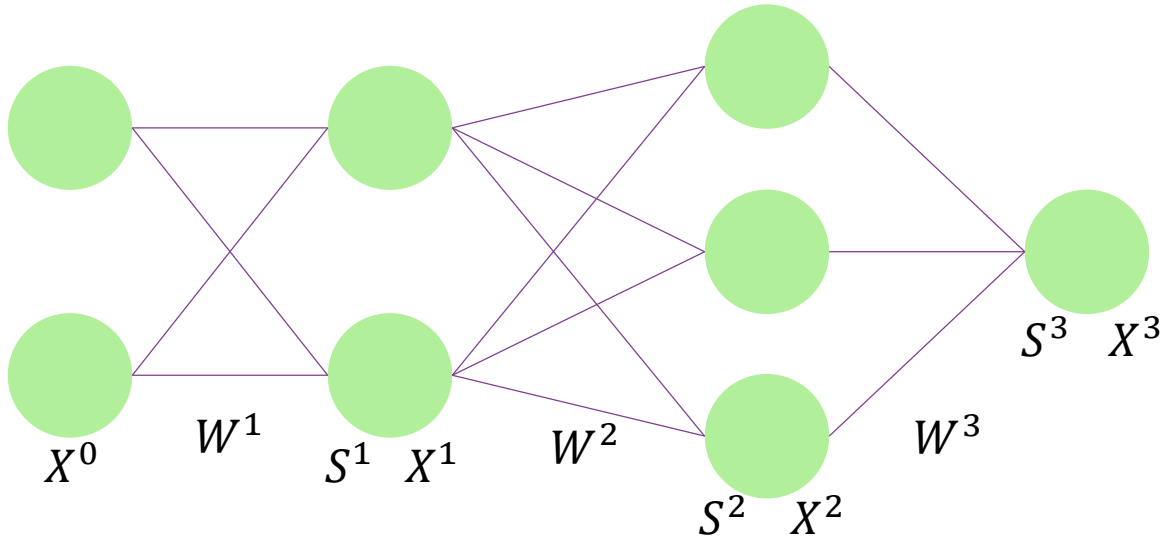
$$\delta_j^{(L)} = -2 * (y_n - s_1^{(L)})$$

$$\delta_j^{(l)} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$

Derivate – summary

$$\delta_j^{(L)} = -2 * (y_n - s_1^{(L)})$$

$$\delta_j^{(l)} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$



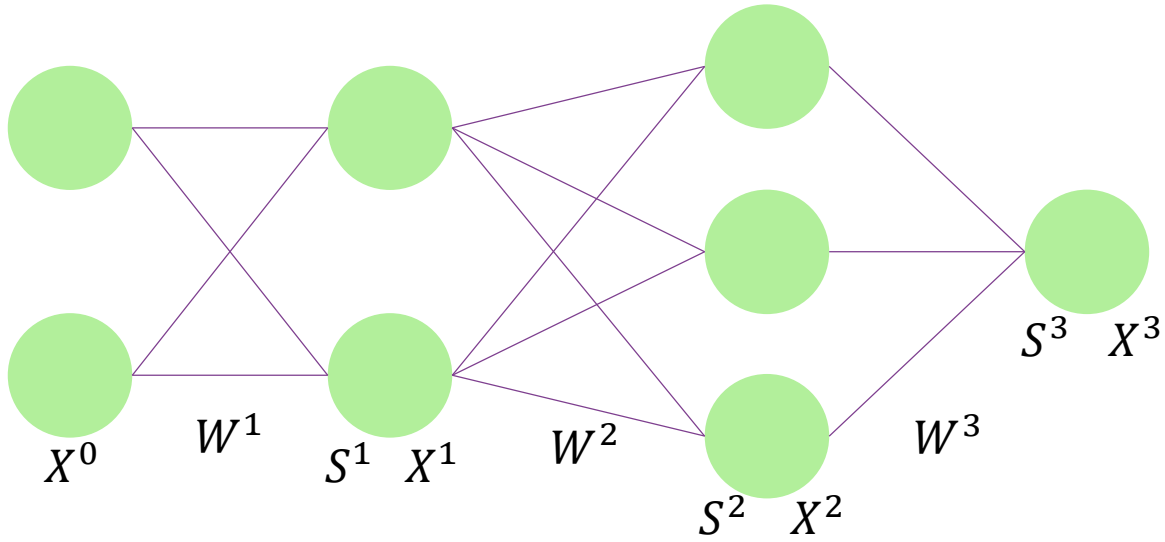
Derivate – summary

$$\delta_j^{(L)} = -2 * (y_n - s_1^{(L)})$$

$$\delta_j^{(l)} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$

$$\frac{\partial e_n}{\partial w_{i1}^{(L)}} = \frac{\partial e_n}{\partial s_1^{(L)}} * \frac{\partial s_1^{(L)}}{\partial w_{i1}^{(L)}} = \delta_1^{(L)} * x_i^{(L-1)}$$

$$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \frac{\partial e_n}{\partial s_j^{(l)}} * \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} * x_i^{(l-1)}$$



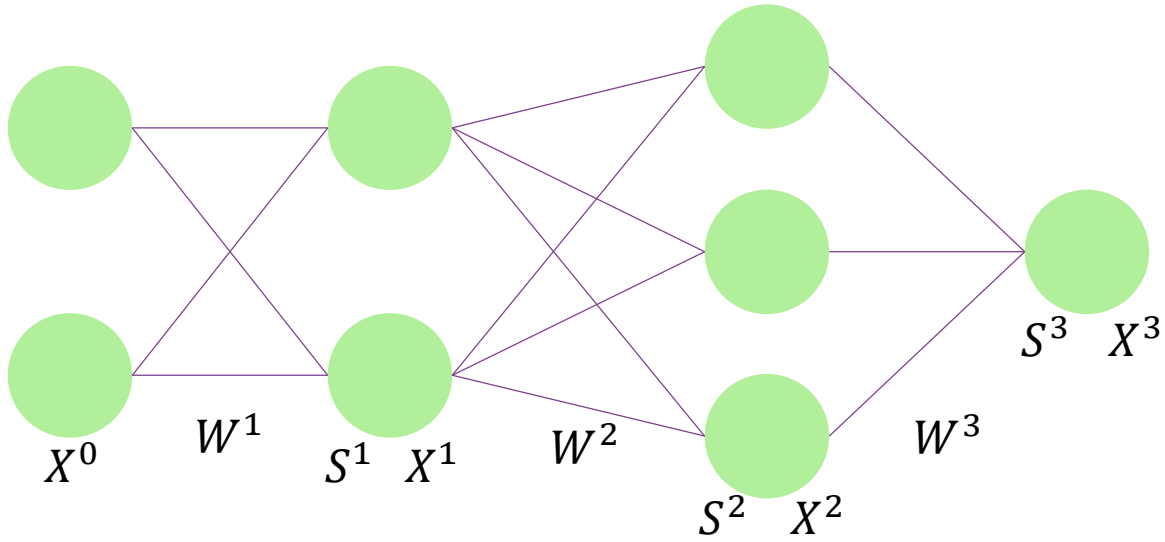
Derivate – summary

$$\delta_j^{(L)} = -2 * (y_n - s_1^{(L)})$$

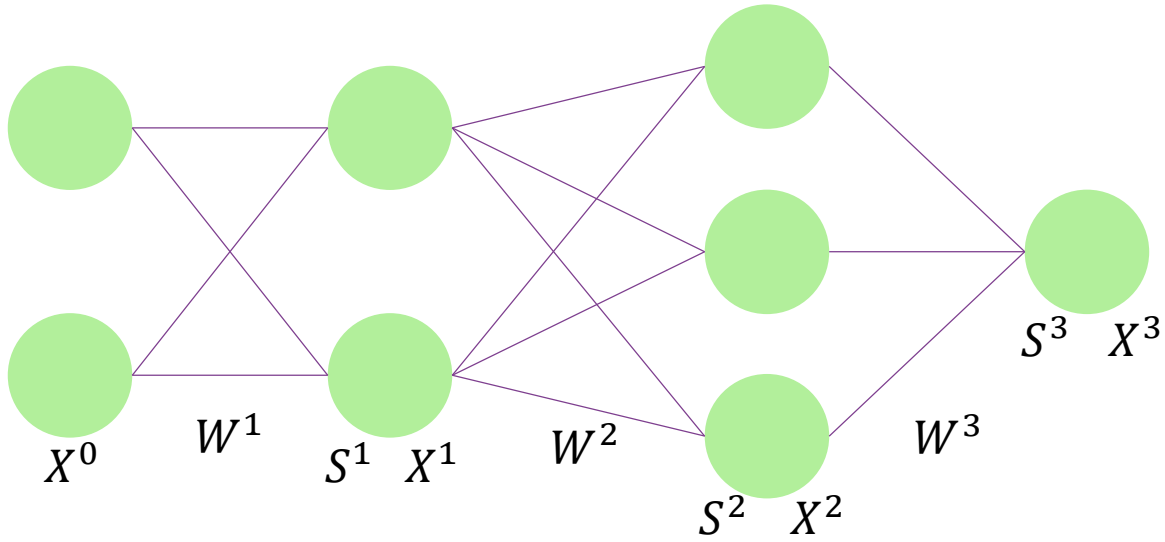
$$\delta_j^{(l)} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$

$$\frac{\partial e_n}{\partial w_{i1}^{(L)}} = \frac{\partial e_n}{\partial s_1^{(L)}} * \frac{\partial s_1^{(L)}}{\partial w_{i1}^{(L)}} = \delta_1^{(L)} * x_i^{(L-1)}$$

$$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \frac{\partial e_n}{\partial s_j^{(l)}} * \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} * x_i^{(l-1)}$$



Derivate – summary



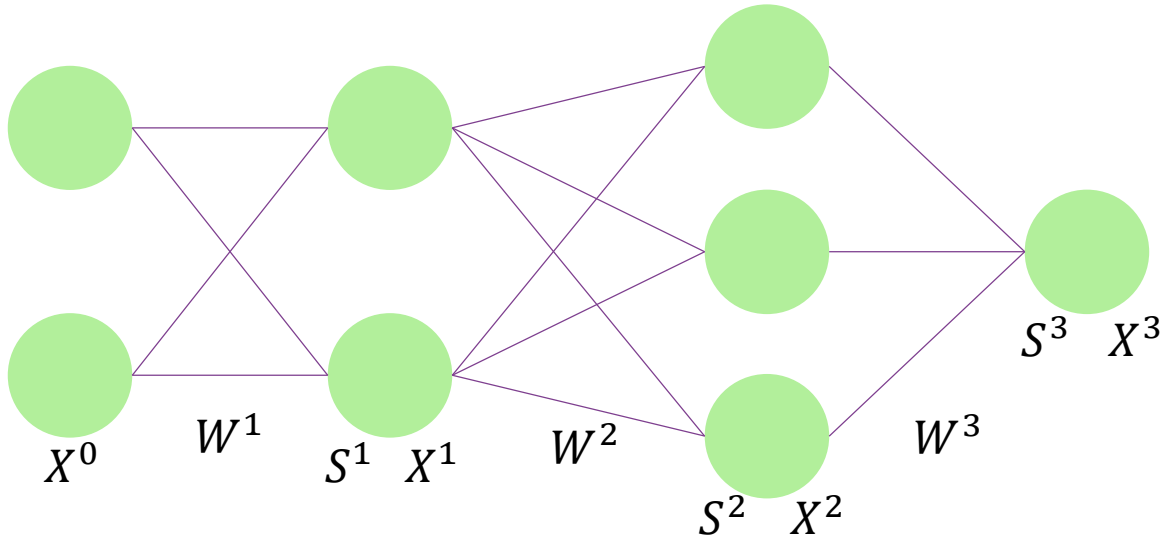
$$\delta_j^{(L)} = -2 * (y_n - s_1^{(L)})$$

$$\delta_j^{(l)} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2(s_j^{(l)})$$

$$\frac{\partial e_n}{\partial w_{i1}^{(L)}} = \frac{\partial e_n}{\partial s_1^{(L)}} * \frac{\partial s_1^{(L)}}{\partial w_{i1}^{(L)}} = \delta_1^{(L)} * x_i^{(L-1)} = -2 * (y_n - s_1^{(L)}) * x_i^{(L-1)}$$

$$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \frac{\partial e_n}{\partial s_j^{(l)}} * \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} * x_i^{(l-1)} = \{\sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2\left(s_j^{(l)}\right)\} * x_i^{(l-1)}$$

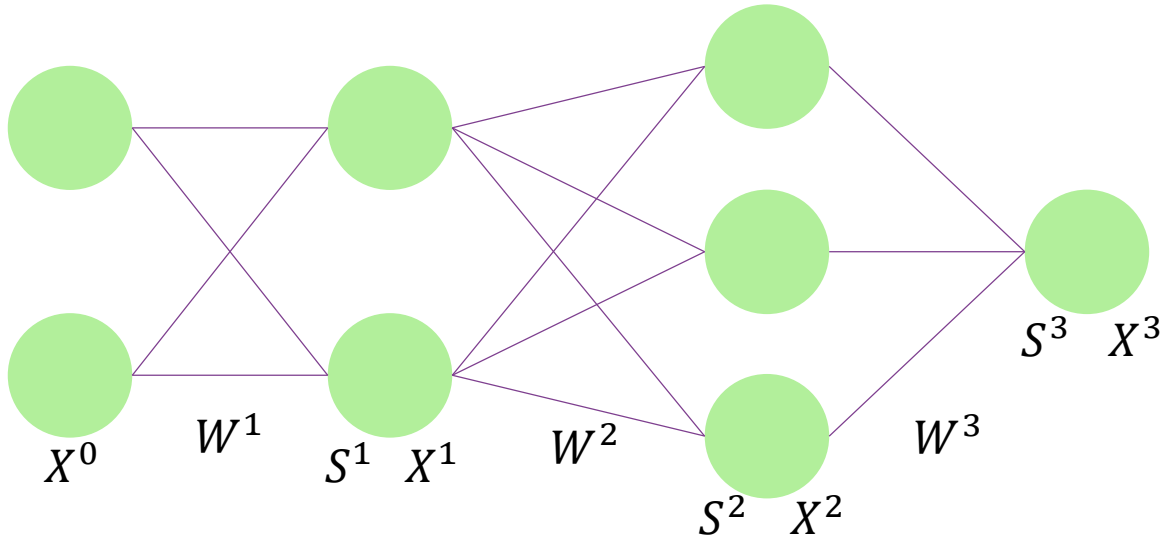
Derivate – summary



$$\frac{\partial e_n}{\partial w_{i1}^{(L)}} = \frac{\partial e_n}{\partial s_1^{(L)}} * \frac{\partial s_1^{(L)}}{\partial w_{i1}^{(L)}} = \delta_1^{(L)} * x_i^{(L-1)} = -2 * (y_n - s_1^{(L)}) * x_i^{(L-1)}$$

$$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \frac{\partial e_n}{\partial s_j^{(l)}} * \frac{\partial s_j^{(l)}}{\partial w_{ij}^{(l)}} = \delta_j^{(l)} * x_i^{(l-1)} = \{\sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2\left(s_j^{(l)}\right)\} * x_i^{(l-1)}$$

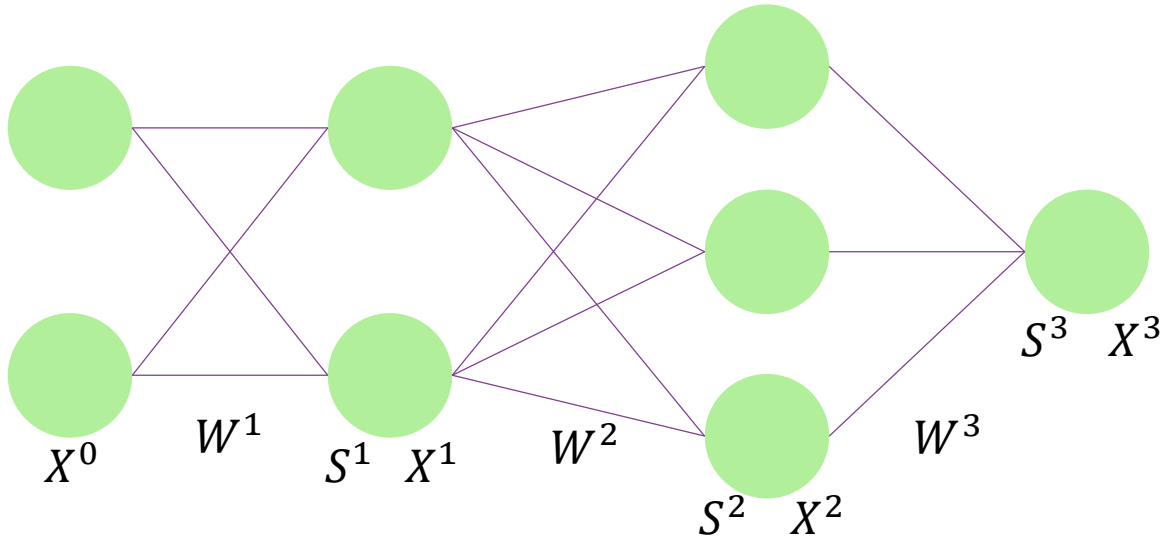
Derivate – summary



$$\frac{\partial e_n}{\partial w_{i1}^{(L)}} = -2 * (y_n - s_1^{(L)}) * x_i^{(L-1)}$$

$$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \{\sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2\left(s_j^{(l)}\right)\} * x_i^{(l-1)}$$

Derivate – summary



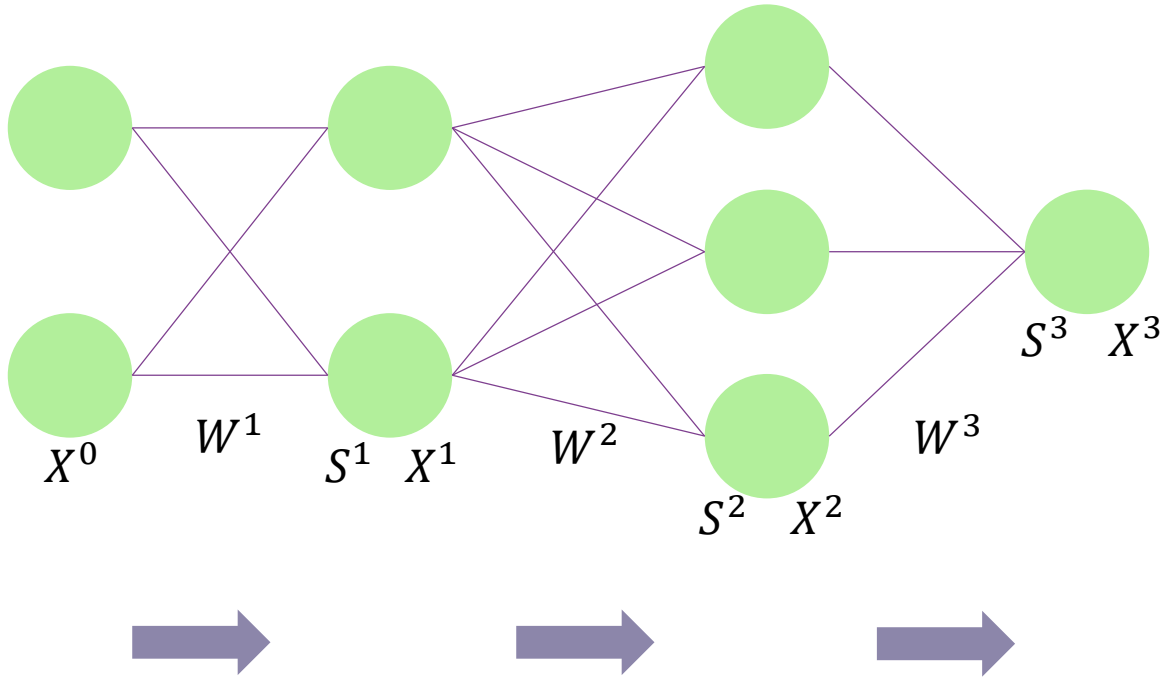
$$\frac{\partial e_n}{\partial w_{i1}^{(L)}} = -2 * (y_n - s_1^{(L)}) * x_i^{(L-1)}$$

$$\frac{\partial e_n}{\partial w_{ij}^{(l)}} = \{\sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2\left(s_j^{(l)}\right)\} * x_i^{(l-1)}$$

Formula summery

Forward

1. Compute score(s)
2. Compute output(x)

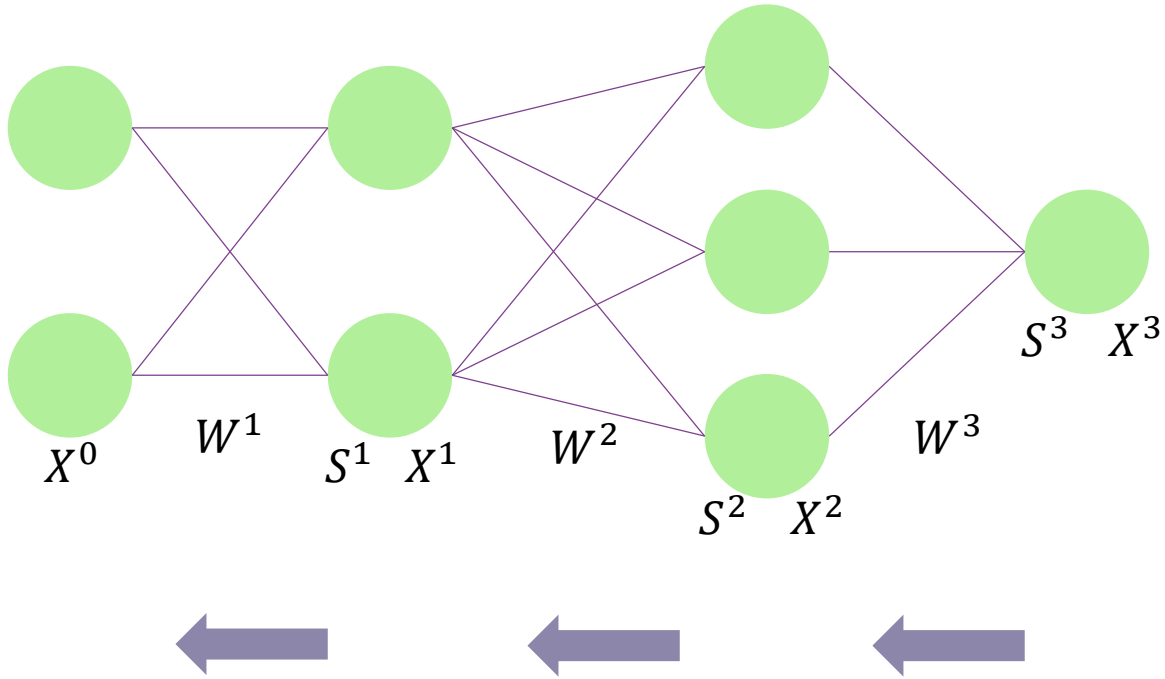


$$s_j^{(l)} = \sum_{i=0}^{d^{l-1}} w_{ij}^{(l)} x_i^{(l-1)} \quad \Rightarrow \quad s^{(l)} = w^{(l)} * x^{(l-1)}$$

$$x_j^{(l)} = \tanh\left(s_j^{(l)}\right) \qquad \Rightarrow \mathrm{x}^{(l)} = \tanh(s^{(l)})$$

Backward

Compute delta(δ)



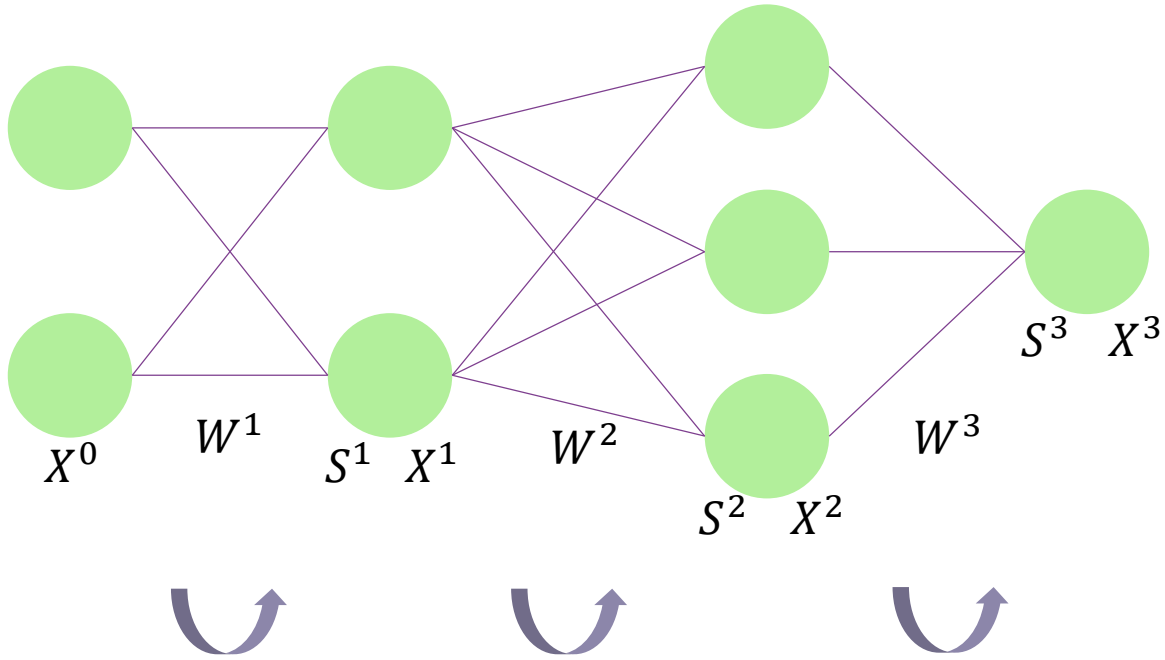
$$\delta_j^{(L)} = -2 * \left(y_n - s_1^{(L)}\right) \qquad \Rightarrow \ \delta^{(L)} = -2(y_n - s^{(L)})$$

$$\delta_j^{(l)} = \sum_{k=1}^{d^{l+1}} \delta_j^{(l+1)} * w_{jk}^{(l)} * sech^2\left(s_j^{(l)}\right) \qquad \Rightarrow \delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \mathrm{sech}^2(s^{(l)})$$

Update

Update the weight by output and delta

$$W_{ij}^{(l)} = W_{ij}^{(l)} - \eta * x_i^{(l-1)} * \delta_j^{(l)}$$



Implementation

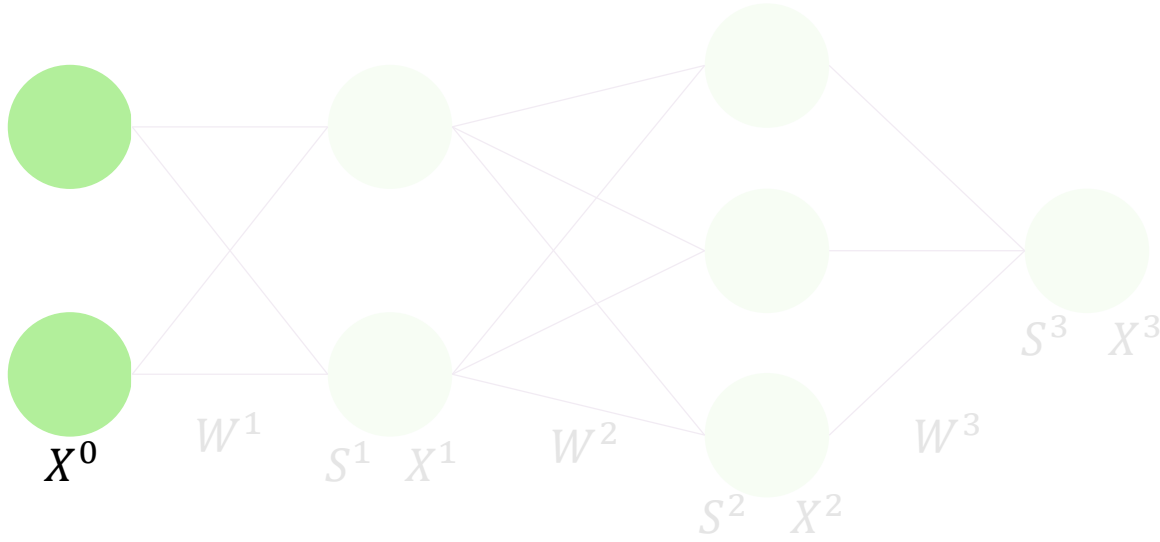
Take (1, -1) as example

idea output is 0

eta is 1.0

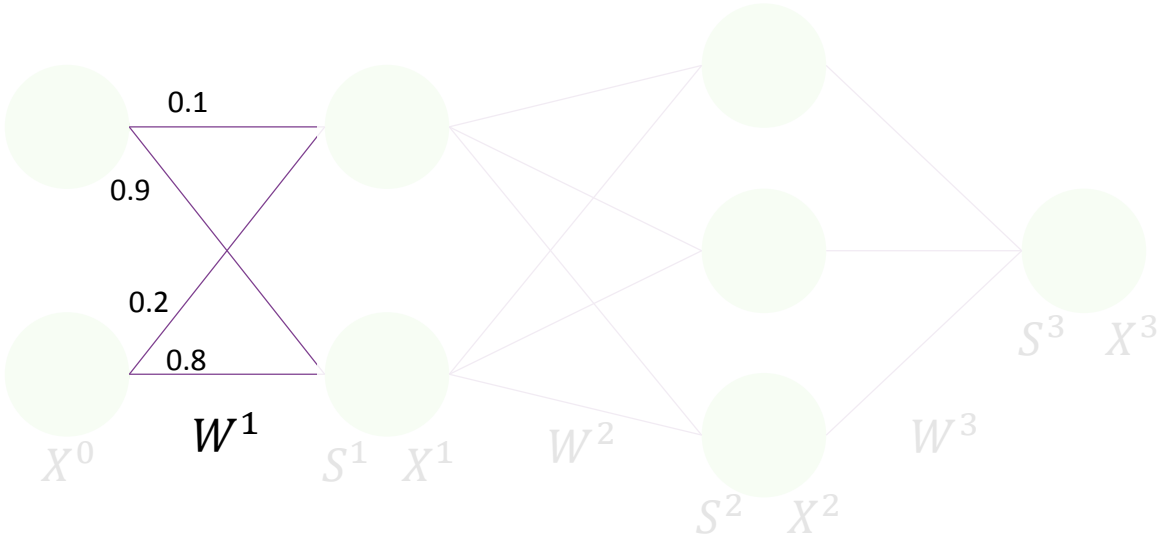
Initialize

$$X^0 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

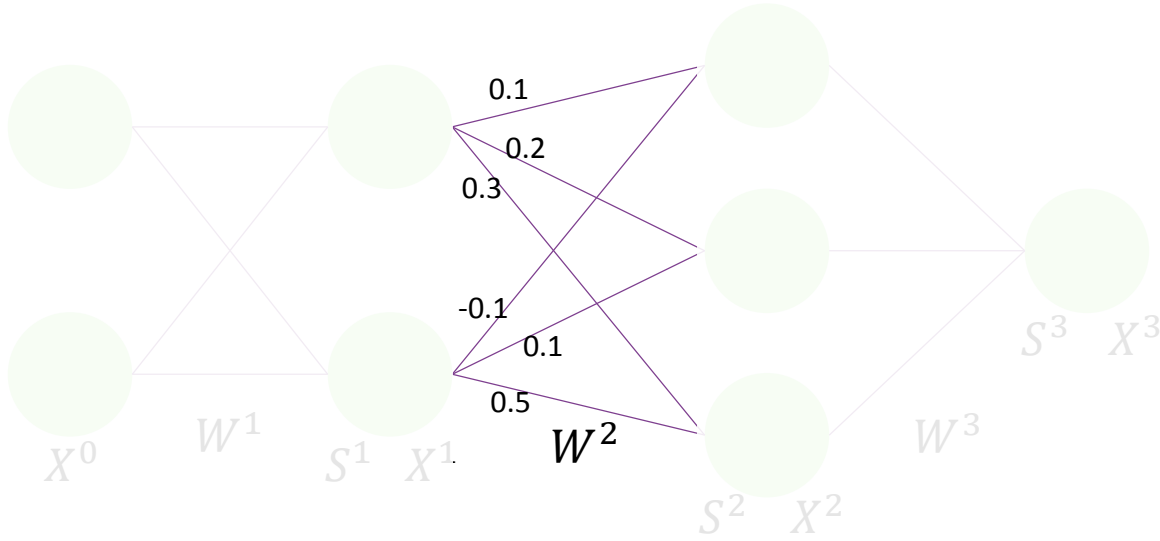


Initialize

$$W^1 = \begin{bmatrix} 0.1 & 0.2 \\ 0.9 & 0.8 \end{bmatrix}$$



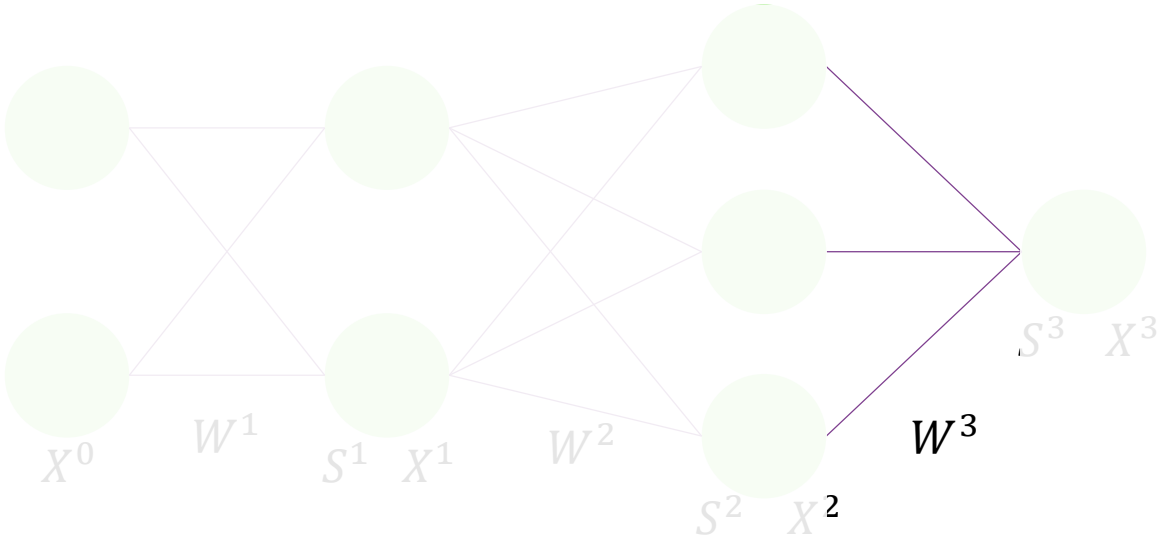
Initialize



$$W^2 = \begin{bmatrix} 0.1 & -0.1 \\ 0.2 & 0.1 \\ 0.3 & 0.5 \end{bmatrix}$$

Initialize

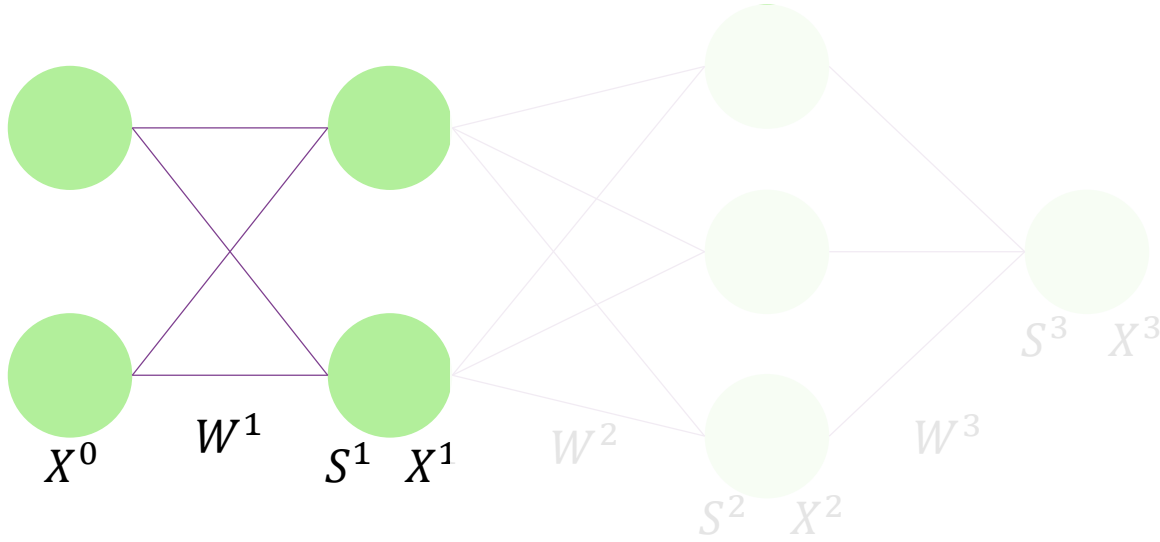
$$W^3 = [-2 \quad 1 \quad 3]$$



1st Round

(1 round) - Forward

$$s^{(l)} = w^{(l)} * x^{(l-1)} \qquad \mathrm{x}^{(l)} = \tanh(s^{(l)})$$

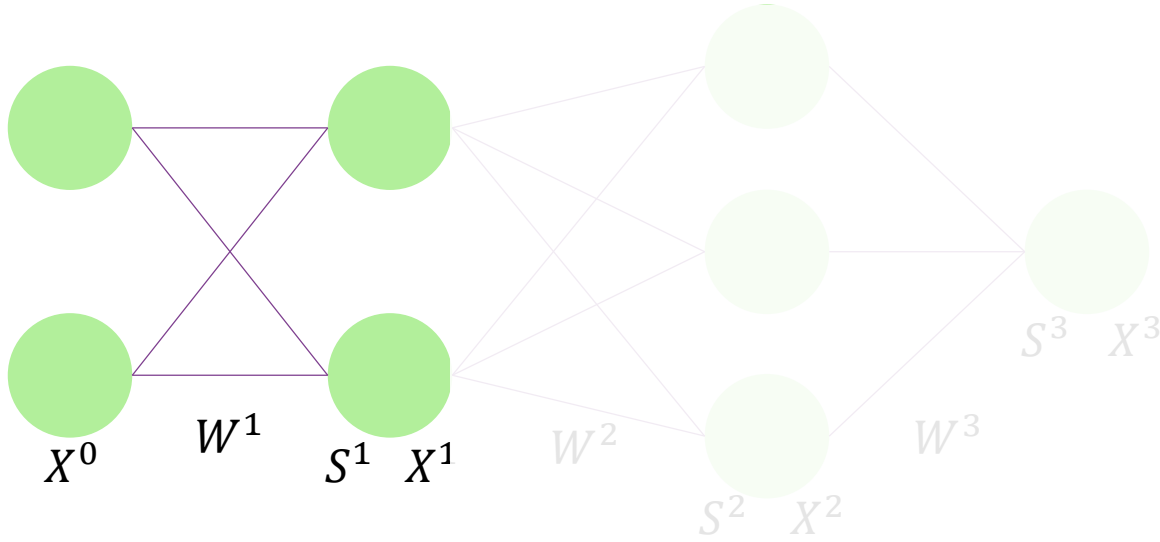


(1 round) - Forward

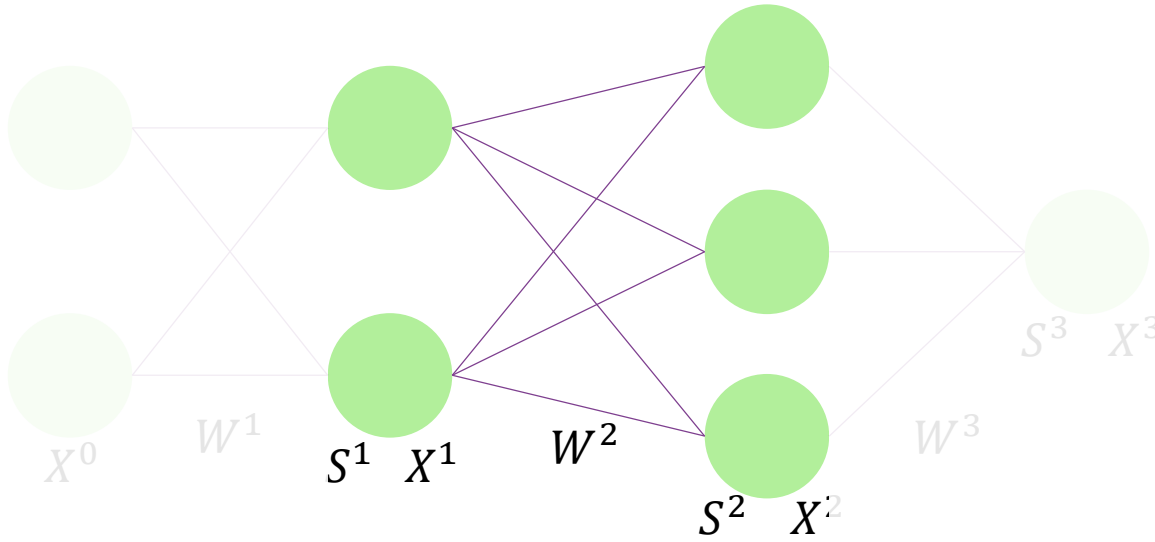
$s^{(l)} = w^{(l)} * x^{(l-1)}$ $\quad$ $\mathrm{x}^{(l)} = \tanh(s^{(l)})$

$$S^{(1)} = \mathrm{W}^{(1)} * X^{(0)} = \begin{bmatrix} 0.1 & 0.2 \\ 0.9 & 0.8 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}$$

$$X^{(1)} = \tanh\left(S^{(1)}\right) = \tanh\left(\begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}\right) = \begin{bmatrix} -0.09966799 \\ 0.09966799 \end{bmatrix}$$



(1 round) - Forward



$$s^{(l)} = w^{(l)} * x^{(l-1)} \qquad \mathrm{x}^{(l)} = \tanh(s^{(l)})$$

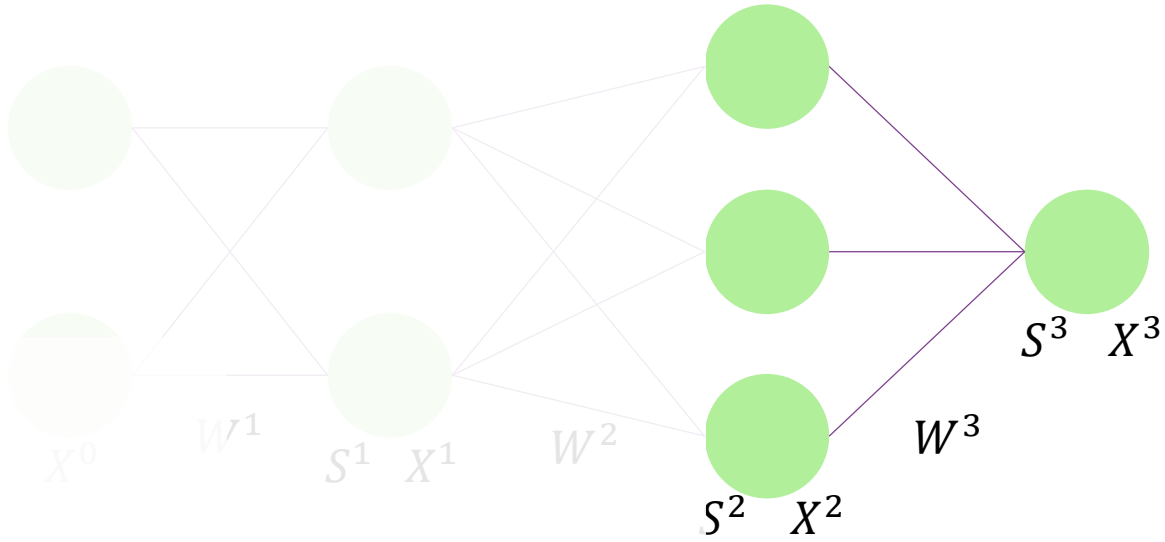
$$S^{(1)} = \mathrm{W}^{(1)} * X^{(0)} = \begin{bmatrix} 0.1 & 0.2 \\ 0.9 & 0.8 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}$$

$$X^{(1)} = \tanh\left(S^{(1)}\right) = \tanh\left(\begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}\right) = \begin{bmatrix} -0.09966799 \\ 0.09966799 \end{bmatrix}$$

$$S^{(2)} = \mathrm{W}^{(2)} * X^{(1)} = \begin{bmatrix} 0.1 & -0.1 \\ 0.2 & 0.1 \\ 0.3 & 0.5 \end{bmatrix} \begin{bmatrix} -0.09966799 \\ 0.09966799 \end{bmatrix} = \begin{bmatrix} -0.0199336 \\ -0.0099668 \\ 0.0199336 \end{bmatrix}$$

$$X^{(2)} = \tanh\left(S^{(2)}\right) = \tanh\left(\begin{bmatrix} -0.0199336 \\ -0.0099668 \\ 0.0199336 \end{bmatrix}\right) = \begin{bmatrix} -0.01993096 \\ -0.00996647 \\ 0.01993096 \end{bmatrix}$$

(1 round) - Forward



$$s^{(l)} = w^{(l)} * x^{(l-1)} \qquad \mathrm{x}^{(l)} = \tanh(s^{(l)})$$

$$S^{(1)} = \mathrm{W}^{(1)} * X^{(0)} = \begin{bmatrix} 0.1 & 0.2 \\ 0.9 & 0.8 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}$$

$$X^{(1)} = \tanh\left(S^{(1)}\right) = \tanh\left(\begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}\right) = \begin{bmatrix} -0.09966799 \\ 0.09966799 \end{bmatrix}$$

$$S^{(2)} = \mathrm{W}^{(2)} * X^{(1)} = \begin{bmatrix} 0.1 & -0.1 \\ 0.2 & 0.1 \\ 0.3 & 0.5 \end{bmatrix} \begin{bmatrix} -0.09966799 \\ 0.09966799 \end{bmatrix} = \begin{bmatrix} -0.0199336 \\ -0.0099668 \\ 0.0199336 \end{bmatrix}$$

$$X^{(2)} = \tanh\left(S^{(2)}\right) = \tanh\left(\begin{bmatrix} -0.0199336 \\ -0.0099668 \\ 0.0199336 \end{bmatrix}\right) = \begin{bmatrix} -0.01993096 \\ -0.00996647 \\ 0.01993096 \end{bmatrix}$$

$$S^{(3)} = \mathrm{W}^{(3)} * X^{(2)} = \begin{bmatrix} -2 & 1 & 3 \end{bmatrix} \begin{bmatrix} -0.01993096 \\ -0.00996647 \\ 0.01993096 \end{bmatrix} = [0.08968833]$$

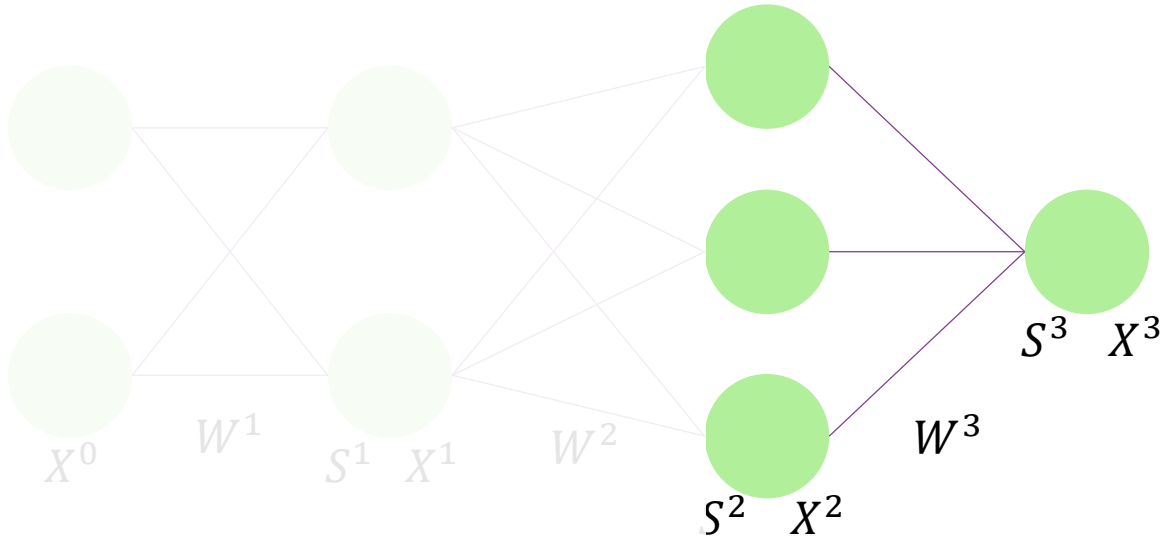
$$X^{(3)} = \mathrm{linear}\left(S^{(3)}\right) = [0.08968833]$$

(1 round) - Backward

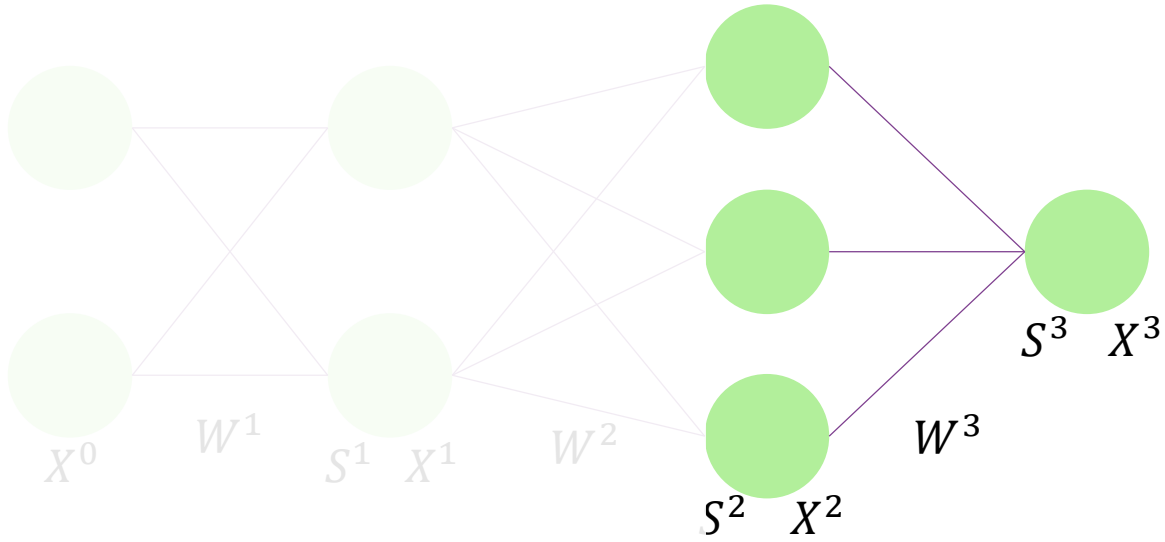
$$\delta^{(L)} = -2(y_n - s^{(L)})$$

$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \operatorname{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \operatorname{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.08968833] \qquad y_n = 0$$



(1 round) - Backward



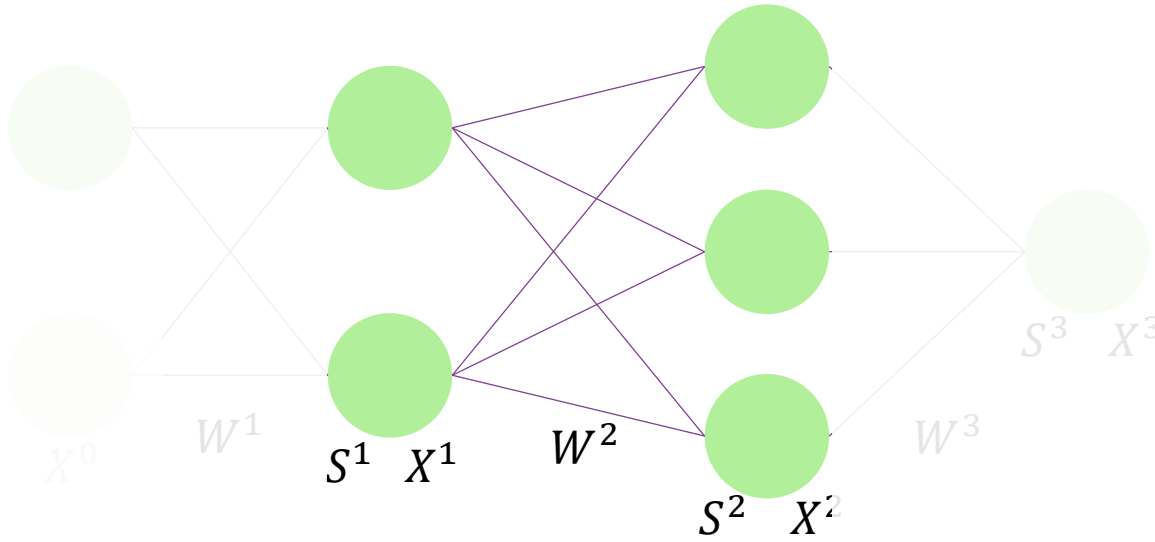
$$\delta^{(L)} = -2(y_n - s^{(L)})$$

$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \operatorname{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \operatorname{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.08968833] \qquad y_n = 0$$

$$\delta^{(3)} = -2(y_n - s^{(3)}) = -2 * (0 - [0.08968833]) = 0.17937665$$

(1 round) - Backward



$$\delta^{(L)} = -2(y_n - s^{(L)})$$

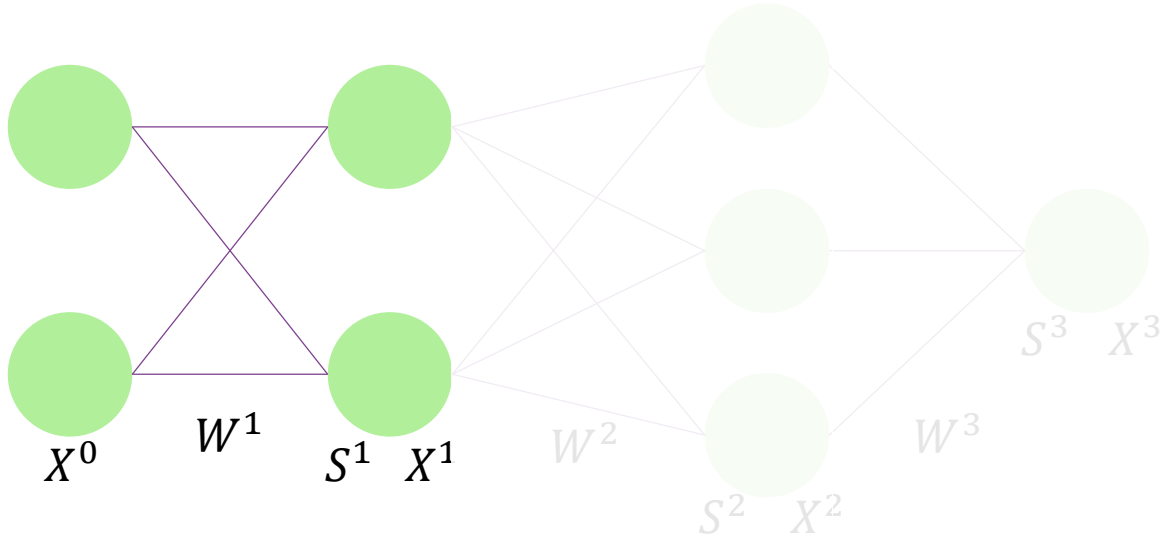
$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \text{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \text{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.08968833] \qquad y_n = 0$$

$$\delta^{(3)} = -2(y_n - s^{(3)}) = -2 * (0 - [0.08968833]) = 0.17937665$$

$$\delta^{(2)} = W^{T(3)} * \delta^{(3)} \otimes \text{sech}^2(s^{(2)}) = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} * [0.17937665\,] \otimes \text{sech}^2\left(\begin{bmatrix} -0.0199336 \\ -0.0099668 \\ 0.0199336 \end{bmatrix}\right) = \begin{bmatrix} -0.35861079 \\ 0.17935884 \\ 0.53791619 \end{bmatrix}$$

(1 round) - Backward



$$\delta^{(L)} = -2(y_n - s^{(L)})$$

$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \operatorname{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \operatorname{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.08968833] \qquad y_n = 0$$

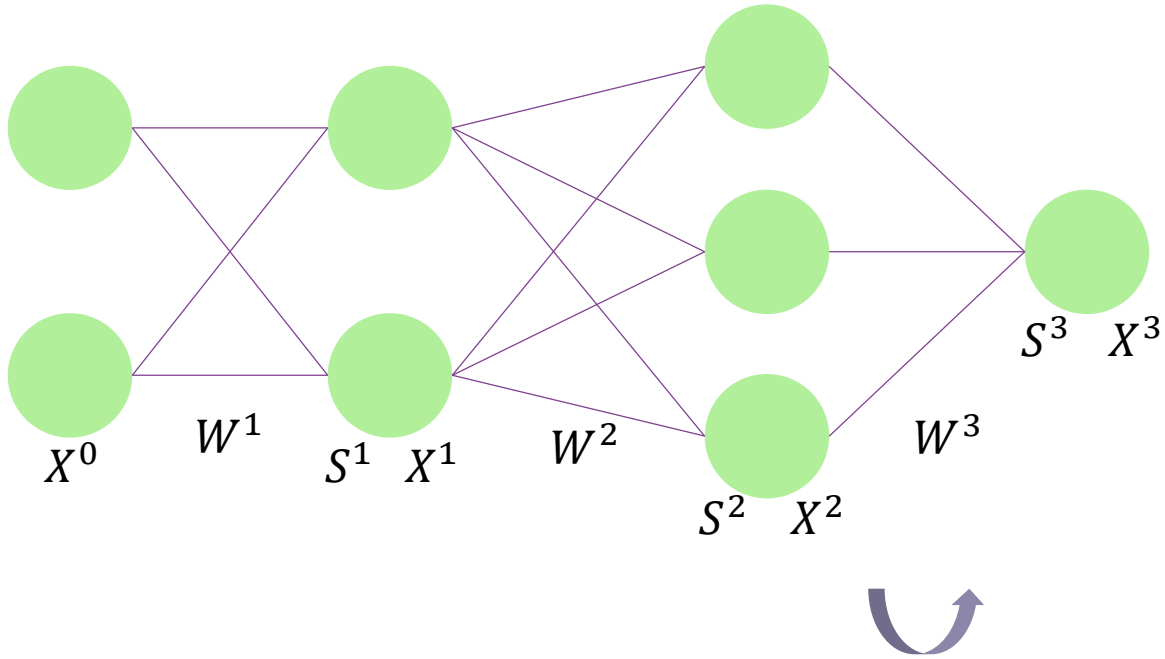
$$\delta^{(3)} = -2(y_n - s^{(3)}) = -2 * (0 - [0.08968833]) = 0.17937665$$

$$\delta^{(2)} = W^{T(3)} * \delta^{(3)} \otimes \operatorname{sech}^2(s^{(2)}) = \begin{bmatrix} -2 \\ 1 \\ 3 \end{bmatrix} * [0.17937665\,] \otimes \operatorname{sech}^2\left(\begin{bmatrix} -0.0199336 \\ -0.0099668 \\ 0.0199336 \end{bmatrix}\right) = \begin{bmatrix} -0.35861079 \\ 0.17935884 \\ 0.53791619 \end{bmatrix}$$

$$\delta^{(1)} = W^{T(2)} * \delta^{(2)} \otimes \operatorname{sech}^2(s^{(1)}) = \begin{bmatrix} 0.1 & 0.2 & 0.3 \\ -0.1 & 0.1 & 0.5 \end{bmatrix} * \begin{bmatrix} -0.35861079 \\ 0.17935884 \\ 0.53791619 \end{bmatrix} \otimes \operatorname{sech}^2\left(\begin{bmatrix} -0.1 \\ 0.1 \end{bmatrix}\right) = \begin{bmatrix} 0.15978239 \\ 0.3195489 \end{bmatrix}$$

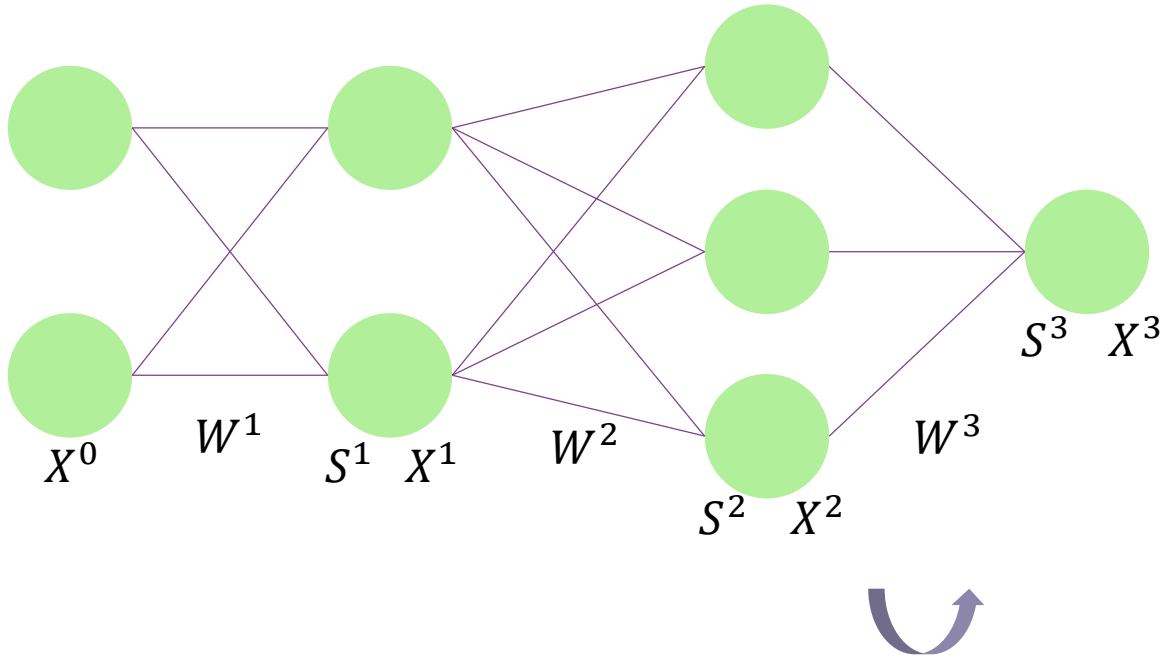
(1 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$



(1 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$

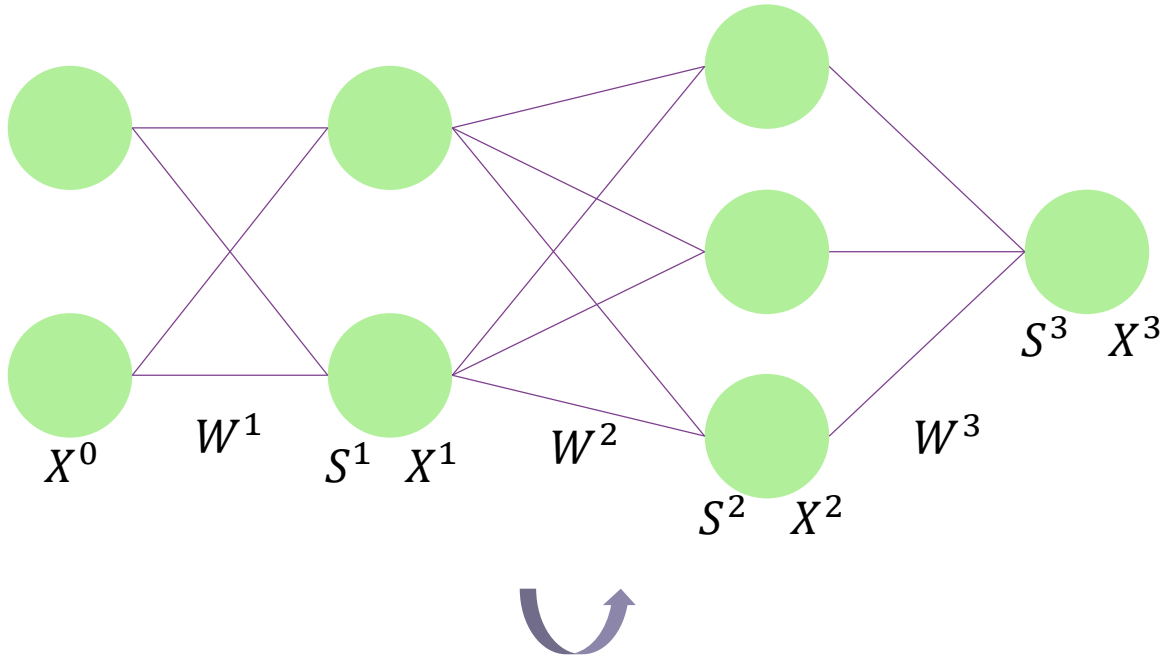


$$W^{(3)} = W^{(3)} - \eta * \delta^{(3)} * X^{T^{(2)}} = [-2 \quad 1 \quad 3] - 1.0 * 0.17937665 * [-0.01993096 \quad -0.00996647 \quad 0.01993096]$$

$$= [-1.99982124 \quad 1.00008939 \quad 2.99982124]$$

(1 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$



$$W^{(3)} = W^{(3)} - \eta * \delta^{(3)} * X^{T^{(2)}} = [-2 \quad 1 \quad 3] - 1.0 * 0.17937665 * [-0.01993096 \quad -0.00996647 \quad 0.01993096]$$

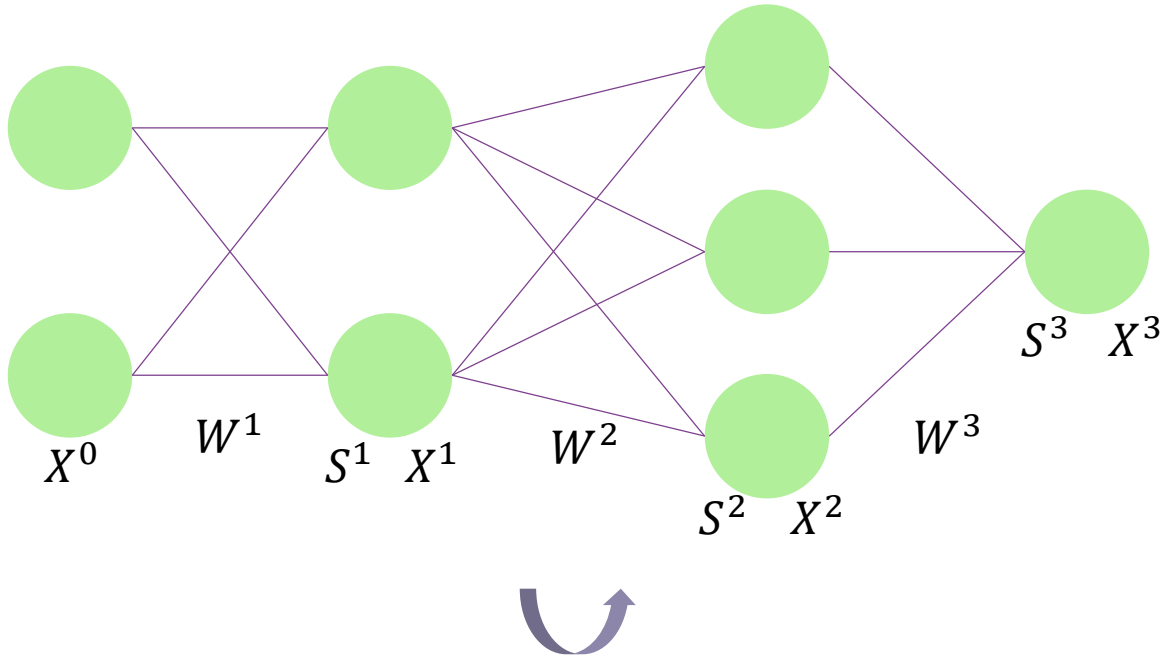
$$= [-1.99982124 \quad 1.00008939 \quad 2.99982124]$$

$$W^{(2)} = W^{(2)} - \eta * \delta^{(2)} * X^{T^{(1)}} = \begin{bmatrix} 0.1 & -0.1 \\ 0.2 & 0.1 \\ 0.3 & 0.5 \end{bmatrix} - 1.0 * \begin{bmatrix} -0.35861079 \\ 0.17935884 \\ 0.53791619 \end{bmatrix} * [-0.09966799 \quad 0.09966799]$$

$$= \begin{bmatrix} 0.0982129 & -0.0982129 \\ 0.2008938 & 0.09910618 \\ 0.3026806 & 0.49731935 \end{bmatrix}$$

(1 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$

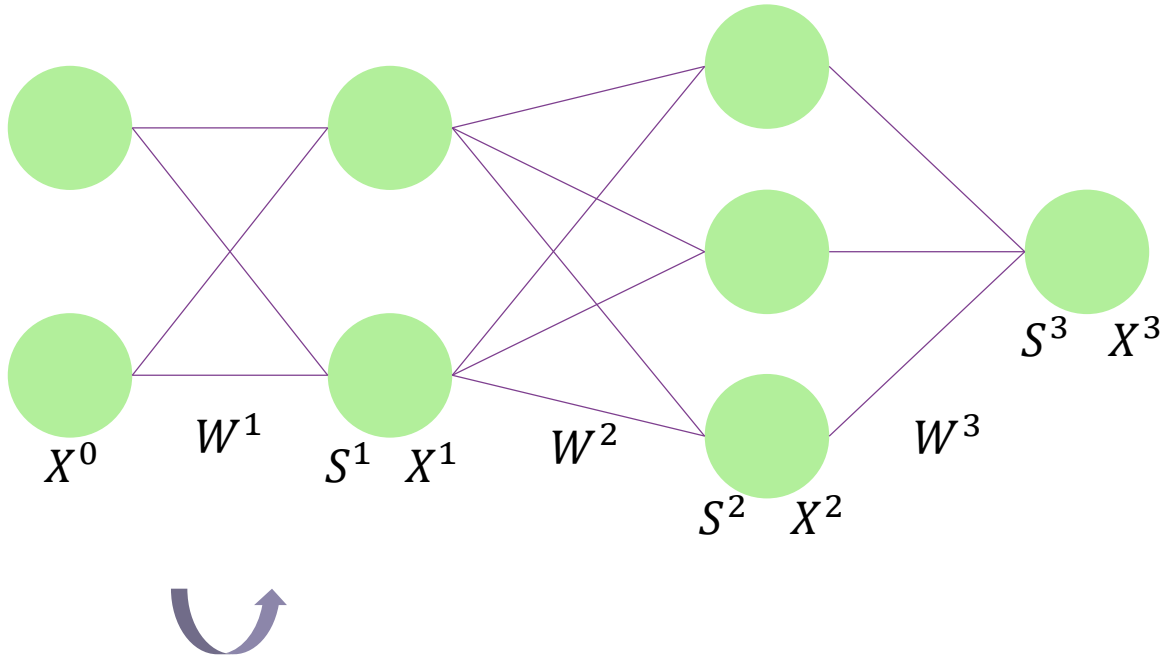


$$W^{(3)} = [-1.99982124 \quad 1.00008939 \quad 2.99982124]$$

$$W^{(2)} = \begin{bmatrix} 0.0982129 & -0.0982129 \\ 0.2008938 & 0.09910618 \\ 0.3026806 & 0.49731935 \end{bmatrix}$$

(1 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$



$$W^{(3)} = [-1.99982124 \quad 1.00008939 \quad 2.99982124]$$

$$W^{(2)} = \begin{bmatrix} 0.0982129 & -0.0982129 \\ 0.2008938 & 0.09910618 \\ 0.3026806 & 0.49731935 \end{bmatrix}$$

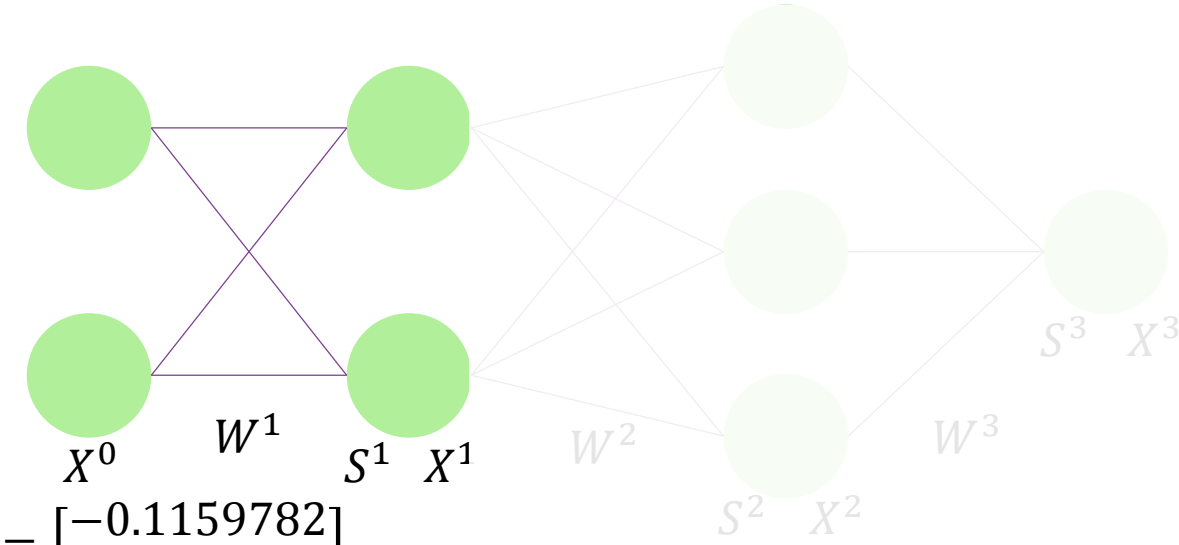
$$W^{(1)} = W^{(1)} - \eta * \delta^{(1)} * X^{T^{(0)}} = \begin{bmatrix} 0.1 & 0.2 \\ 0.9 & 0.8 \end{bmatrix} - 1.0 * \begin{bmatrix} 0.15978239 \\ 0.3195489 \end{bmatrix} * [1 \quad -1]$$

$$= \begin{bmatrix} 0.09201088 & 0.20798912 \\ 0.88402255 & 0.81597745 \end{bmatrix}$$

2nd Round

(2 round) - Forward

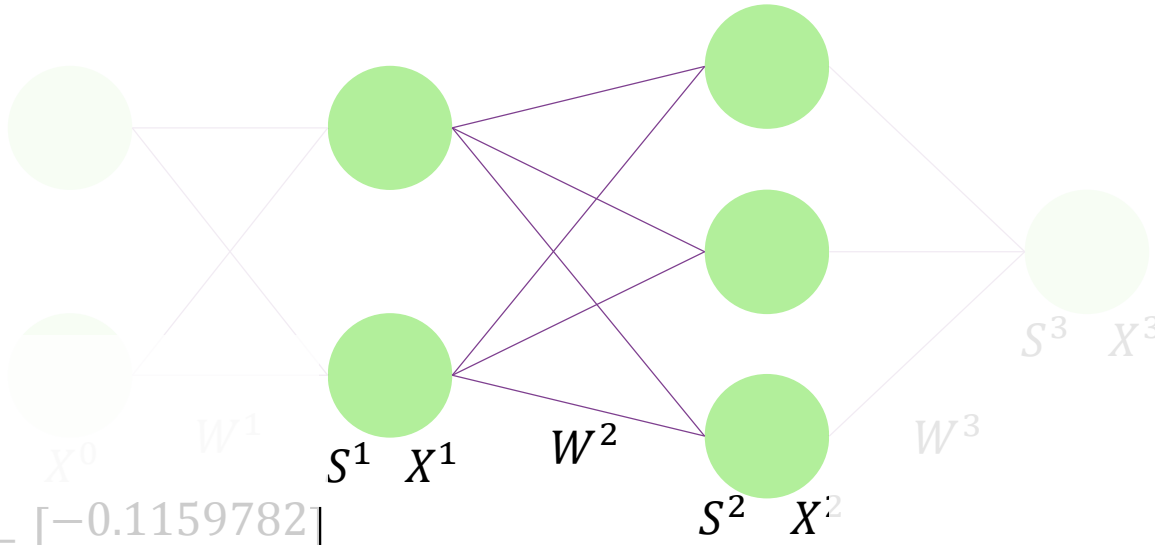
$s^{(l)} = w^{(l)} * x^{(l-1)}$ $\quad$ $\mathrm{x}^{(l)} = \tanh(s^{(l)})$



$$S^{(1)} = \mathrm{W}^{(1)} * X^{(0)} = \begin{bmatrix} 0.09201088 & 0.20798912 \\ 0.88402255 & 0.81597745 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.1159782 \\ 0.06804511 \end{bmatrix}$$

$$X^{(1)} = \tanh\left(S^{(1)}\right) = \tanh\left(\begin{bmatrix} -0.1159782 \\ 0.06804511 \end{bmatrix}\right) = \begin{bmatrix} -0.11546102 \\ 0.06794028 \end{bmatrix}$$

(2 round) - Forward



$$s^{(l)} = w^{(l)} * x^{(l-1)} \qquad \mathrm{x}^{(l)} = \tanh(s^{(l)})$$

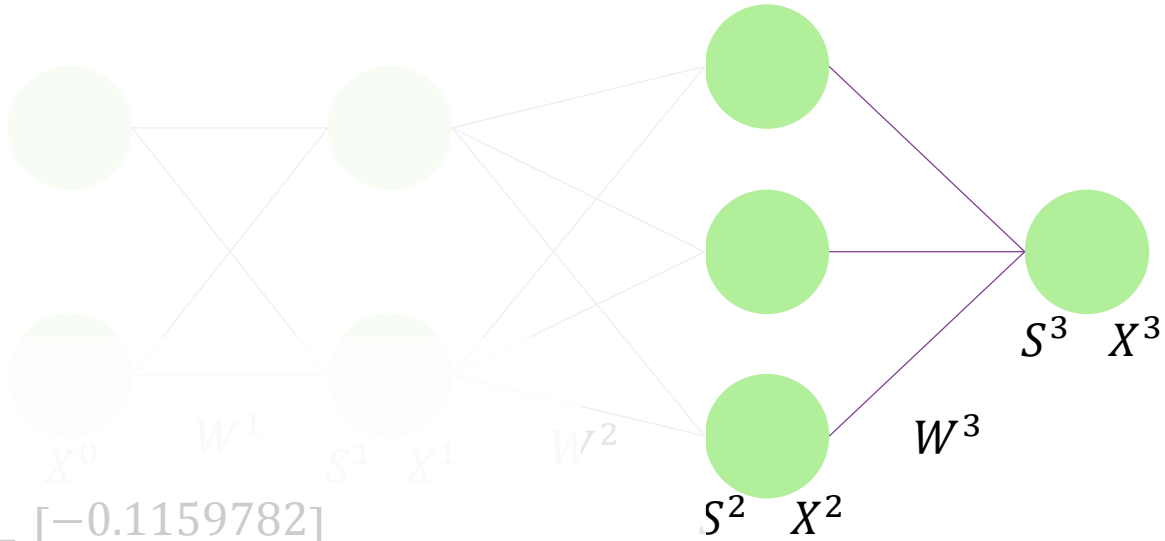
$$S^{(1)} = \mathrm{W}^{(1)} * X^{(0)} = \begin{bmatrix} 0.09201088 & 0.20798912 \\ 0.88402255 & 0.81597745 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.1159782 \\ 0.06804511 \end{bmatrix}$$

$$X^{(1)} = \tanh\left(S^{(1)}\right) = \tanh\left(\begin{bmatrix} -0.1159782 \\ 0.06804511 \end{bmatrix}\right) = \begin{bmatrix} -0.11546102 \\ 0.06794028 \end{bmatrix}$$

$$S^{(2)} = \mathrm{W}^{(2)} * X^{(1)} = \begin{bmatrix} 0.0982129 & -0.0982129 \\ 0.2008938 & 0.09910618 \\ 0.3026806 & 0.49731935 \end{bmatrix} \begin{bmatrix} -0.11546102 \\ 0.06794028 \end{bmatrix} = \begin{bmatrix} -0.01801237 \\ -0.0164621 \\ -0.0011598 \end{bmatrix}$$

$$X^{(2)} = \tanh\left(S^{(2)}\right) = \tanh\left(\begin{bmatrix} -0.01801237 \\ -0.0164621 \\ -0.0011598 \end{bmatrix}\right) = \begin{bmatrix} -0.01801043 \\ -0.01646061 \\ -0.0011598 \end{bmatrix}$$

(2 round) - Forward



$$s^{(l)} = w^{(l)} * x^{(l-1)} \qquad \mathrm{x}^{(l)} = \tanh(s^{(l)})$$

$$S^{(1)} = \mathrm{W}^{(1)} * X^{(0)} = \begin{bmatrix} 0.09201088 & 0.20798912 \\ 0.88402255 & 0.81597745 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.1159782 \\ 0.06804511 \end{bmatrix}$$

$$X^{(1)} = \tanh\left(S^{(1)}\right) = \tanh\left(\begin{bmatrix} -0.1159782 \\ 0.06804511 \end{bmatrix}\right) = \begin{bmatrix} -0.11546102 \\ 0.06794028 \end{bmatrix}$$

$$S^{(2)} = \mathrm{W}^{(2)} * X^{(1)} = \begin{bmatrix} 0.0982129 & -0.0982129 \\ 0.2008938 & 0.09910618 \\ 0.3026806 & 0.49731935 \end{bmatrix} \begin{bmatrix} -0.11546102 \\ 0.06794028 \end{bmatrix} = \begin{bmatrix} -0.01801237 \\ -0.0164621 \\ -0.0011598 \end{bmatrix}$$

$$X^{(2)} = \tanh\left(S^{(2)}\right) = \tanh\left(\begin{bmatrix} -0.01801237 \\ -0.0164621 \\ -0.0011598 \end{bmatrix}\right) = \begin{bmatrix} -0.01801043 \\ -0.01646061 \\ -0.0011598 \end{bmatrix}$$

$$S^{(3)} = \mathrm{W}^{(3)} * X^{(2)} = [-1.99982124 \quad 1.00008939 \quad 2.99982124] \begin{bmatrix} -0.01801043 \\ -0.01646061 \\ -0.0011598 \end{bmatrix} = [0.01607636]$$

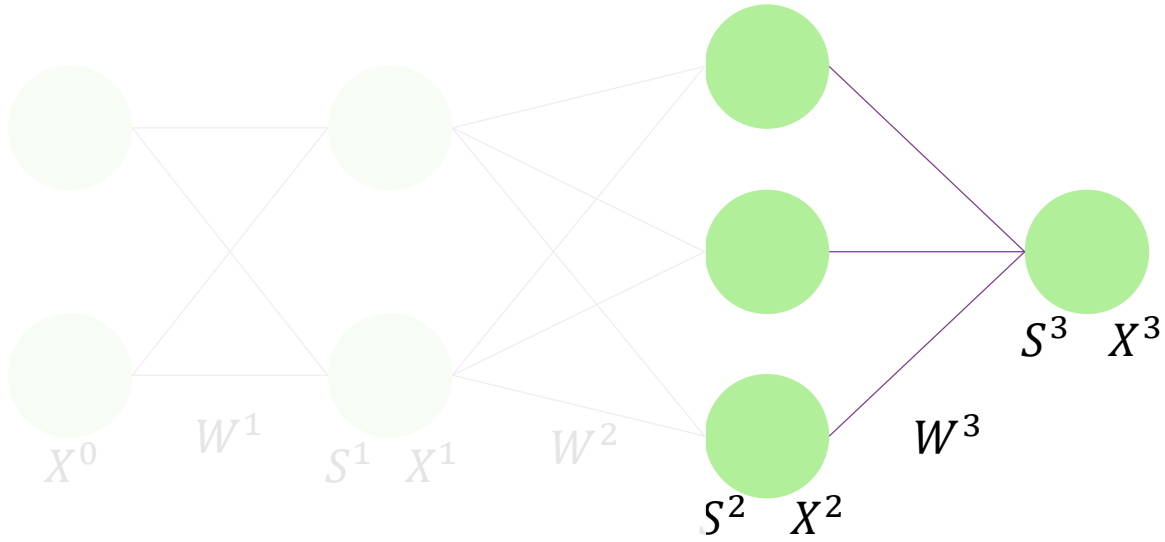
$$X^{(3)} = \mathrm{linear}\left(S^{(3)}\right) = [0.01607636]$$

(2 round) - Backward

$$\delta^{(L)} = -2(y_n - s^{(L)})$$

$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \operatorname{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \operatorname{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.01607636] \qquad y_n = 0$$



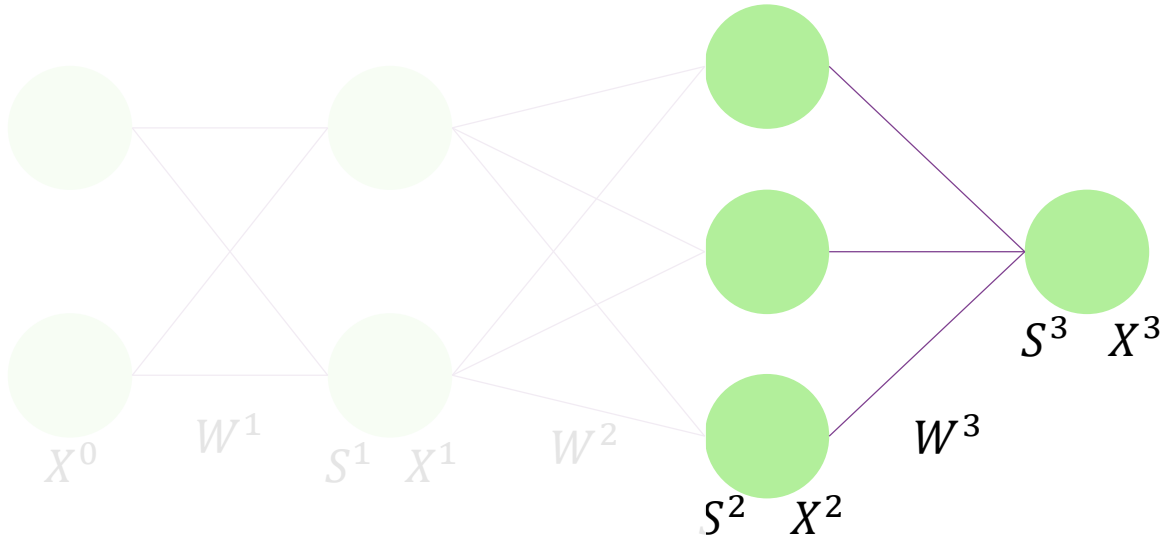
(2 round) - Backward

$$\delta^{(L)} = -2(y_n - s^{(L)})$$

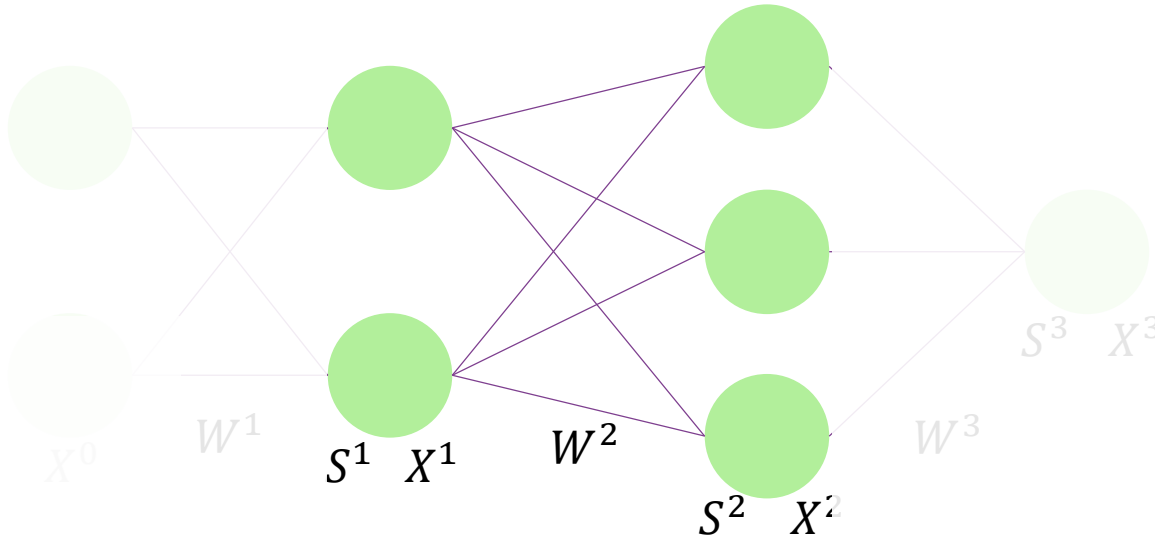
$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \operatorname{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \operatorname{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.01607636] \qquad y_n = 0$$

$$\delta^{(3)} = -2(y_n - s^{(3)}) = -2 * (0 - [0.01607636]) = 0.03215272$$



(2 round) - Backward



$$\delta^{(L)} = -2(y_n - s^{(L)})$$

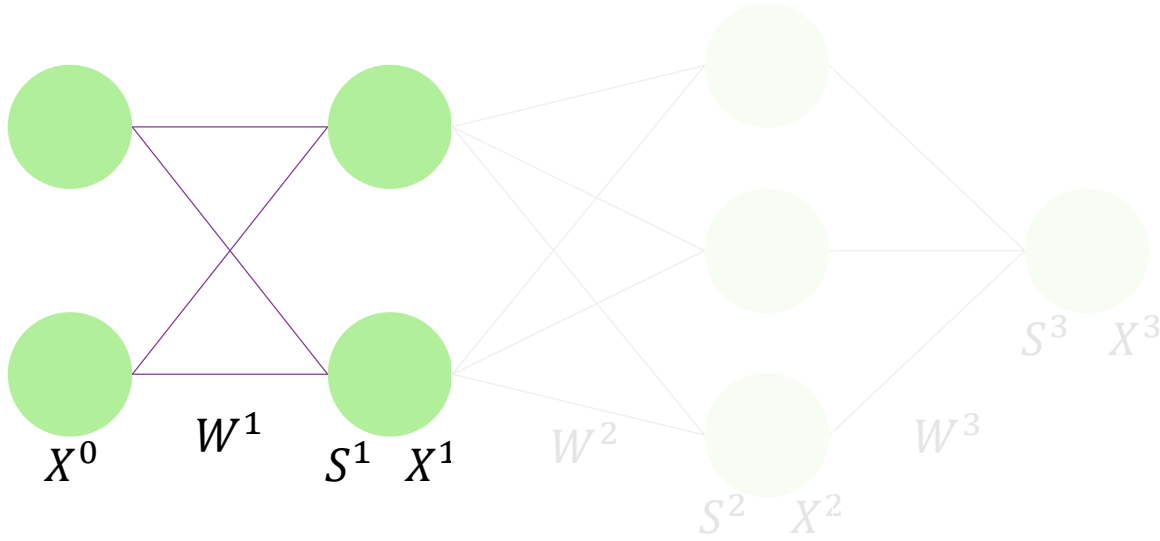
$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \text{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \text{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.01607636] \qquad y_n = 0$$

$$\delta^{(3)} = -2(y_n - s^{(3)}) = -2 * (0 - [0.01607636]) = 0.03215272$$

$$\delta^{(2)} = W^{T(3)} * \delta^{(3)} \otimes \text{sech}^2(s^{(2)}) = \begin{bmatrix} -1.99982124 \\ 1.00008939 \\ 2.99982124 \end{bmatrix} * [0.03215272] \otimes \text{sech}^2\left(\begin{bmatrix} -0.01801237 \\ -0.0164621 \\ -0.0011598 \end{bmatrix}\right) = \begin{bmatrix} -0.06427884 \\ 0.03214689 \\ 0.0964523 \end{bmatrix}$$

(2 round) - Backward



$$\delta^{(L)} = -2(y_n - s^{(L)})$$

$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \operatorname{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \operatorname{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.01607636] \qquad y_n = 0$$

$$\delta^{(3)} = -2(y_n - s^{(3)}) = -2 * (0 - [0.01607636]) = 0.03215272$$

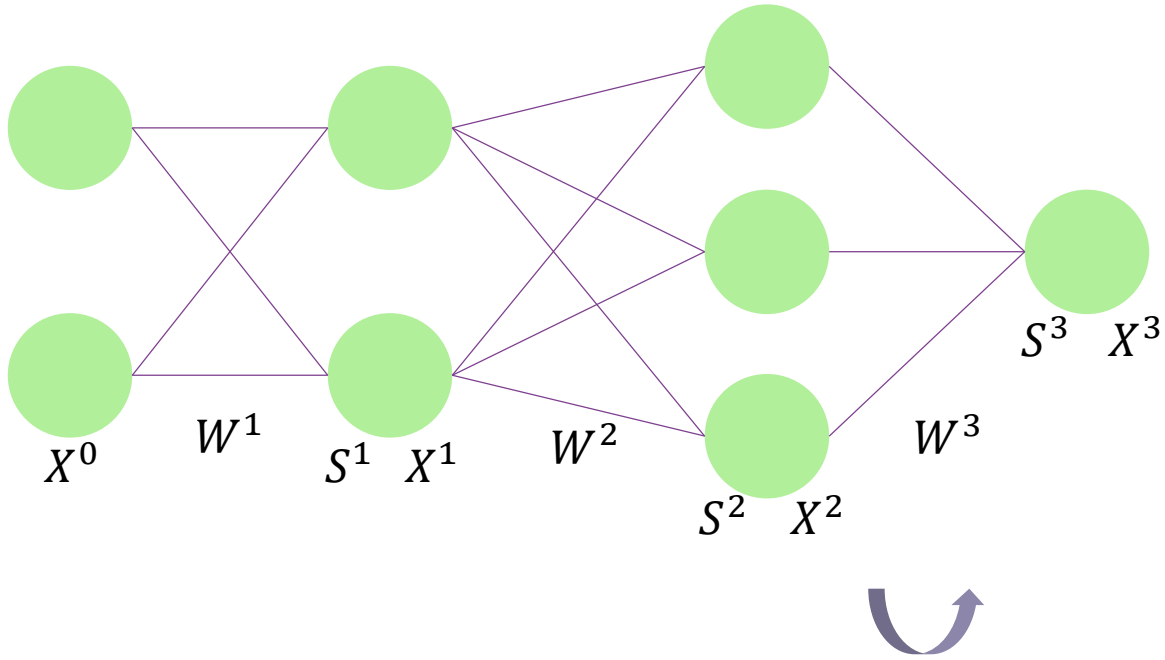
$$\delta^{(2)} = W^{T(3)} * \delta^{(3)} \otimes \operatorname{sech}^2(s^{(2)}) = \begin{bmatrix} -1.99982124 \\ 1.00008939 \\ 2.99982124 \end{bmatrix} * [0.03215272] \otimes \operatorname{sech}^2\left(\begin{bmatrix} -0.01801237 \\ -0.0164621 \\ -0.0011598 \end{bmatrix}\right) = \begin{bmatrix} -0.06427884 \\ 0.03214689 \\ 0.0964523 \end{bmatrix}$$

$$\delta^{(1)} = W^{T(2)} * \delta^{(2)} \otimes \operatorname{sech}^2(s^{(1)}) = \begin{bmatrix} 0.0982129 & 0.2008938 & 0.3026806 \\ -0.0982129 & 0.09910618 & 0.49731935 \end{bmatrix} * \begin{bmatrix} -0.06427884 \\ 0.03214689 \\ 0.0964523 \end{bmatrix}$$

$$\otimes \operatorname{sech}^2\left(\begin{bmatrix} -0.1159782 \\ 0.06804511 \end{bmatrix}\right) = \begin{bmatrix} 0.02894821 \\ 0.0572013 \end{bmatrix}$$

(2 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$

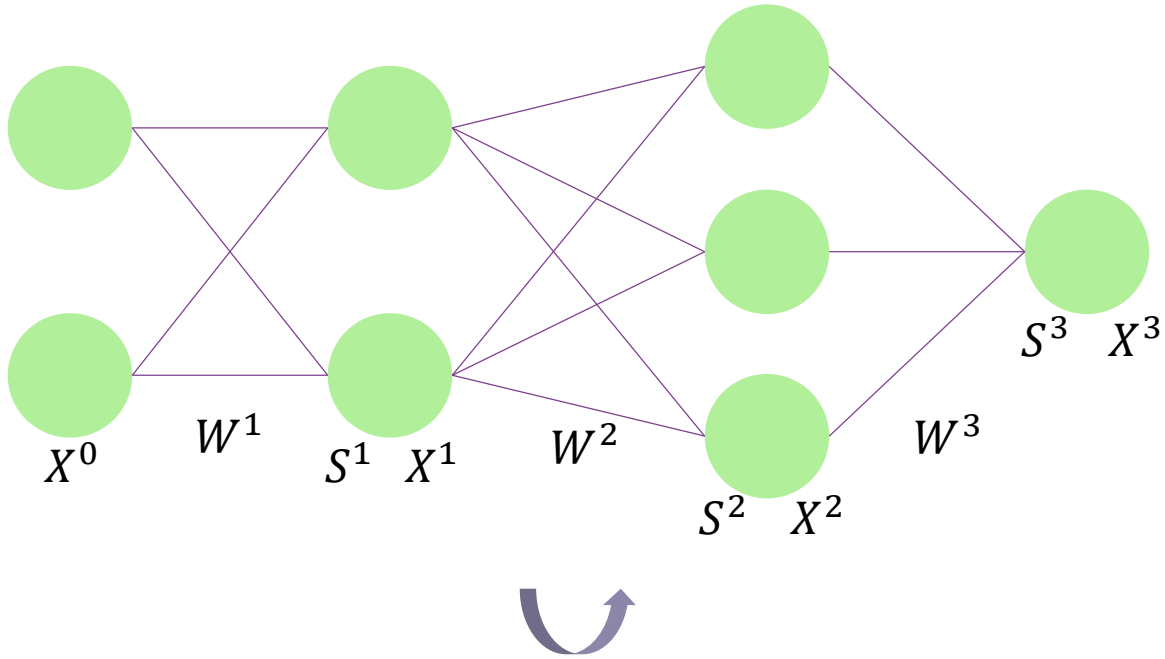


$$W^{(3)} = W^{(3)} - \eta * \delta^{(3)} * X^{T^{(2)}} = [-1.99982124 \quad 1.00008939 \quad 2.99982124] - 1.0 * 0.03215272 * [-0.01801043 \quad -0.01646061 \quad 0.0011598]$$

$$= [-1.99979229 \quad 1.00011585 \quad 2.99982311]$$

(2 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$



$$W^{(3)} = W^{(3)} - \eta * \delta^{(3)} * X^{T^{(2)}} = [-1.99982124 \quad 1.00008939 \quad 2.99982124] - 1.0 * 0.03215272 * [-0.01801043 \quad -0.01646061 \quad 0.0011598]$$

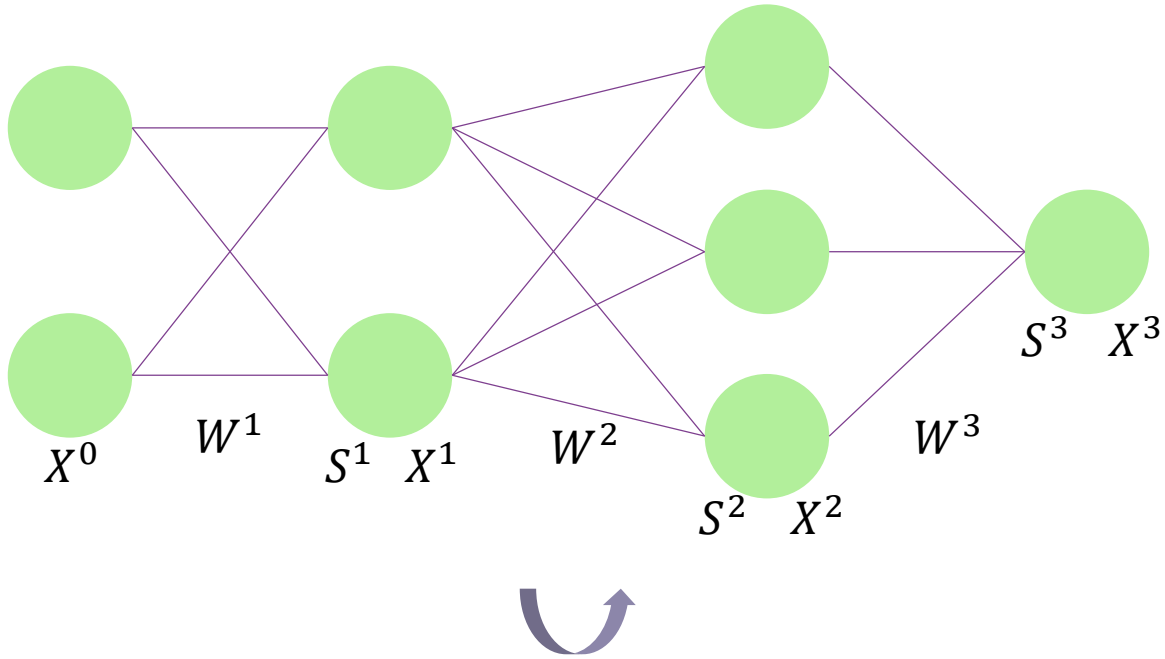
$$= [-1.99979229 \quad 1.00011585 \quad 2.99982311]$$

$$W^{(2)} = W^{(2)} - \eta * \delta^{(2)} * X^{T^{(1)}} = \begin{bmatrix} 0.0982129 & -0.0982129 \\ 0.2008938 & 0.09910618 \\ 0.3026806 & 0.49731935 \end{bmatrix} - 1.0 * \begin{bmatrix} -0.06427884 \\ 0.03214689 \\ 0.0964523 \end{bmatrix} * [-0.11546102 \quad 0.06794028]$$

$$= \begin{bmatrix} 0.09784181 & -0.09799454 \\ 0.2010794 & 0.09899698 \\ 0.30323748 & 0.4969917 \end{bmatrix}$$

(2 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$

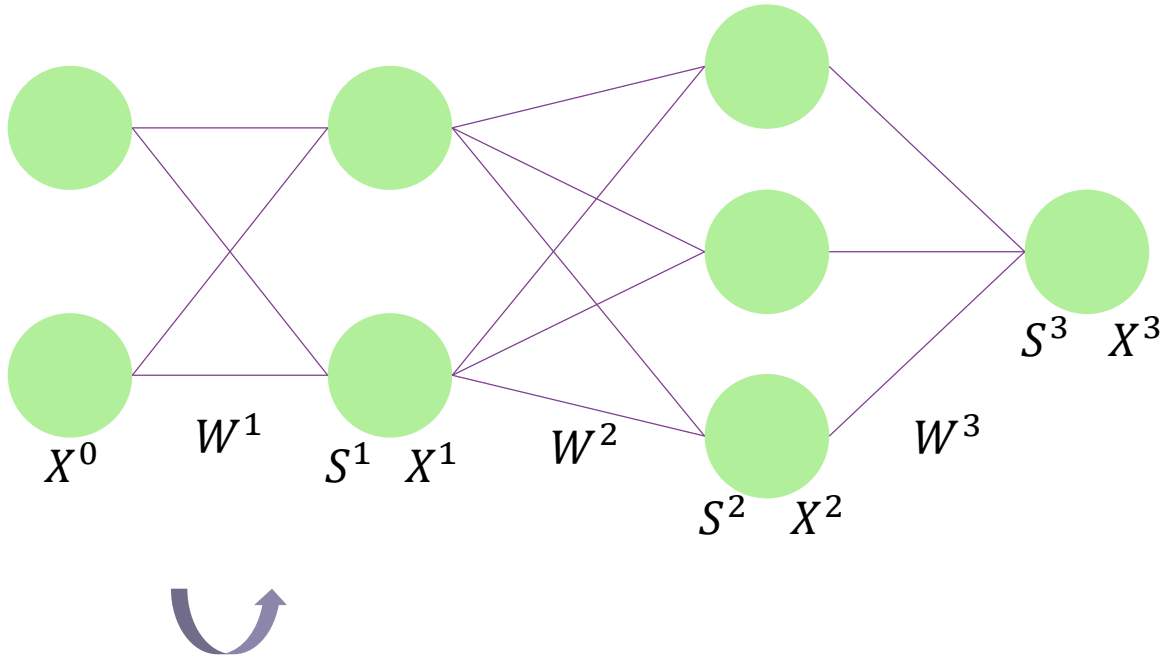


$$W^{(3)} = [-1.99979229 \quad 1.00011585 \quad 2.99982311]$$

$$W^{(2)} = \begin{bmatrix} 0.09784181 & -0.09799454 \\ 0.2010794 & 0.09899698 \\ 0.30323748 & 0.4969917 \end{bmatrix}$$

(2 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$



$$W^{(3)} = [-1.99979229 \quad 1.00011585 \quad 2.99982311]$$

$$W^{(2)} = \begin{bmatrix} 0.09784181 & -0.09799454 \\ 0.2010794 & 0.09899698 \\ 0.30323748 & 0.4969917 \end{bmatrix}$$

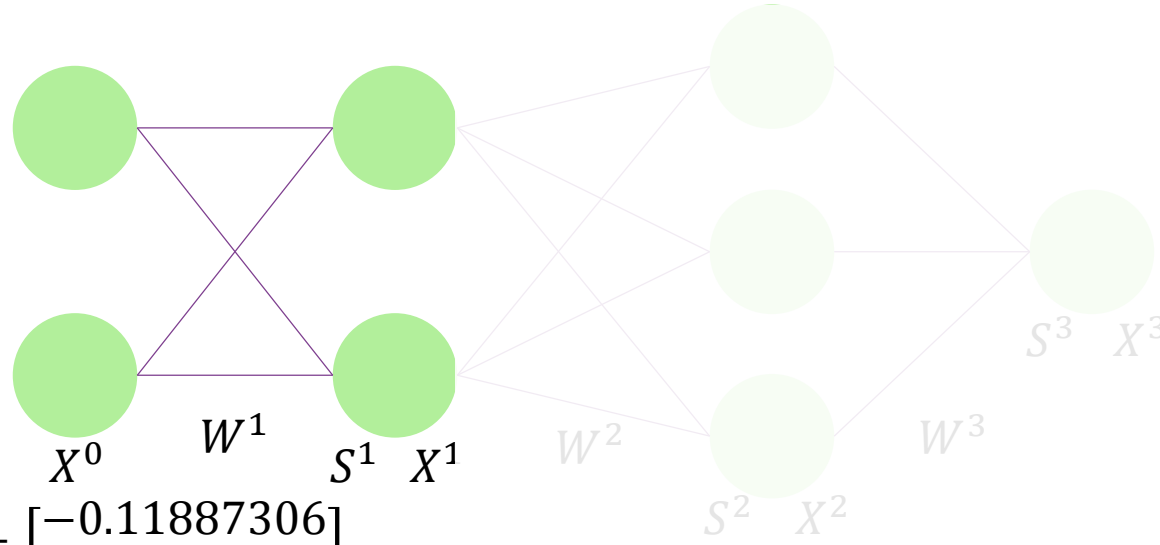
$$W^{(1)} = W^{(1)} - \eta * \delta^{(1)} * X^{T^{(0)}} = \begin{bmatrix} 0.09201088 & 0.20798912 \\ 0.88402255 & 0.81597745 \end{bmatrix} - 1.0 * \begin{bmatrix} 0.02894821 \\ 0.0572013 \end{bmatrix} * [1 \quad -1]$$

$$= \begin{bmatrix} 0.09056347 & 0.20943653 \\ 0.88116249 & 0.81883751 \end{bmatrix}$$

3rd Round

(3 round) - Forward

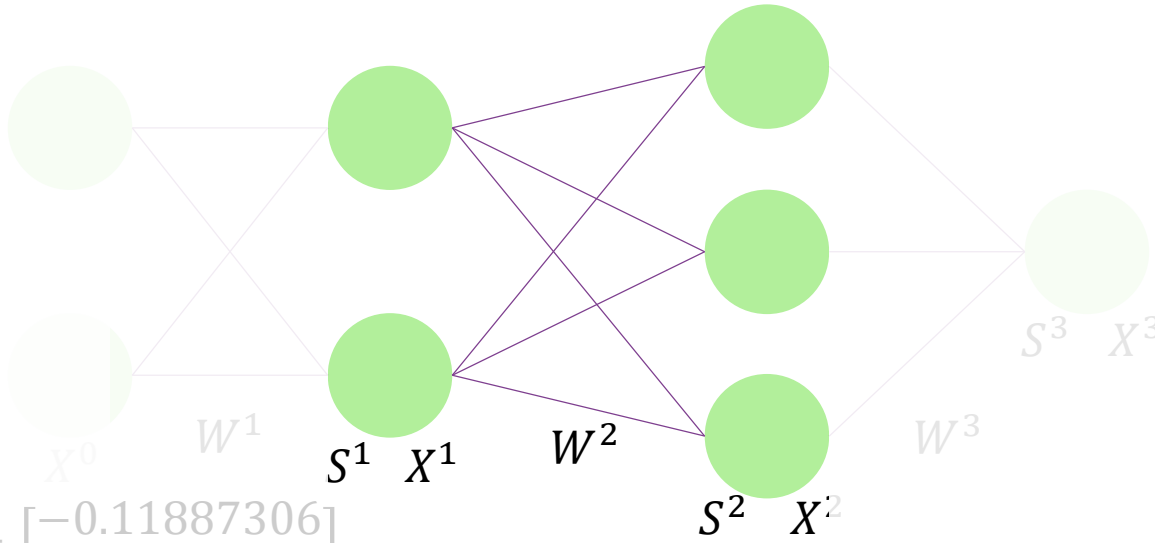
$$s^{(l)} = w^{(l)} * x^{(l-1)} \qquad \mathrm{x}^{(l)} = \tanh(s^{(l)})$$



$$S^{(1)} = \mathrm{W}^{(1)} * X^{(0)} = \begin{bmatrix} 0.09056347 & 0.20943653 \\ 0.88116249 & 0.81883751 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.11887306 \\ 0.06232498 \end{bmatrix}$$

$$X^{(1)} = \tanh\left(S^{(1)}\right) = \tanh\left(\begin{bmatrix} -0.11887306 \\ 0.06232498 \end{bmatrix}\right) = \begin{bmatrix} -0.11831628 \\ 0.06224441 \end{bmatrix}$$

(3 round) - Forward



$$s^{(l)} = w^{(l)} * x^{(l-1)} \qquad \mathrm{x}^{(l)} = \tanh(s^{(l)})$$

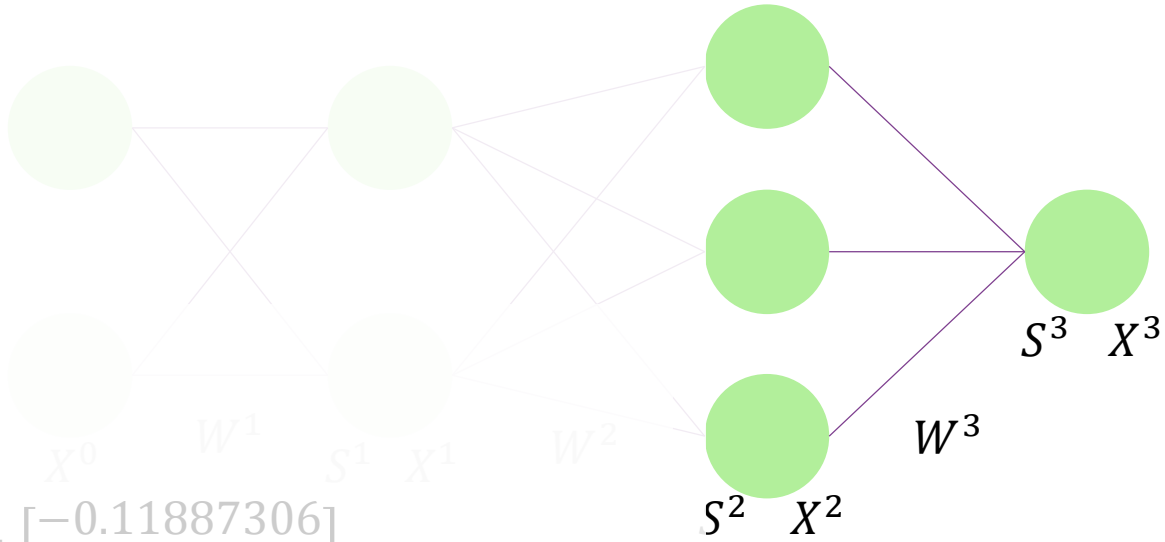
$$S^{(1)} = \mathrm{W}^{(1)} * X^{(0)} = \begin{bmatrix} 0.09056347 & 0.20943653 \\ 0.88116249 & 0.81883751 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.11887306 \\ 0.06232498 \end{bmatrix}$$

$$X^{(1)} = \tanh\left(S^{(1)}\right) = \tanh\left(\begin{bmatrix} -0.11887306 \\ 0.06232498 \end{bmatrix}\right) = \begin{bmatrix} -0.11831628 \\ 0.06224441 \end{bmatrix}$$

$$S^{(2)} = \mathrm{W}^{(2)} * X^{(1)} = \begin{bmatrix} 0.09784181 & -0.09799454 \\ 0.2010794 & 0.09899698 \\ 0.30323748 & 0.4969917 \end{bmatrix} \begin{bmatrix} -0.11831628 \\ 0.06224441 \end{bmatrix} = \begin{bmatrix} -0.01767589 \\ -0.01762896 \\ -0.00494298 \end{bmatrix}$$

$$X^{(2)} = \tanh\left(S^{(2)}\right) = \tanh\left(\begin{bmatrix} -0.01767589 \\ -0.01762896 \\ -0.00494298 \end{bmatrix}\right) = \begin{bmatrix} -0.01767405 \\ -0.01762713 \\ -0.00494294 \end{bmatrix}$$

(3 round) - Forward



$$s^{(l)} = w^{(l)} * x^{(l-1)} \qquad \mathrm{x}^{(l)} = \tanh(s^{(l)})$$

$$S^{(1)} = \mathrm{W}^{(1)} * X^{(0)} = \begin{bmatrix} 0.09056347 & 0.20943653 \\ 0.88116249 & 0.81883751 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -0.11887306 \\ 0.06232498 \end{bmatrix}$$

$$X^{(1)} = \tanh\left(S^{(1)}\right) = \tanh\left(\begin{bmatrix} -0.11887306 \\ 0.06232498 \end{bmatrix}\right) = \begin{bmatrix} -0.11831628 \\ 0.06224441 \end{bmatrix}$$

$$S^{(2)} = \mathrm{W}^{(2)} * X^{(1)} = \begin{bmatrix} 0.09784181 & -0.09799454 \\ 0.2010794 & 0.09899698 \\ 0.30323748 & 0.4969917 \end{bmatrix} \begin{bmatrix} -0.11831628 \\ 0.06224441 \end{bmatrix} = \begin{bmatrix} -0.01767589 \\ -0.01762896 \\ -0.00494298 \end{bmatrix}$$

$$X^{(2)} = \tanh\left(S^{(2)}\right) = \tanh\left(\begin{bmatrix} -0.01767589 \\ -0.01762896 \\ -0.00494298 \end{bmatrix}\right) = \begin{bmatrix} -0.01767405 \\ -0.01762713 \\ -0.00494294 \end{bmatrix}$$

$$S^{(3)} = \mathrm{W}^{(3)} * X^{(2)} = [-1.99979229 \quad 1.00011585 \quad 2.99982311] \begin{bmatrix} -0.01767405 \\ -0.01762713 \\ -0.00494294 \end{bmatrix} = [0.00288732]$$

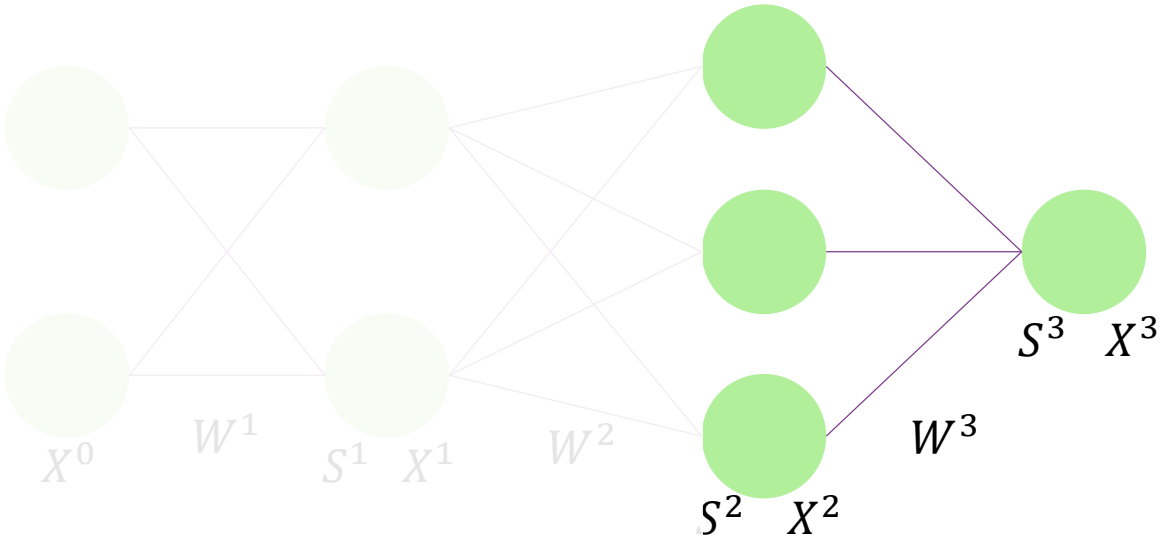
$$X^{(3)} = \mathrm{linear}\left(S^{(3)}\right) = [0.00288732]$$

(3 round) - Backward

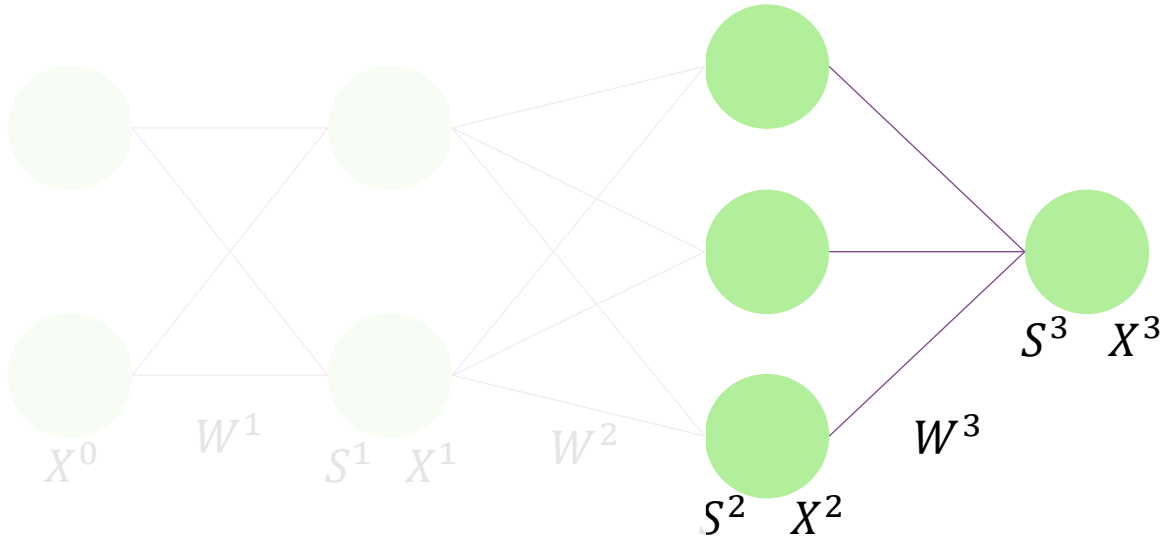
$$\delta^{(L)} = -2(y_n - s^{(L)})$$

$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \operatorname{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \operatorname{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.00288732] \qquad y_n = 0$$



(3 round) - Backward



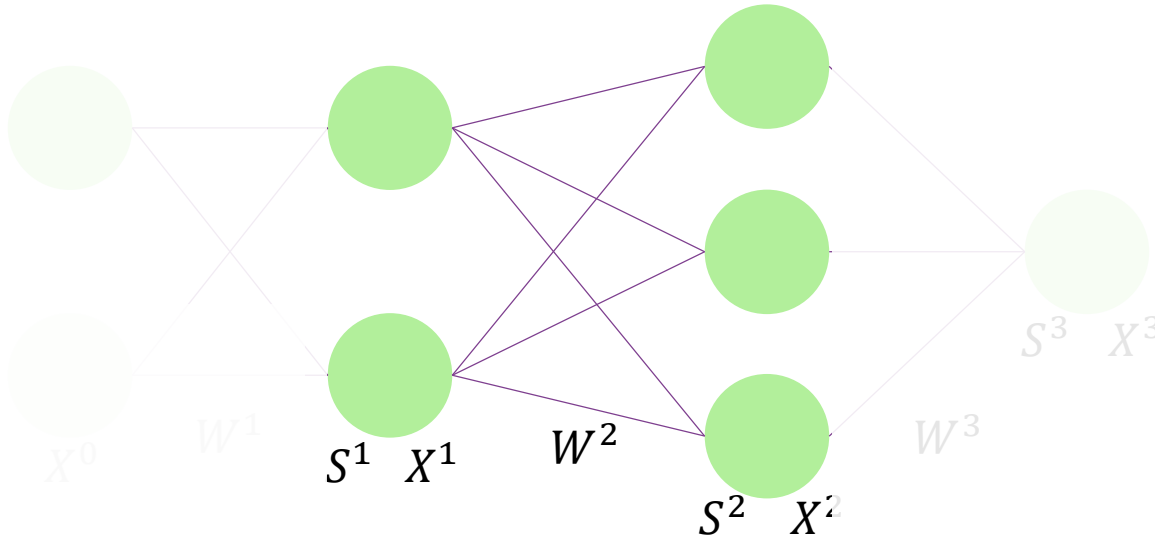
$$\delta^{(L)} = -2(y_n - s^{(L)})$$

$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \operatorname{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \operatorname{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.00288732] \qquad y_n = 0$$

$$\delta^{(3)} = -2(y_n - s^{(3)}) = -2 * (0 - [0.00288732]) = 0.00577464$$

(3 round) - Backward



$$\delta^{(L)} = -2(y_n - s^{(L)})$$

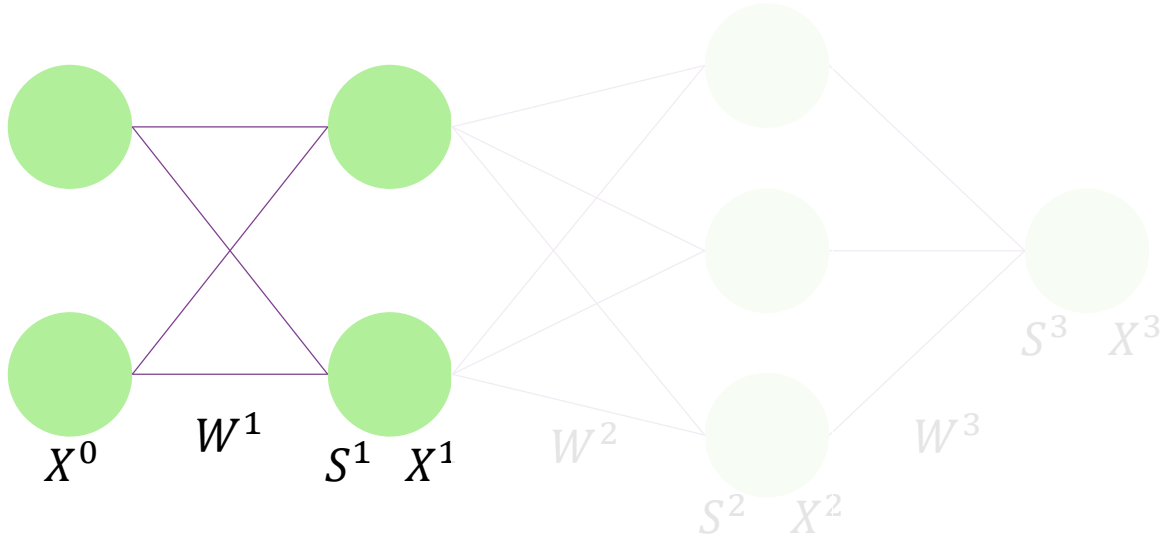
$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \text{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \text{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.00288732] \qquad y_n = 0$$

$$\delta^{(3)} = -2(y_n - s^{(3)}) = -2 * (0 - [0.00288732]) = 0.00577464$$

$$\delta^{(2)} = W^{T(3)} * \delta^{(3)} \otimes \text{sech}^2(s^{(2)}) = \begin{bmatrix} -1.99979229 \\ 1.00011585 \\ 2.99982311 \end{bmatrix} * [0.00577464] \otimes \text{sech}^2\left(\begin{bmatrix} -0.01767589 \\ -0.01762896 \\ -0.00494298 \end{bmatrix}\right) = \begin{bmatrix} -0.01154447 \\ 0.00577351 \\ 0.01732246 \end{bmatrix}$$

(3 round) - Backward



$$\delta^{(L)} = -2(y_n - s^{(L)})$$

$$\delta^{(l)} = \delta^{(l+1)} * W^{(l+1)} * \operatorname{sech}^2(s^{(l)})$$
$$= W^{T(l+1)} * \delta^{(l+1)} \otimes \operatorname{sech}^2(s^{(l)})$$

$$S^{(3)} = [0.00288732] \qquad y_n = 0$$

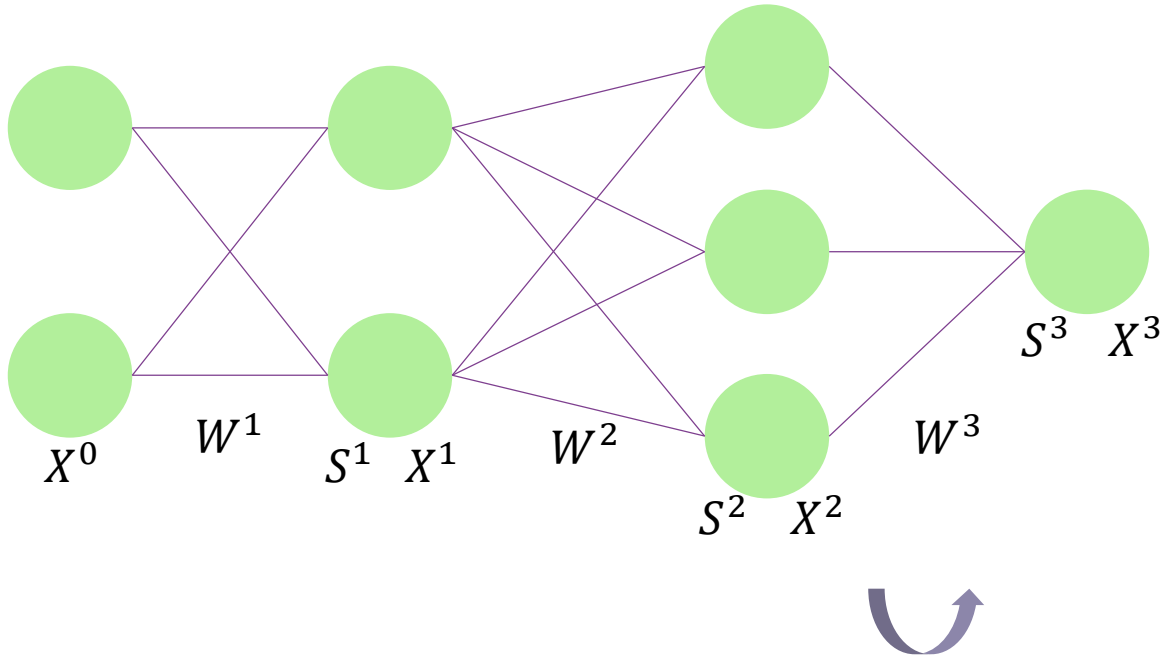
$$\delta^{(3)} = -2(y_n - s^{(3)}) = -2 * (0 - [0.00288732]) = 0.00577464$$

$$\delta^{(2)} = W^{T(3)} * \delta^{(3)} \otimes \operatorname{sech}^2(s^{(2)}) = \begin{bmatrix} -1.99979229 \\ 1.00011585 \\ 2.99982311 \end{bmatrix} * [0.00577464] \otimes \operatorname{sech}^2\left(\begin{bmatrix} -0.01767589 \\ -0.01762896 \\ -0.00494298 \end{bmatrix}\right) = \begin{bmatrix} -0.01154447 \\ 0.00577351 \\ 0.01732246 \end{bmatrix}$$

$$\delta^{(1)} = W^{T(2)} * \delta^{(2)} \otimes \operatorname{sech}^2(s^{(1)}) = \begin{bmatrix} 0.09784181 & 0.2010794 & 0.30323748 \\ -0.09799454 & 0.09899698 & 0.4969917 \end{bmatrix} * \begin{bmatrix} -0.01154447 \\ 0.00577351 \\ 0.01732246 \end{bmatrix}$$
$$\otimes \operatorname{sech}^2\left(\begin{bmatrix} -0.11887306 \\ 0.06232498 \end{bmatrix}\right) = \begin{bmatrix} 0.00521025 \\ 0.01027202 \end{bmatrix}$$

(3 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$

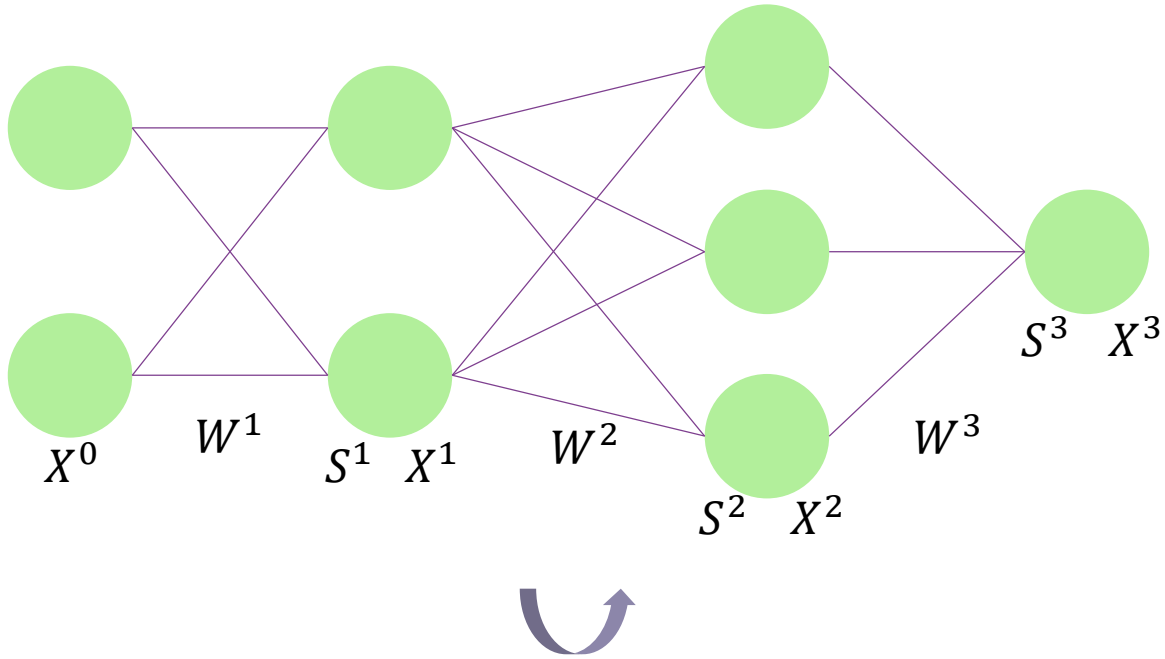


$$W^{(3)} = W^{(3)} - \eta * \delta^{(3)} * X^{T^{(2)}} = [-1.99979229 \quad 1.00011585 \quad 2.99982311] - 1.0 * 0.00577464 * [-0.01767405 \quad -0.01762713 \quad 0.00494294]$$

$$= [-1.99978719 \quad 1.00012094 \quad 2.99982453]$$

(3 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$



$$W^{(3)} = W^{(3)} - \eta * \delta^{(3)} * X^{T^{(2)}} = [-1.99979229 \quad 1.00011585 \quad 2.99982311] - 1.0 * 0.00577464 * [-0.01767405 \quad -0.01762713 \quad 0.00494294]$$

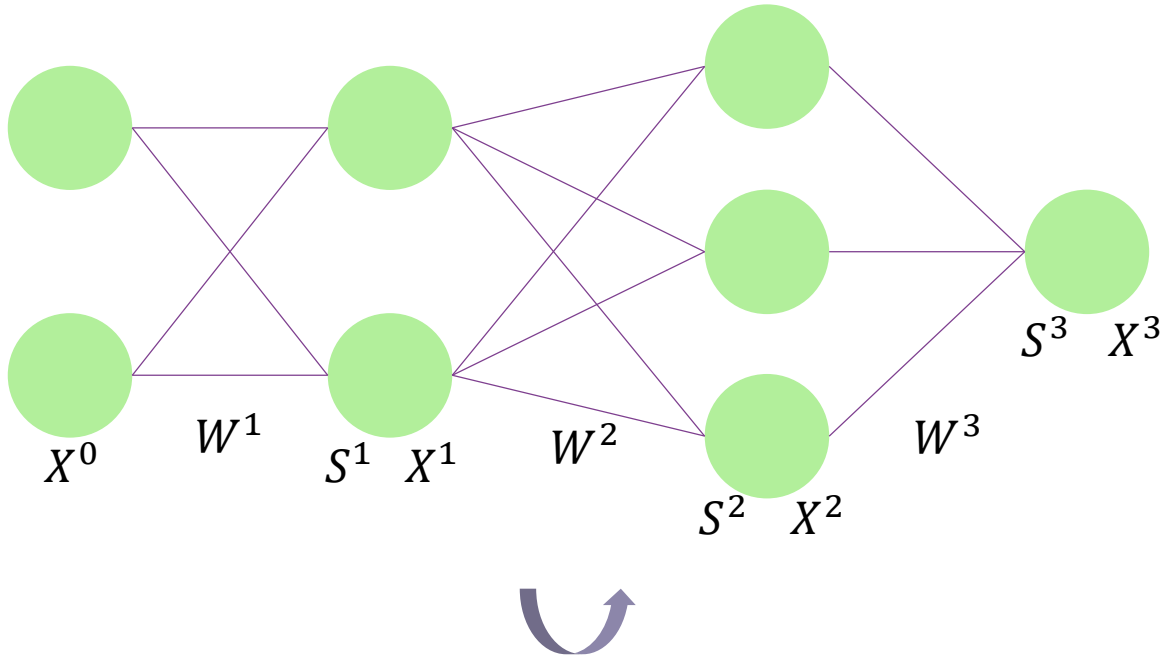
$$= [-1.99978719 \quad 1.00012094 \quad 2.99982453]$$

$$W^{(2)} = W^{(2)} - \eta * \delta^{(2)} * X^{T^{(1)}} = \begin{bmatrix} 0.09784181 & -0.09799454 \\ 0.2010794 & 0.09899698 \\ 0.30323748 & 0.4969917 \end{bmatrix} - 1.0 * \begin{bmatrix} -0.01154447 \\ 0.00577351 \\ 0.01732246 \end{bmatrix} * [-0.11831628 \quad 0.06224441]$$

$$= \begin{bmatrix} 0.09777352 & -0.09795861 \\ 0.20111356 & 0.09897901 \\ 0.30333995 & 0.49693779 \end{bmatrix}$$

(3 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$

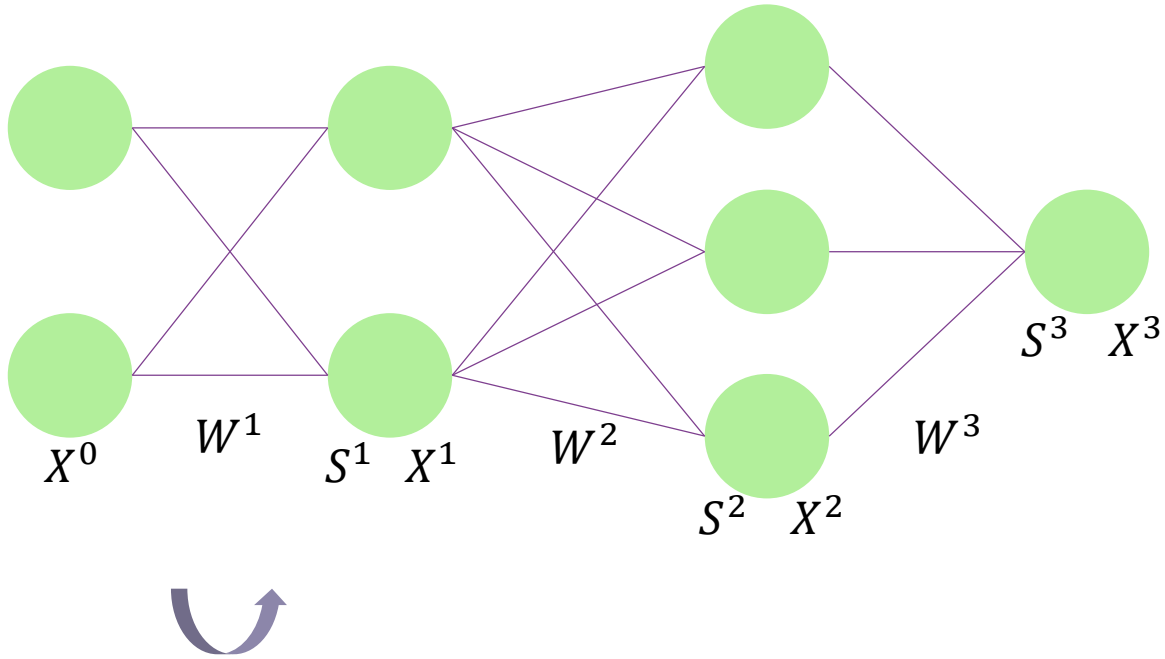


$$W^{(3)} = [-1.99978719 \quad 1.00012094 \quad 2.99982453]$$

$$W^{(2)} = \begin{bmatrix} 0.09777352 & -0.09795861 \\ 0.20111356 & 0.09897901 \\ 0.30333995 & 0.49693779 \end{bmatrix}$$

(3 round) - Update

$$W^{(l)} = W^{(l)} - \eta * X^{(l-1)} * \delta^{(l)}$$
$$\Rightarrow W^{(l)} - \eta * \delta^{(l)} * X^{T^{(l-1)}}$$



$$W^{(3)} = [-1.99978719 \quad 1.00012094 \quad 2.99982453]$$

$$W^{(2)} = \begin{bmatrix} 0.09777352 & -0.09795861 \\ 0.20111356 & 0.09897901 \\ 0.30333995 & 0.49693779 \end{bmatrix}$$

$$W^{(1)} = W^{(1)} - \eta * \delta^{(1)} * X^{T^{(0)}} = \begin{bmatrix} 0.09056347 & 0.20943653 \\ 0.88116249 & 0.81883751 \end{bmatrix} - 1.0 * \begin{bmatrix} 0.00521025 \\ 0.01027202 \end{bmatrix} * [1 \quad -1]$$

$$= \begin{bmatrix} 0.09030296 & 0.20969704 \\ 0.88064889 & 0.81935111 \end{bmatrix}$$

Error comparison

Error measure

Round1: $S^{(3)} = [0.08968833]$ $err = (0 - 0.08968833)^2 = 0.00804399654$

Round2: $S^{(3)} = [0.01607636]$ $err = (0 - 0.01607636)^2 = 0.000258449351$

Round3: $S^{(3)} = [0.00288732]$ $err = (0 - 0.00288732)^2 = 0.000008336616$

Decay!