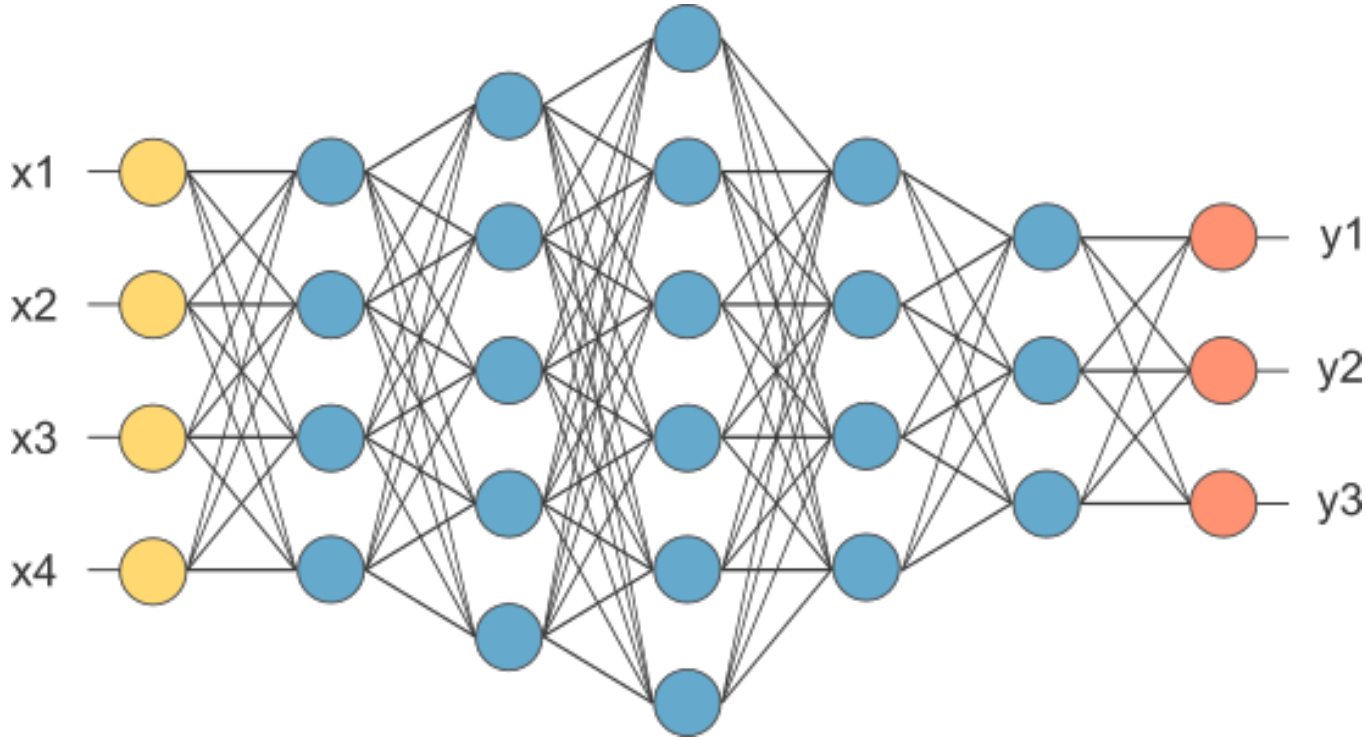
MAC5921 – Deep Learning

Aula 03 – 17/08/2023

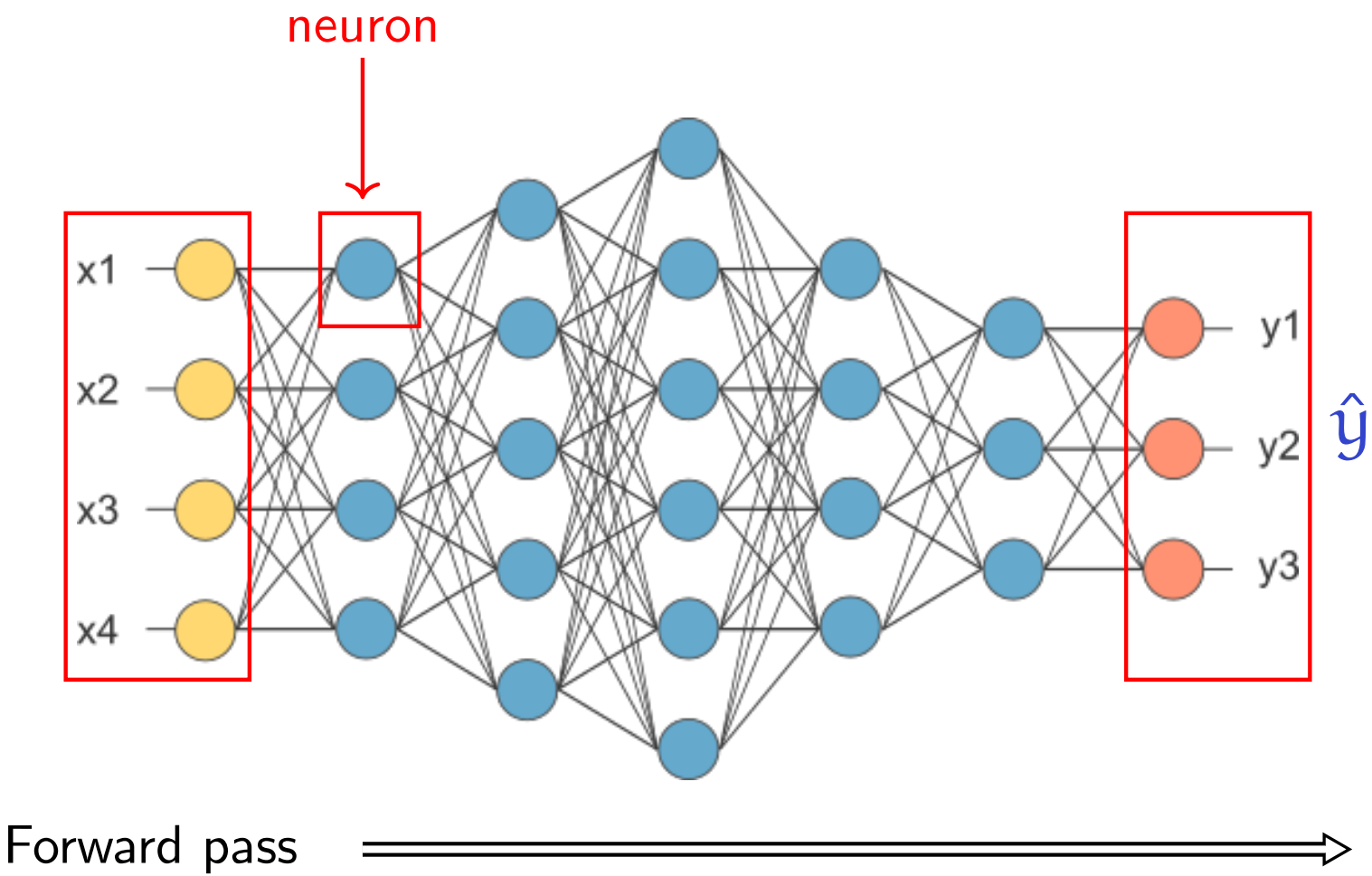
Nina S. T. Hirata

Rede neural fully connected



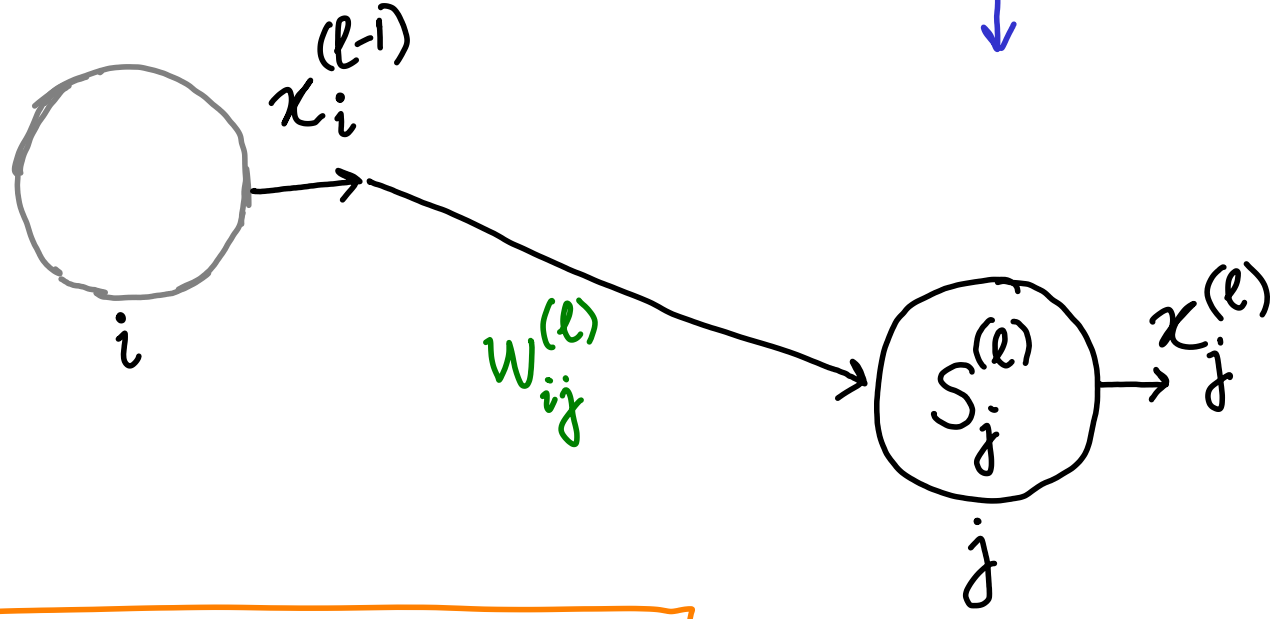
https://medium.com/ravenprotocol/everything-you-need-to-know-about-neural-networks-6fcc7a15cb4

Forward pass



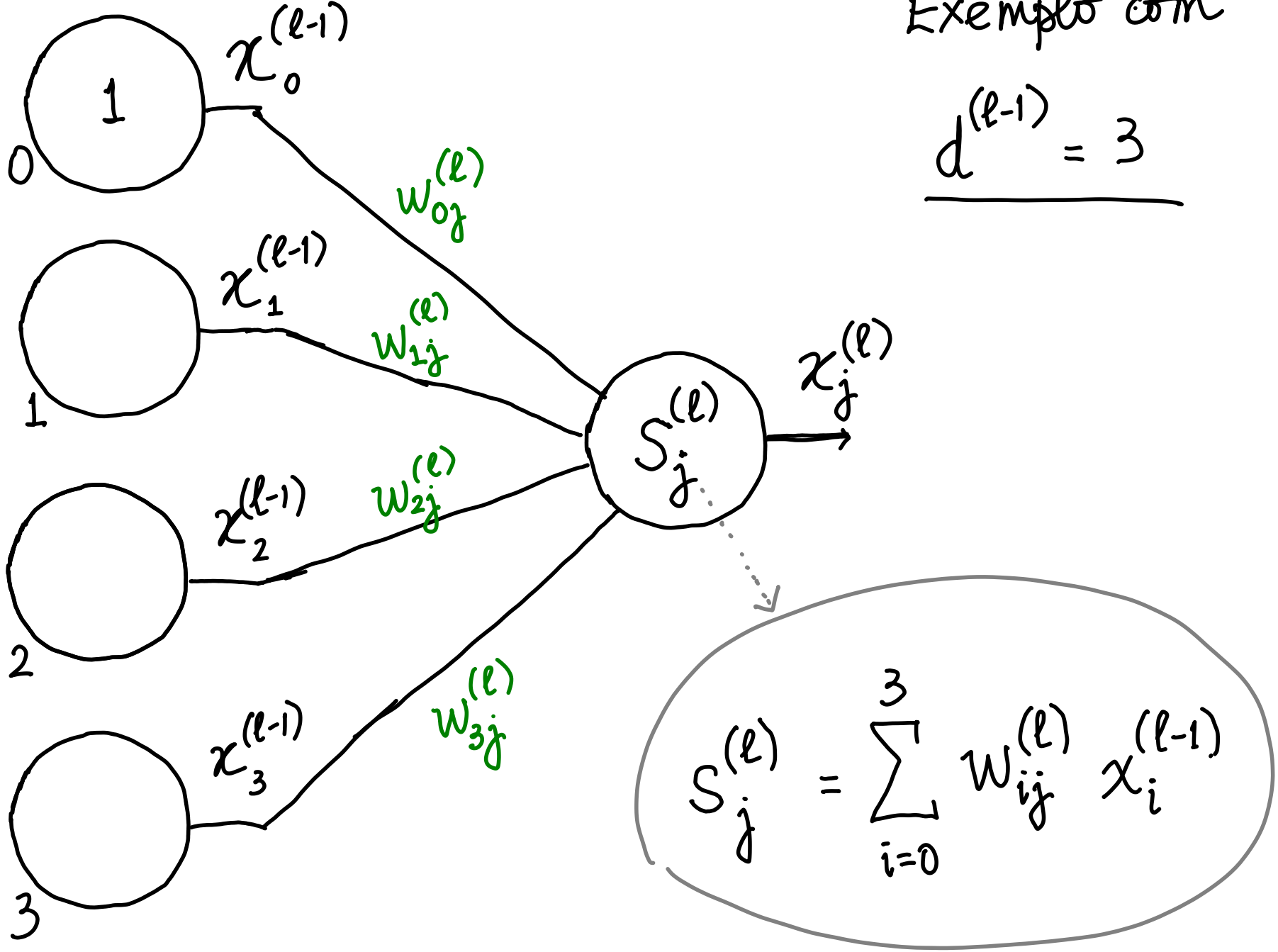
Camada $\ell-1$

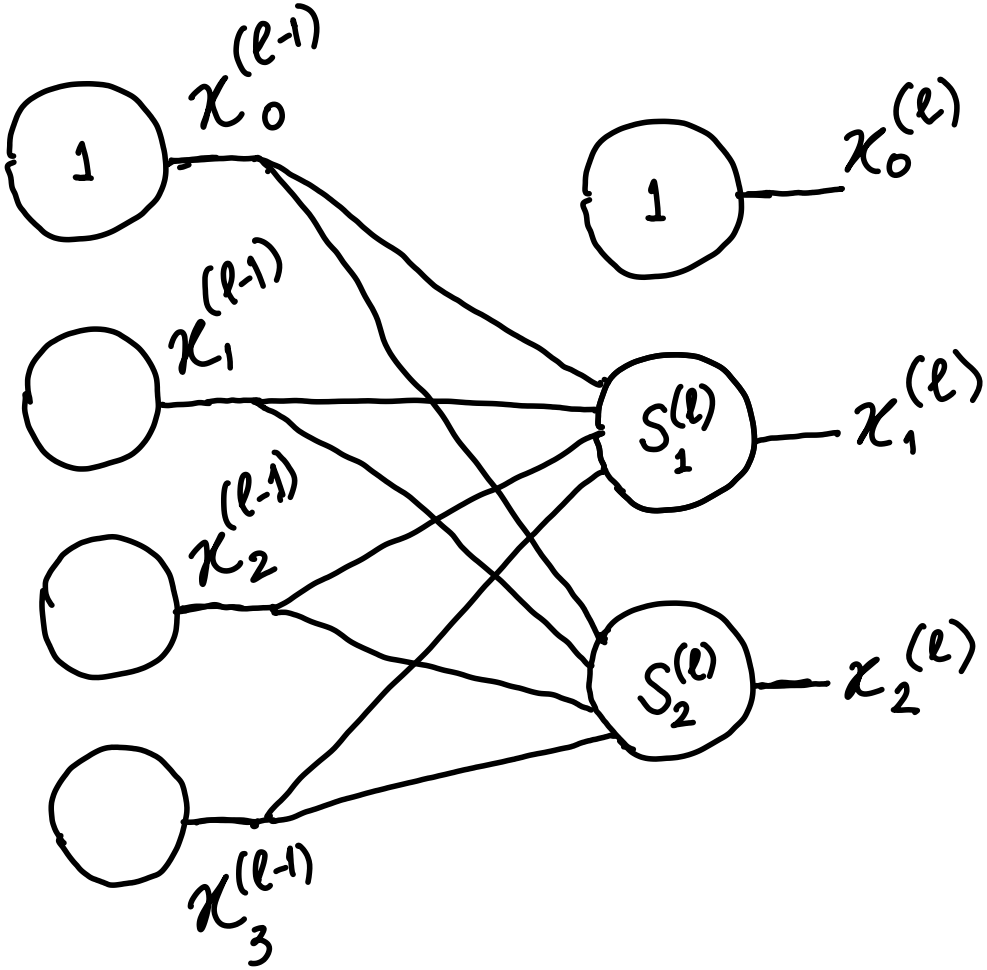
Camada ℓ

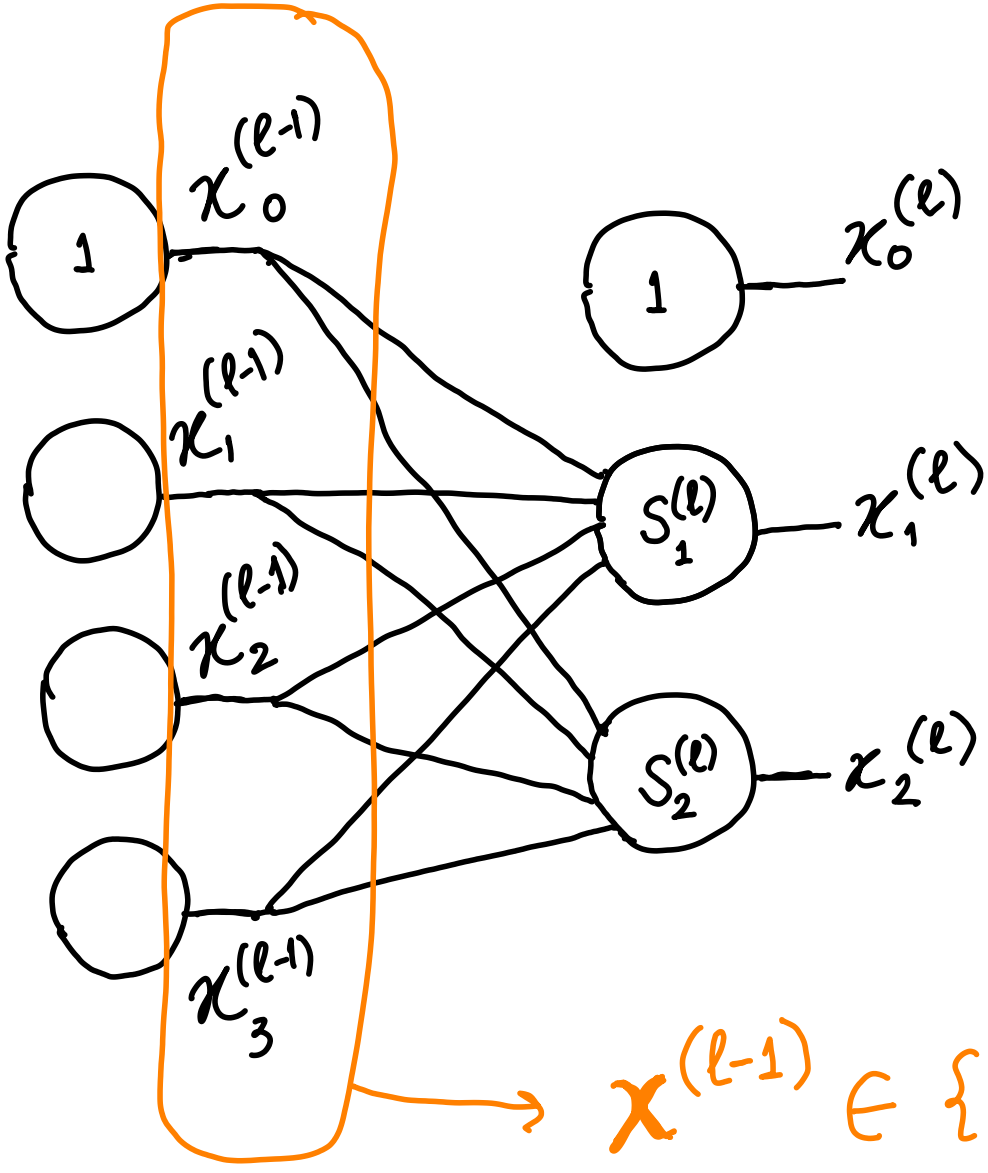


$$s_j^{(\ell)} = \sum_{i=0}^{d^{(\ell-1)}} w_{ij}^{(\ell)} \, x_i^{(\ell-1)}$$

$$x_j^{(\ell)} = \theta\left(s_j^{(\ell)}\right)$$


Exemplo com
$d^{(\ell-1)} = 3$
1
0
$x_0^{(\ell-1)}$
$w_{0j}^{(\ell)}$
$x_1^{(\ell-1)}$
1
$w_{1j}^{(\ell)}$
$x_2^{(\ell-1)}$
2
$w_{2j}^{(\ell)}$
$x_3^{(\ell-1)}$
3
$w_{3j}^{(\ell)}$
$s_j^{(\ell)}$
$x_j^{(\ell)}$
$s_j^{(\ell)} = \sum_{i=0}^{3} w_{ij}^{(\ell)} \, x_i^{(\ell-1)}$


1
$x_0^{(l-1)}$
$x_1^{(l-1)}$
$x_2^{(l-1)}$
$x_3^{(l-1)}$
1
$x_0^{(l)}$
$s_1^{(l)}$
$x_1^{(l)}$
$s_2^{(l)}$
$x_2^{(l)}$

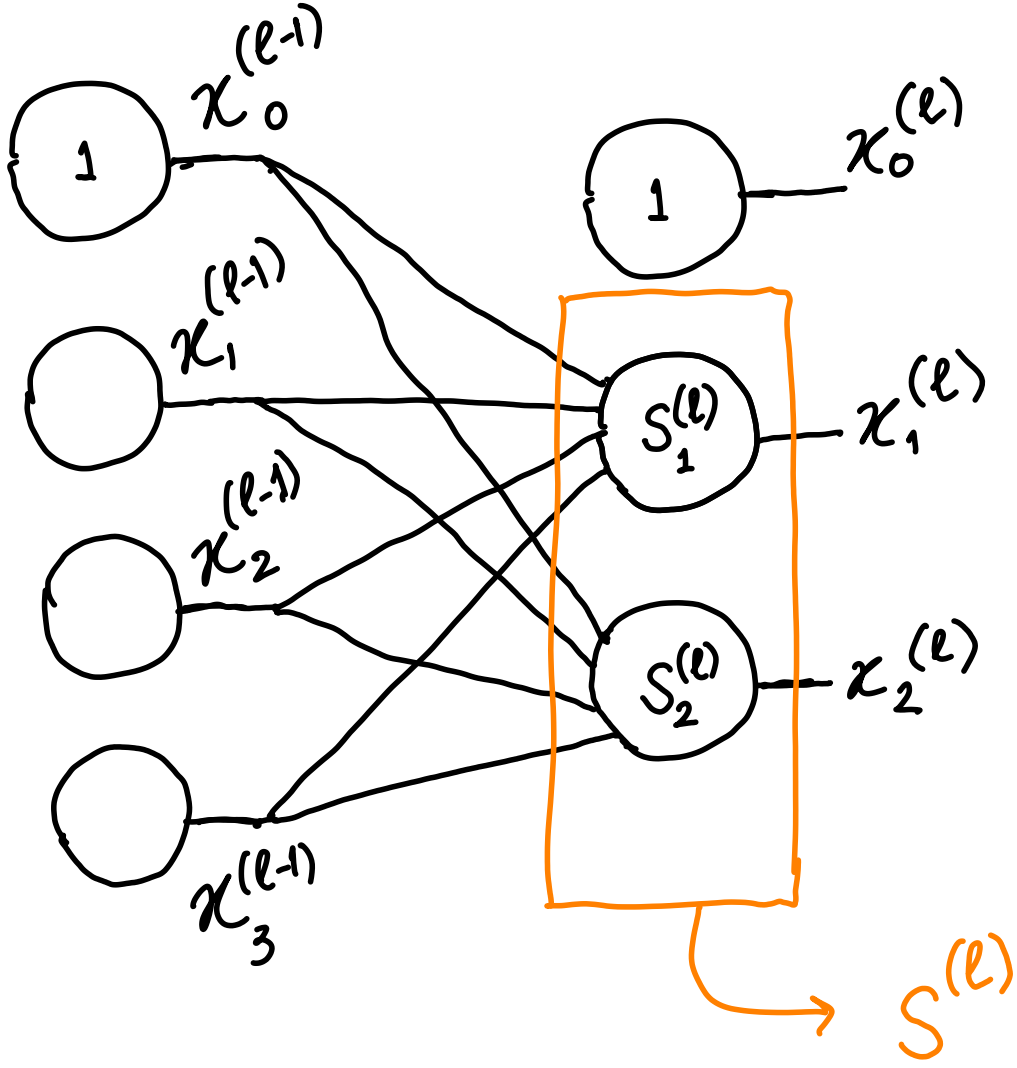


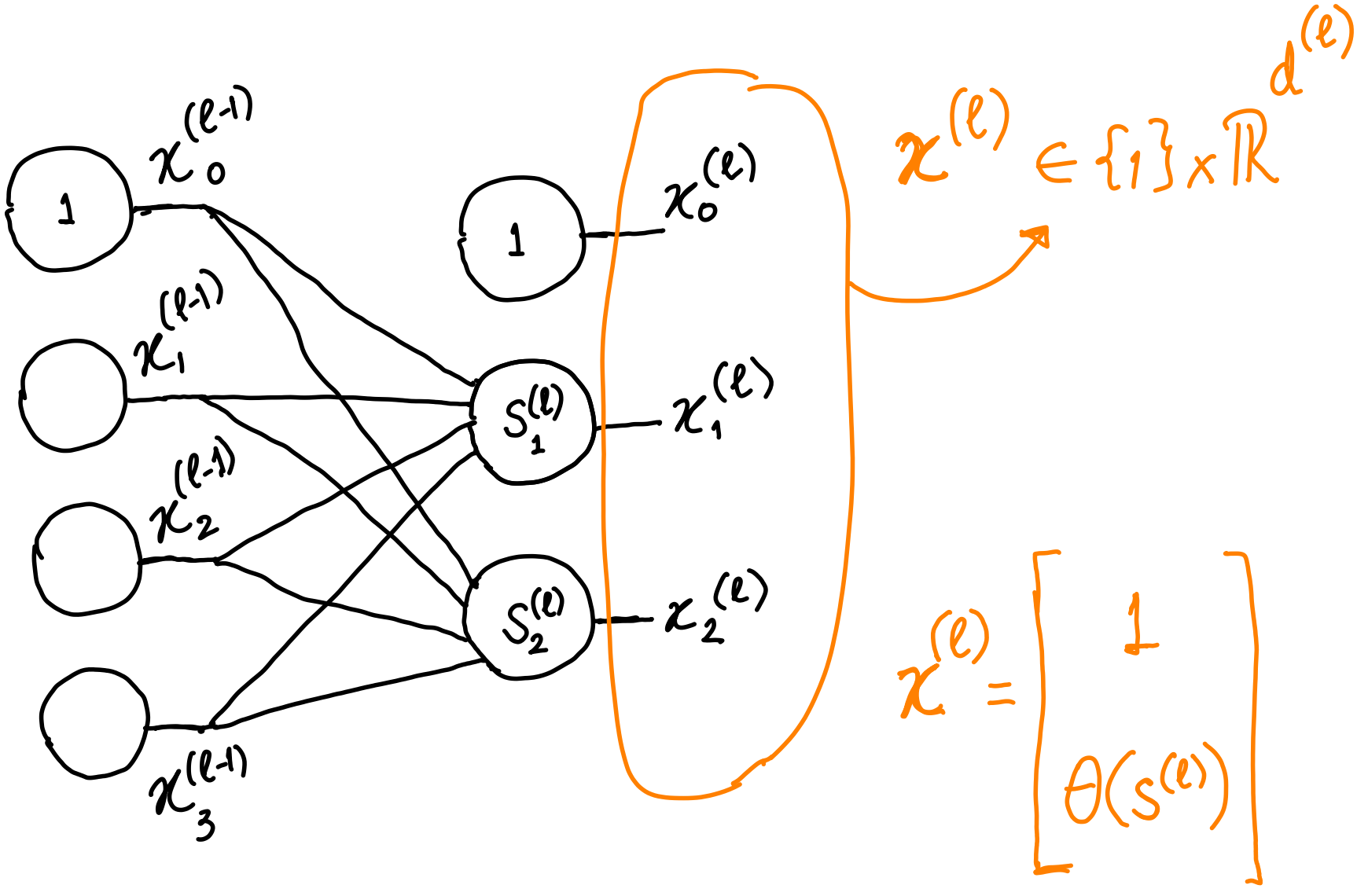
Exemplo com

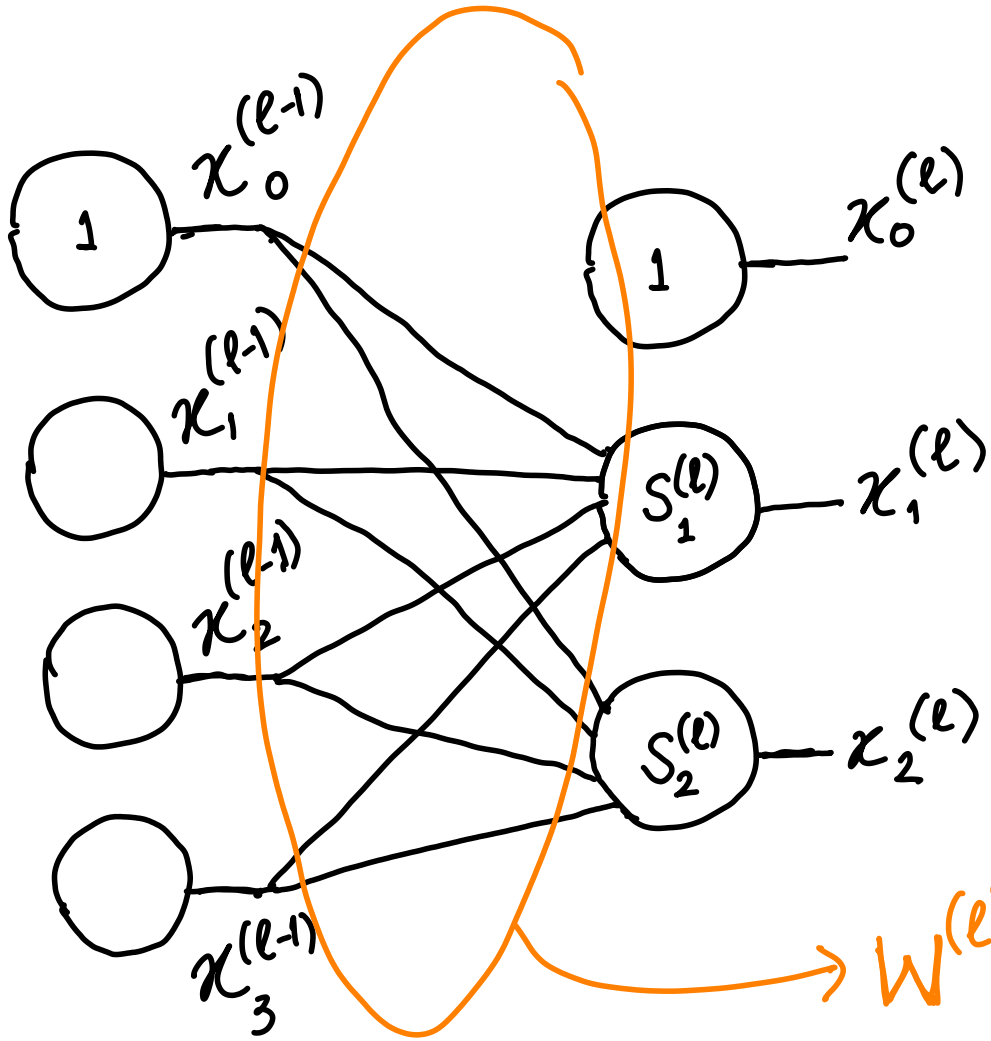
$d^{(\ell-1)} = 3$

$d^{(\ell)} = 2$

$$\mathbf{x}^{(\ell-1)} \in \{1\} \times \mathbb{R}^{d^{(\ell-1)}}$$

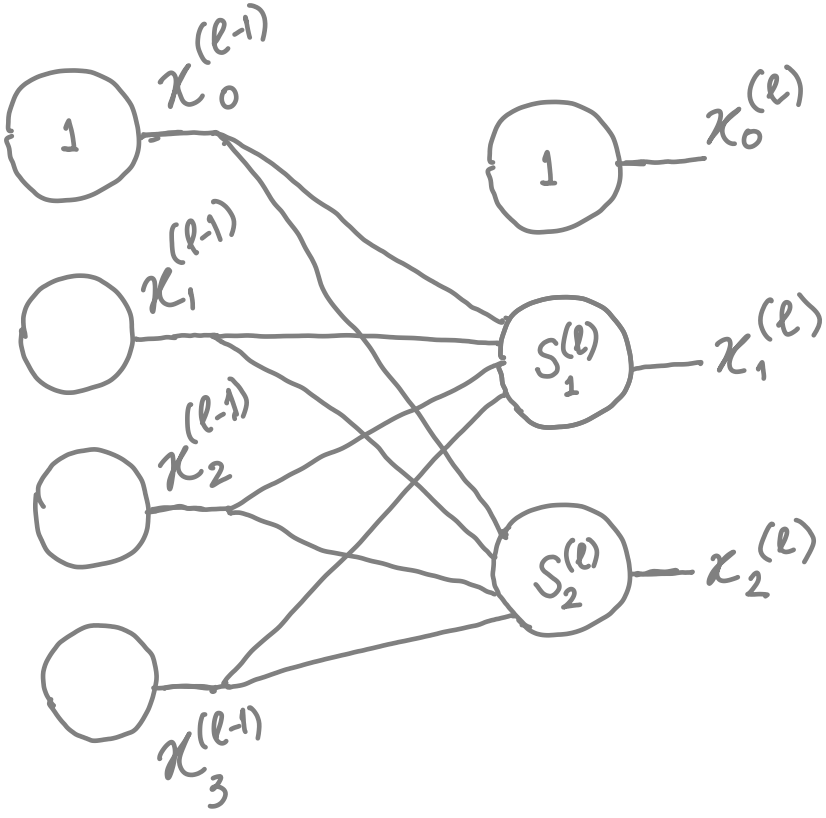

1
$x_0^{(\ell-1)}$
$x_1^{(\ell-1)}$
$x_2^{(\ell-1)}$
$x_3^{(\ell-1)}$
1
$x_0^{(\ell)}$
$s_1^{(\ell)}$
$x_1^{(\ell)}$
$s_2^{(\ell)}$
$x_2^{(\ell)}$
$s^{(\ell)}$


1
$x_0^{(\ell-1)}$
$x_1^{(\ell-1)}$
$x_2^{(\ell-1)}$
$x_3^{(\ell-1)}$
1
$s_1^{(\ell)}$
$s_2^{(\ell)}$
$x_0^{(\ell)}$
$x_1^{(\ell)}$
$x_2^{(\ell)}$
$x^{(\ell)} \in \{1\} \times \mathbb{R}^{d^{(\ell)}}$
$x^{(\ell)} = \begin{bmatrix} 1 \\ \theta(s^{(\ell)}) \end{bmatrix}$



matriz de pesos

$(d^{(\ell-1)} + 1) \times d^{(\ell)}$



$$s^{(\ell)} = \left(W^{(\ell)}\right)^T x^{(\ell-1)}$$

$$x^{(\ell)} = \begin{bmatrix} 1 \\ \theta(s^{(\ell)}) \end{bmatrix}$$

$$x = x^{(0)} \xrightarrow{W^{(1)}} s^{(1)} \xrightarrow{\theta} x^{(1)} \rightarrow \cdots \xrightarrow{W^{(L)}} s^{(L)} \xrightarrow{\theta} x^{(L)} = h(x) = \hat{y}$$

Forward propagation to compute $h(\mathbf{x})$:

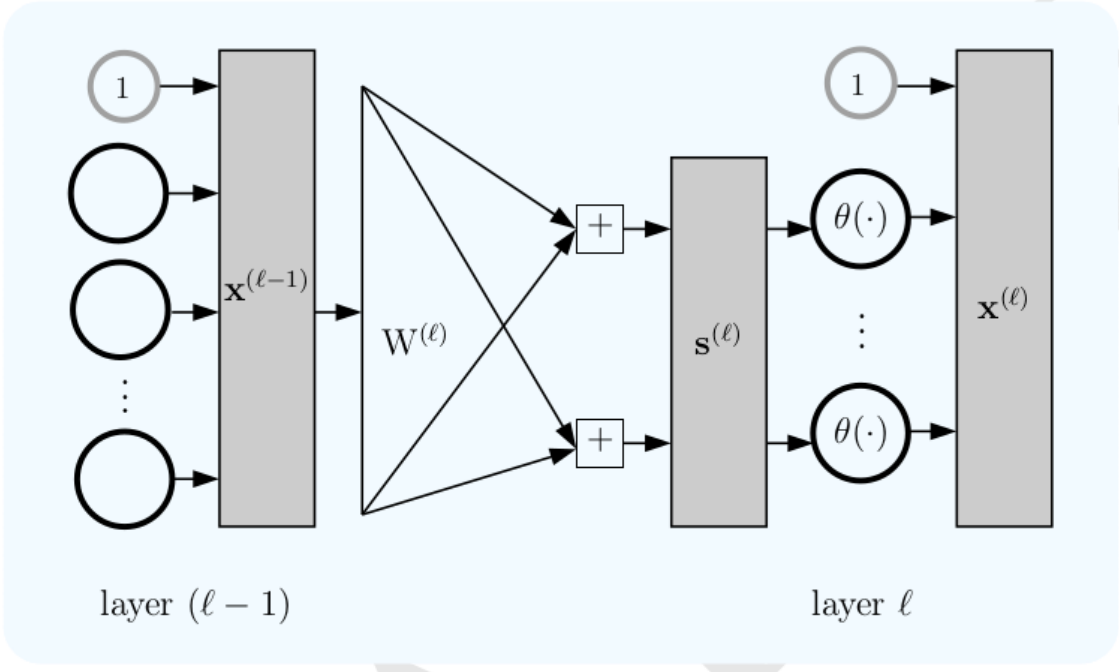
1: $\mathbf{x}^{(0)} \leftarrow \mathbf{x}$ **[Initialization]**

2: **for** $\ell = 1$ to L **do** **[Forward Propagation]**

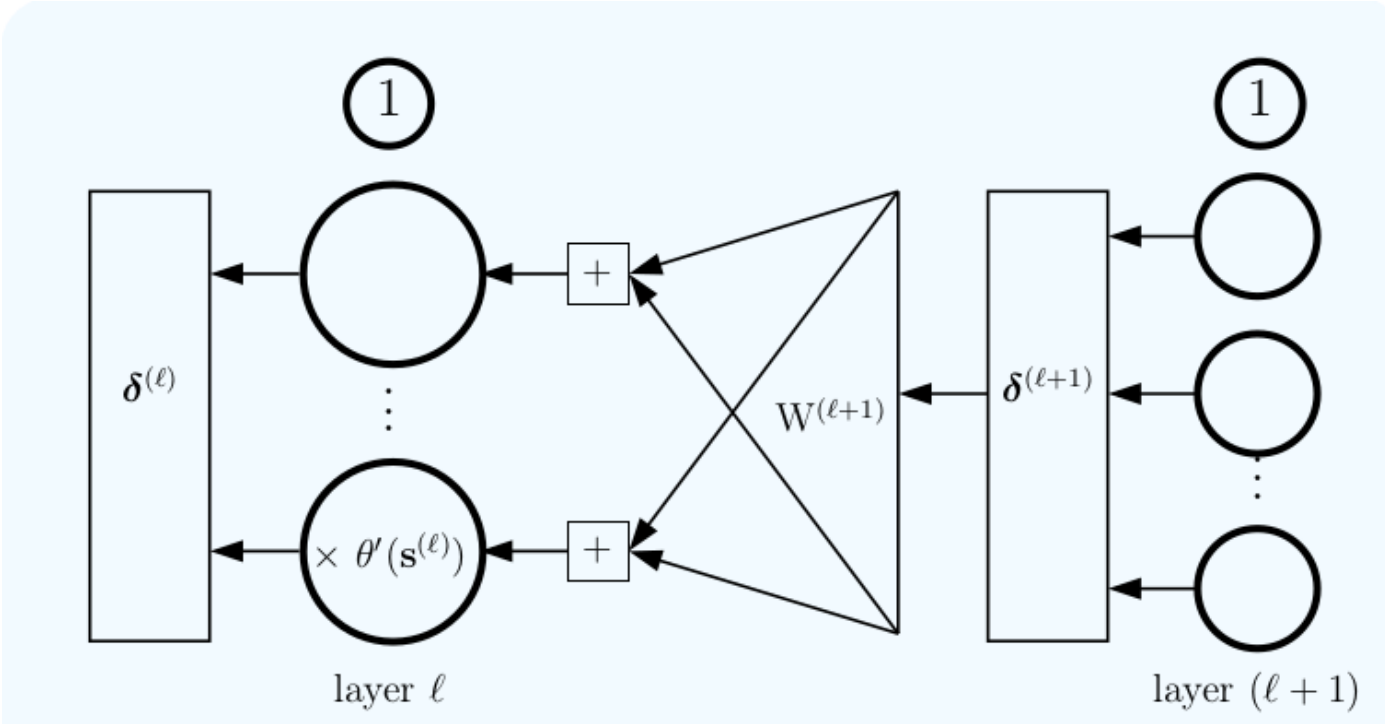
3: $\quad \mathbf{s}^{(\ell)} \leftarrow (\mathrm{W}^{(\ell)})^{\mathrm{T}} \mathbf{x}^{(\ell-1)}$

4: $\quad \mathbf{x}^{(\ell)} \leftarrow \begin{bmatrix} 1 \\ \theta(\mathbf{s}^{(\ell)}) \end{bmatrix}$

5: $h(\mathbf{x}) = \mathbf{x}^{(L)}$ **[Output]**



Backpropagation


1
1
$\boldsymbol{\delta}^{(\ell)}$
$\mathrm{W}^{(\ell+1)}$
$\boldsymbol{\delta}^{(\ell+1)}$
+
+
$\times\ \theta'(\mathbf{s}^{(\ell)})$
layer ℓ
layer $(\ell+1)$