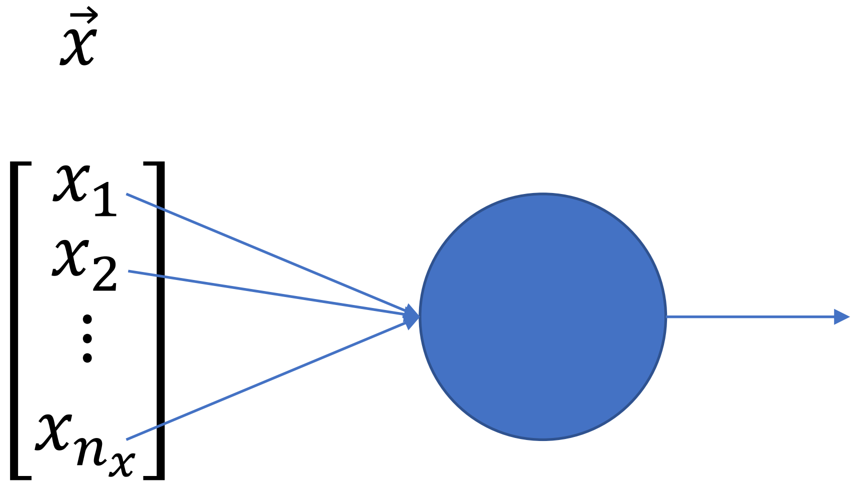
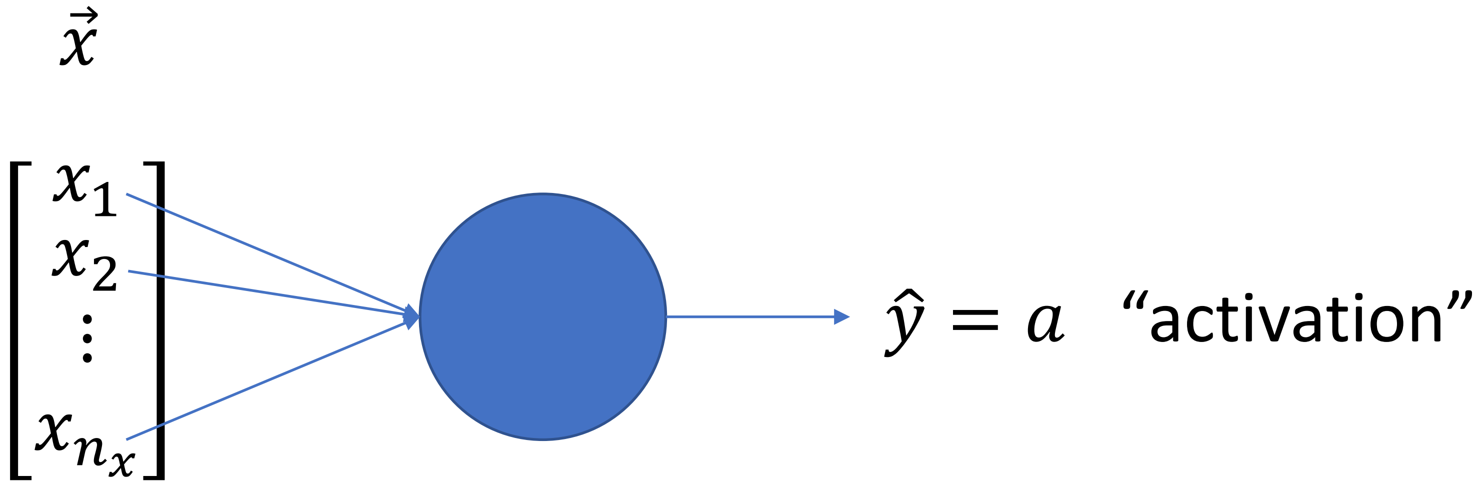
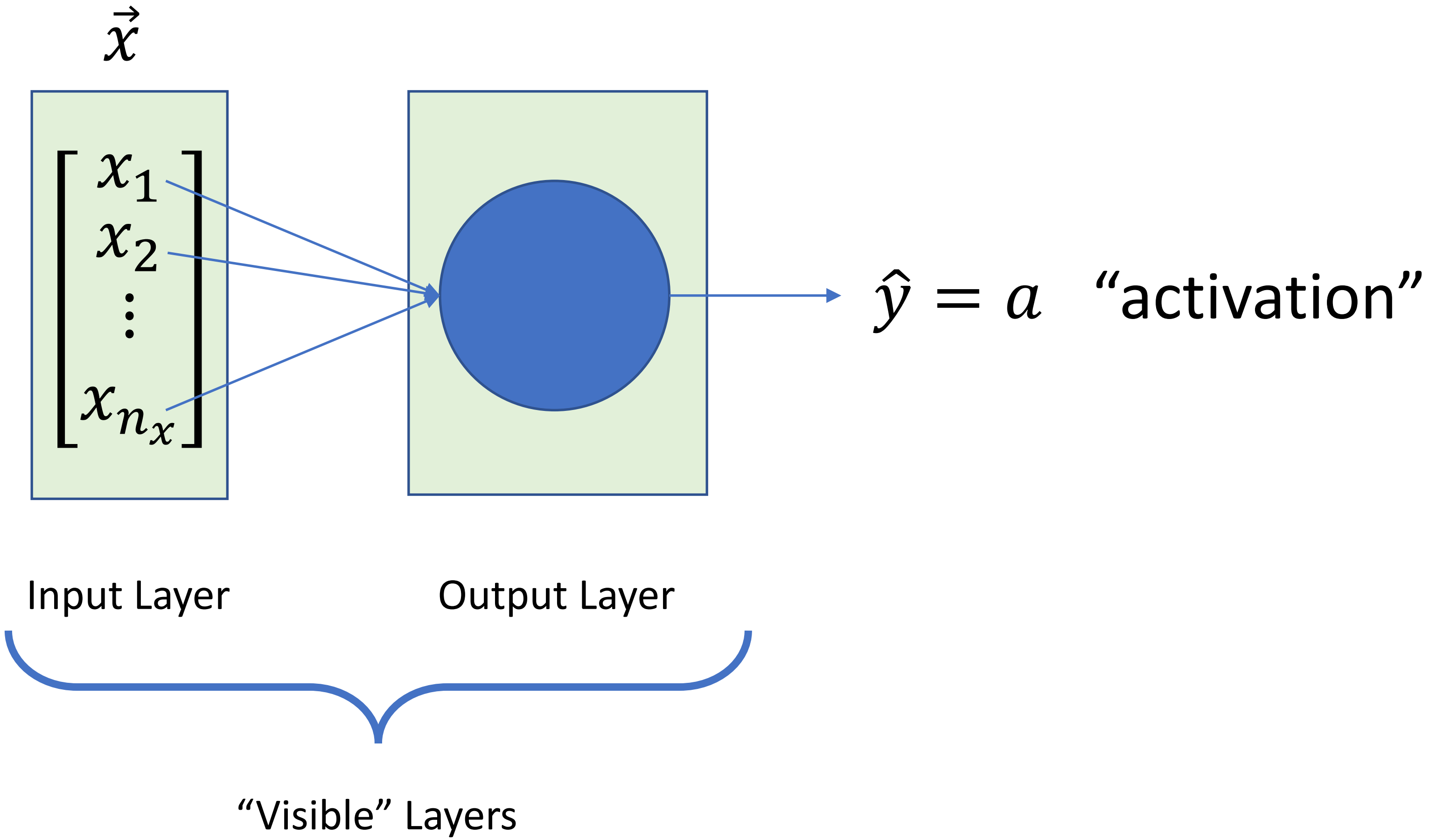
Neural Networks

$$\vec{x}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_x} \end{bmatrix}$$

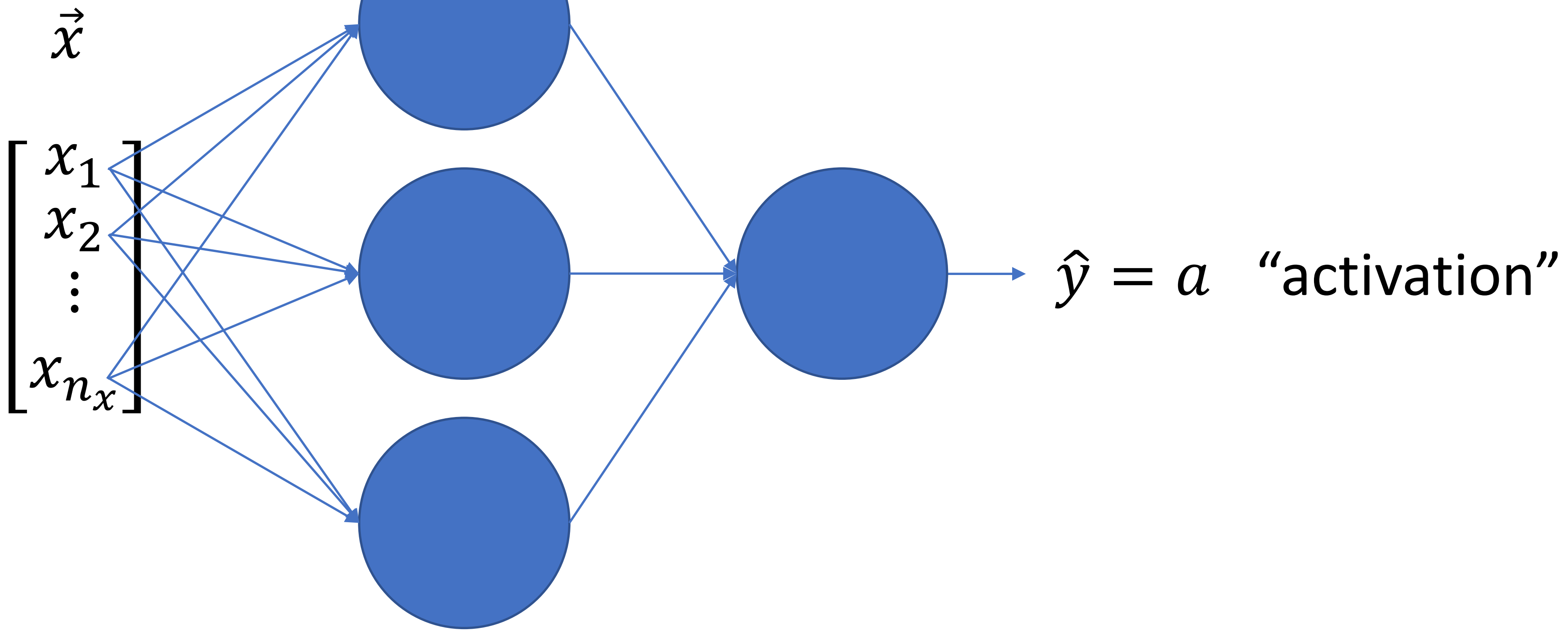

$\vec{x}$
x_1
x_2
$\vdots$
x_{n_x}

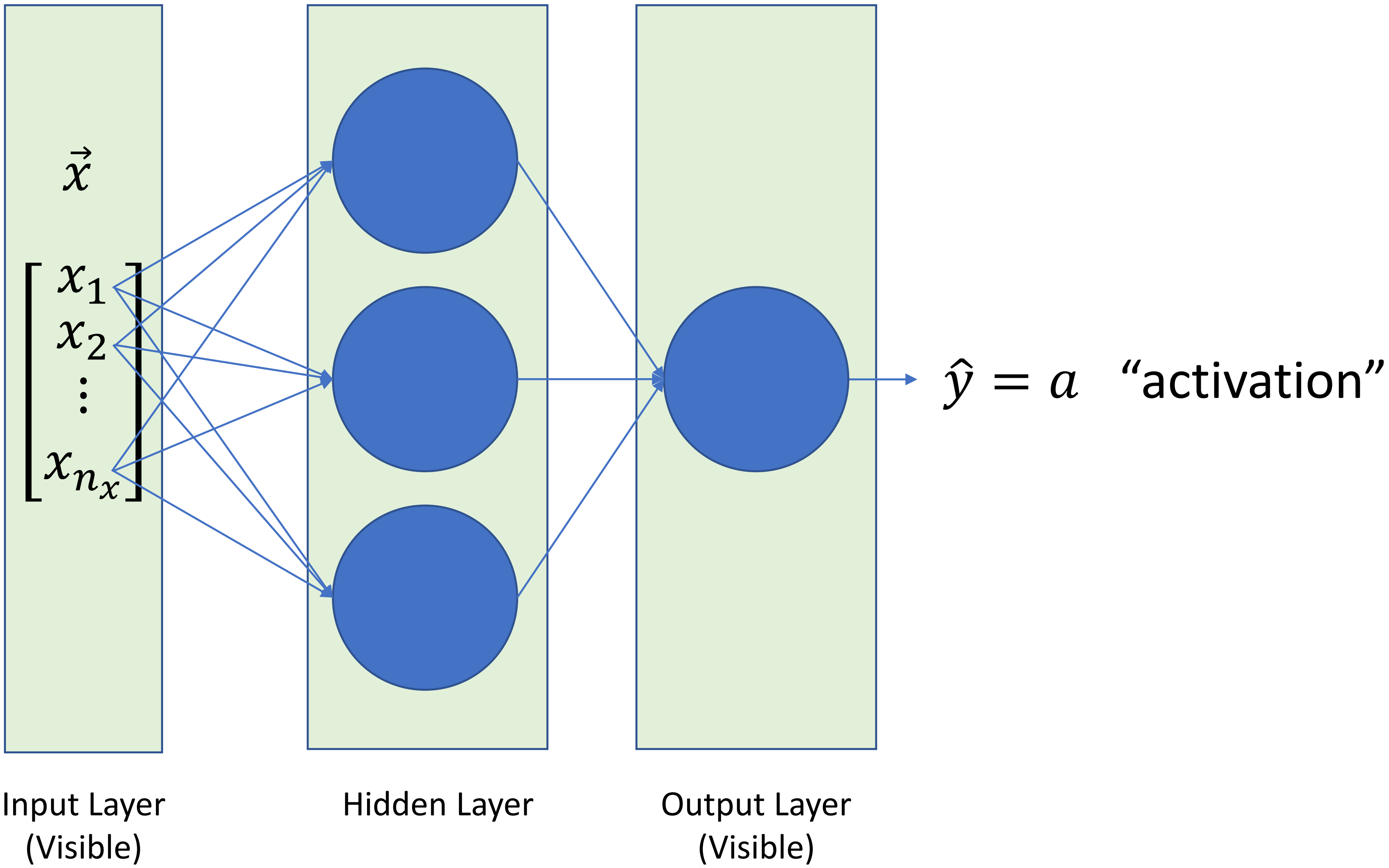

$\vec{x}$
$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_x} \end{bmatrix}$
$\hat{y} = a$ “activation”



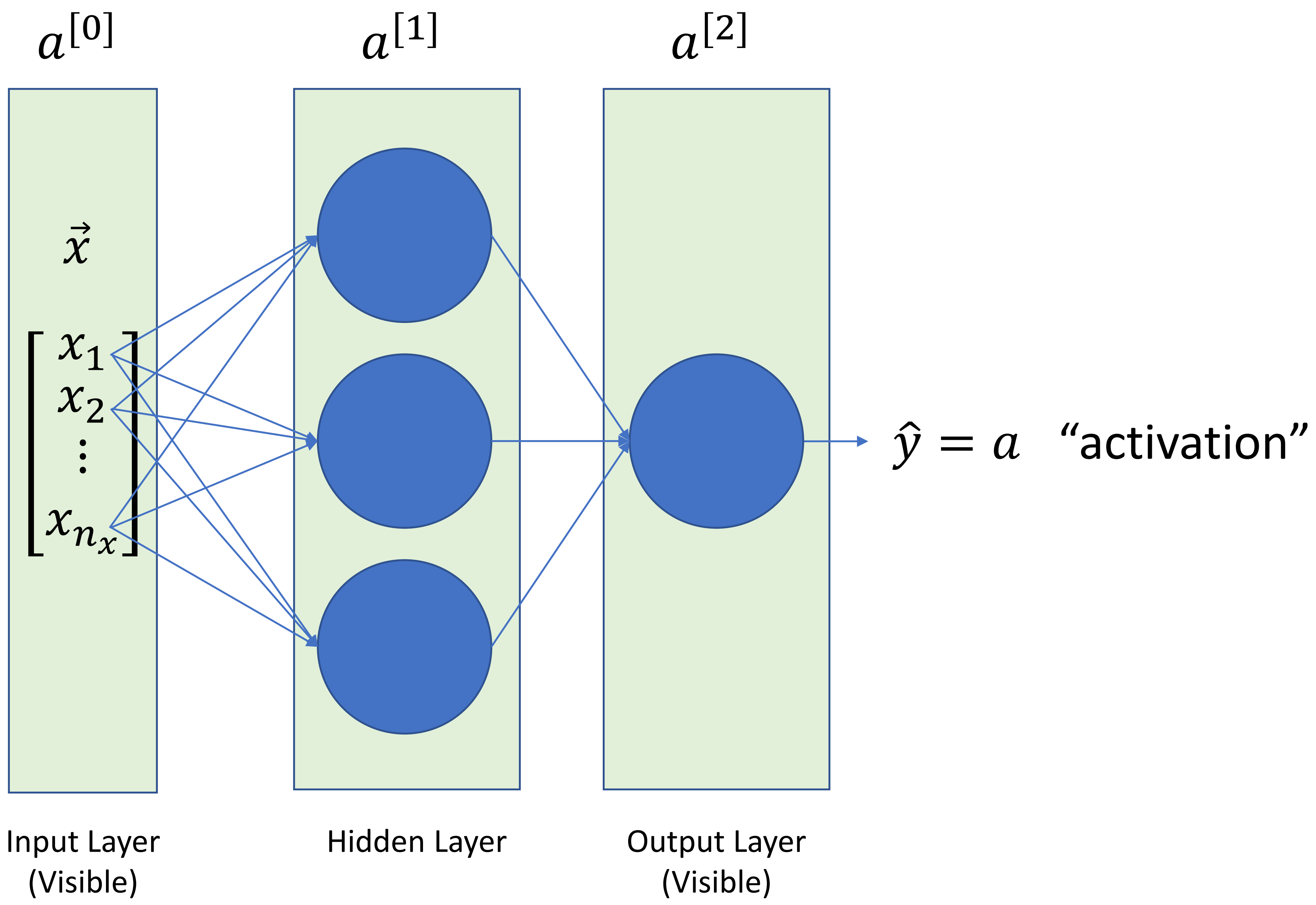
$\vec{x}$
x_1
x_2
$\vdots$
x_{n_x}
$\hat{y} = a$ "activation"
Input Layer
Output Layer
"Visible" Layers

One hidden layer

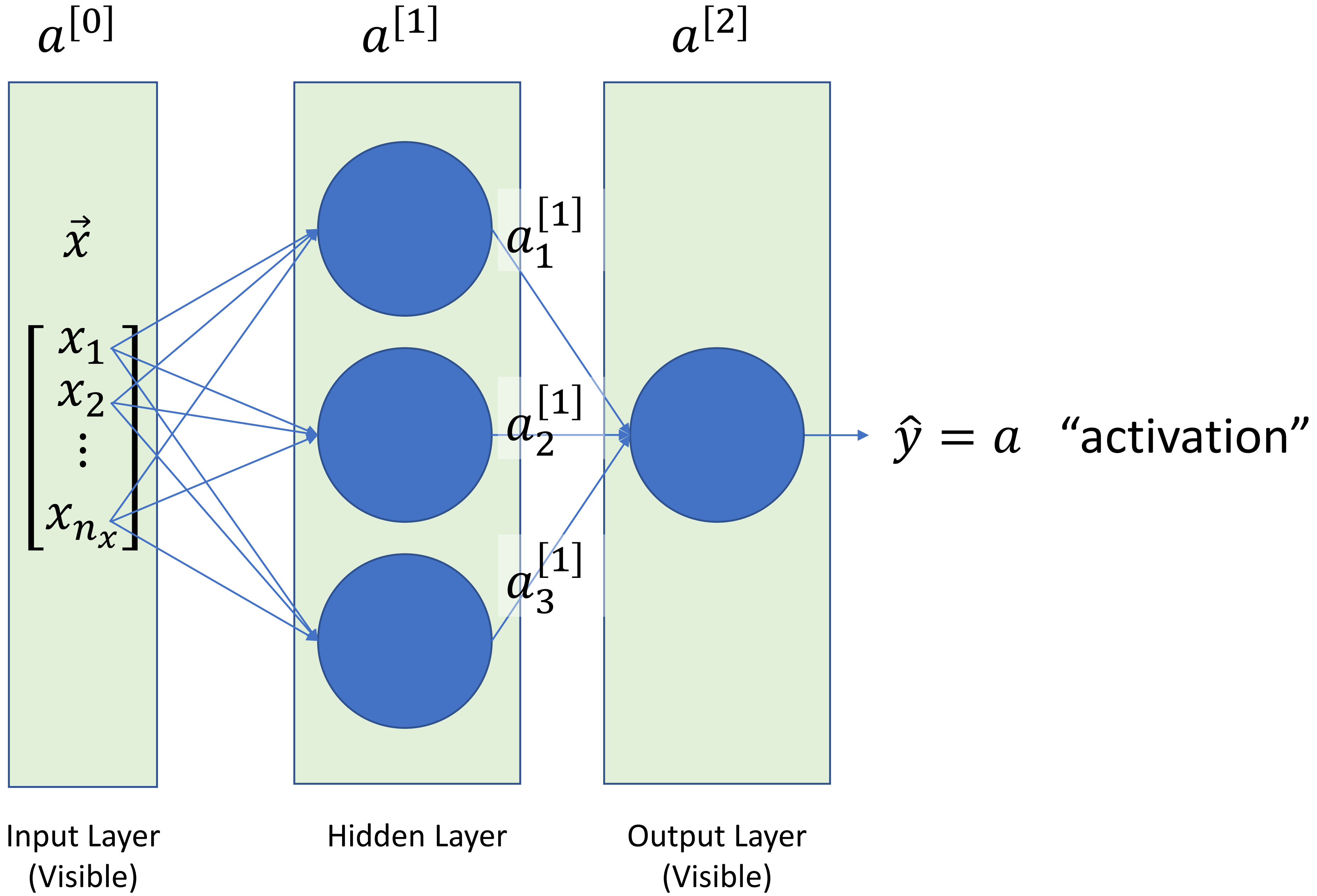

$\vec{x}$
$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_x} \end{bmatrix}$
$\hat{y} = a$ “activation”

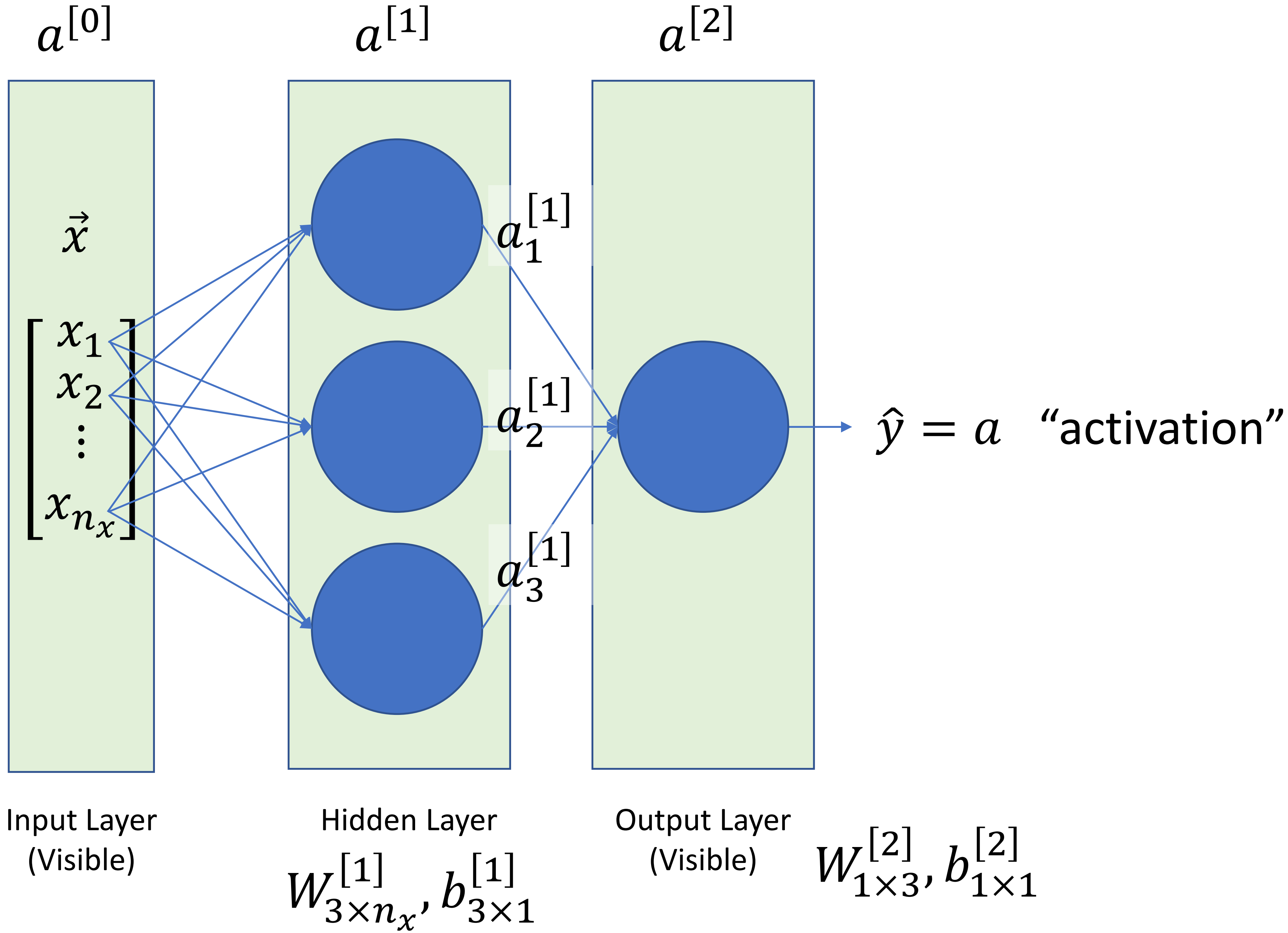


$\vec{x}$
$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_x} \end{bmatrix}$
$\hat{y} = a$ “activation”
Input Layer
(Visible)
Hidden Layer
Output Layer
(Visible)

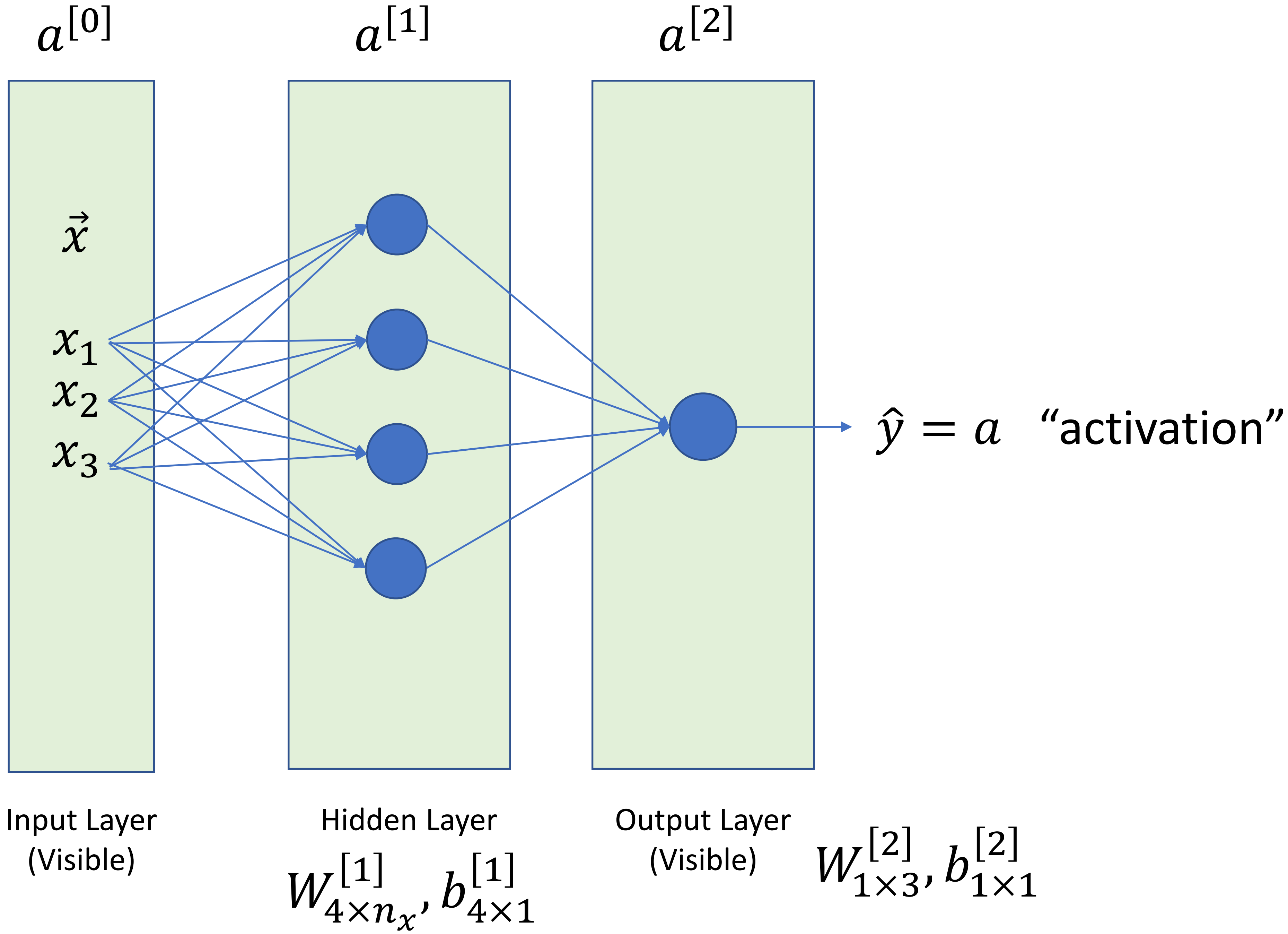


$a^{[0]}$
$a^{[1]}$
$a^{[2]}$
$\vec{x}$
x_1
x_2
$\vdots$
x_{n_x}
$\hat{y} = a$ "activation"
Input Layer (Visible)
Hidden Layer
Output Layer (Visible)

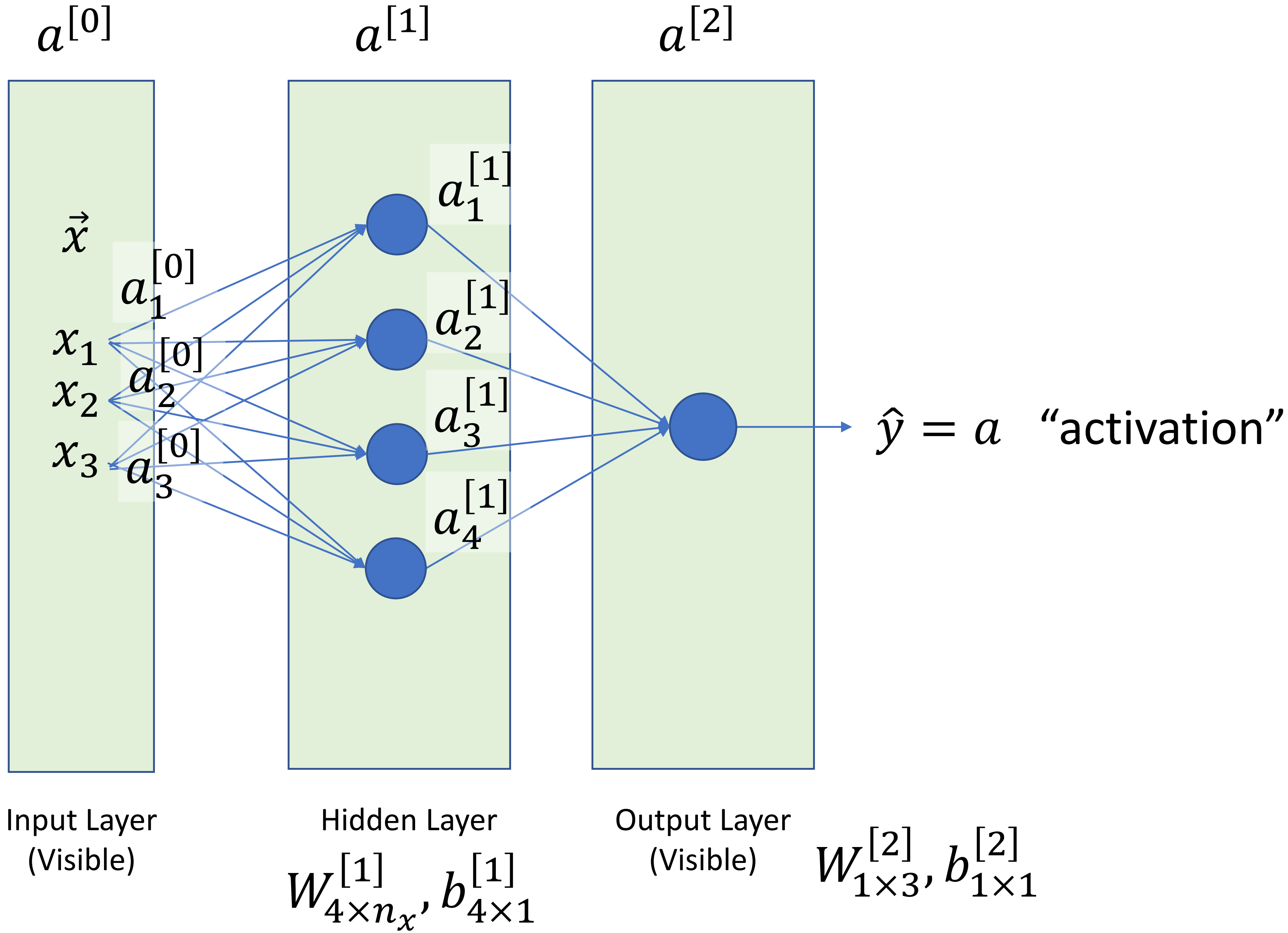

$a^{[0]}$
$a^{[1]}$
$a^{[2]}$
$\vec{x}$
$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_x} \end{bmatrix}$
$a_1^{[1]}$
$a_2^{[1]}$
$a_3^{[1]}$
$\hat{y} = a$ "activation"
Input Layer
(Visible)
Hidden Layer
Output Layer
(Visible)


$a^{[0]}$
$a^{[1]}$
$a^{[2]}$
$\vec{x}$
$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_x} \end{bmatrix}$
$a_1^{[1]}$
$a_2^{[1]}$
$a_3^{[1]}$
$\hat{y} = a$ "activation"
Input Layer
(Visible)
Hidden Layer
$W_{3\times n_x}^{[1]}, b_{3\times 1}^{[1]}$
Output Layer
(Visible)
$W_{1\times 3}^{[2]}, b_{1\times 1}^{[2]}$

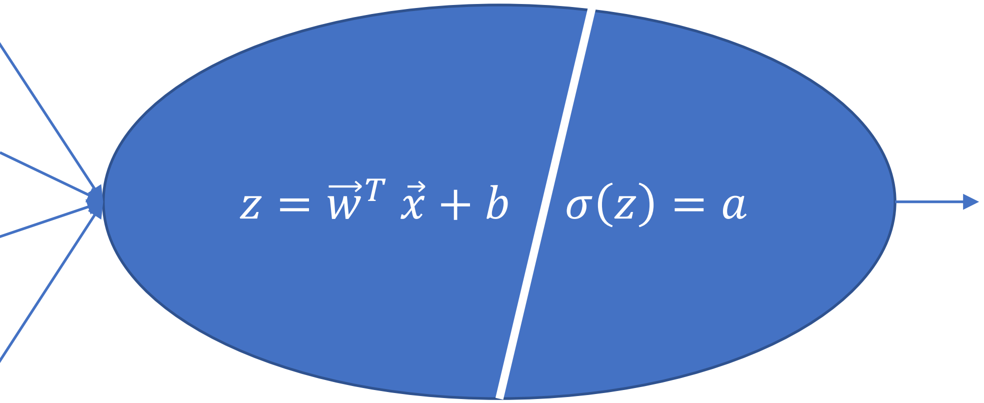
Computing an NN’s output



$a^{[0]}$
$a^{[1]}$
$a^{[2]}$
$\vec{x}$
x_1
x_2
x_3
$\hat{y} = a$ "activation"
Input Layer
(Visible)
Hidden Layer
$W^{[1]}_{4\times n_x}, b^{[1]}_{4\times 1}$
Output Layer
(Visible)
$W^{[2]}_{1\times 3}, b^{[2]}_{1\times 1}$


$a^{[0]}$
$a^{[1]}$
$a^{[2]}$
$\vec{x}$
$a_1^{[0]}$
x_1
$a_2^{[0]}$
x_2
x_3
$a_3^{[0]}$
$a_1^{[1]}$
$a_2^{[1]}$
$a_3^{[1]}$
$a_4^{[1]}$
$\hat{y} = a$ "activation"
Input Layer
(Visible)
Hidden Layer
$W_{4\times n_x}^{[1]}, b_{4\times 1}^{[1]}$
Output Layer
(Visible)
$W_{1\times 3}^{[2]}, b_{1\times 1}^{[2]}$

Each Neuron:

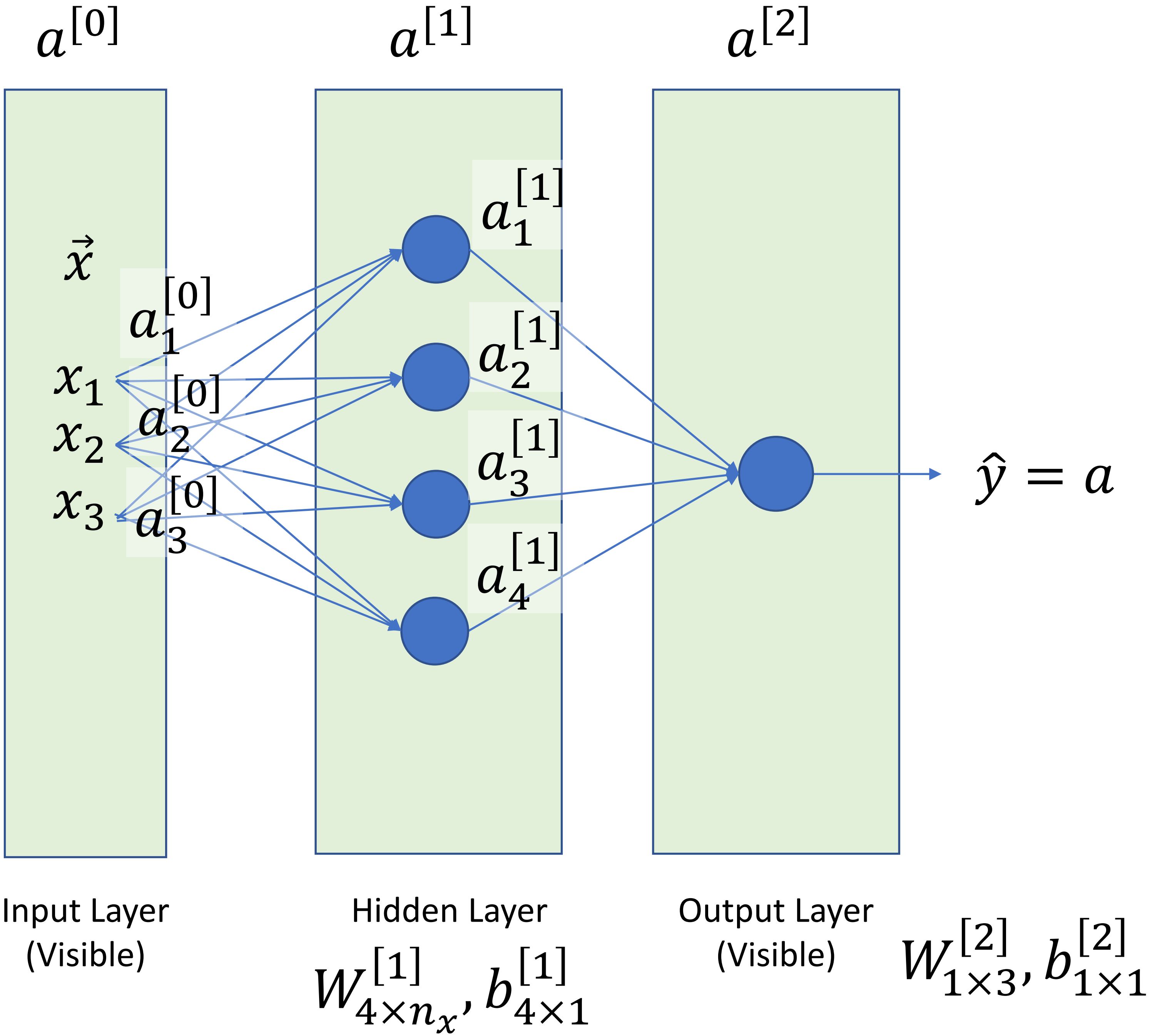


$$z_1^{[1]} = \vec{w}_1^{[1]T}\,\vec{x} + b_1^{[1]}; \qquad a_1^{[1]} = \sigma\left(z_1^{[1]}\right) + b_1^{[1]}$$

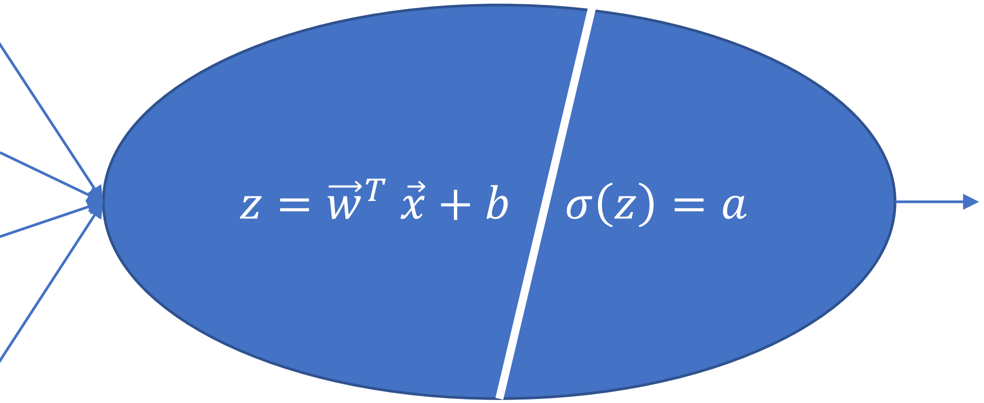
$$z_2^{[1]} = \vec{w}_2^{[1]T}\,\vec{x} + b_2^{[1]}; \qquad a_2^{[1]} = \sigma\left(z_2^{[1]}\right) + b_2^{[1]}$$

$$z_3^{[1]} = \vec{w}_3^{[1]T}\,\vec{x} + b_3^{[1]}; \qquad a_3^{[1]} = \sigma\left(z_3^{[1]}\right) + b_3^{[1]}$$

$$z_4^{[1]} = \vec{w}_4^{[1]T}\,\vec{x} + b_4^{[1]}; \qquad a_4^{[1]} = \sigma\left(z_4^{[1]}\right) + b_4^{[1]}$$



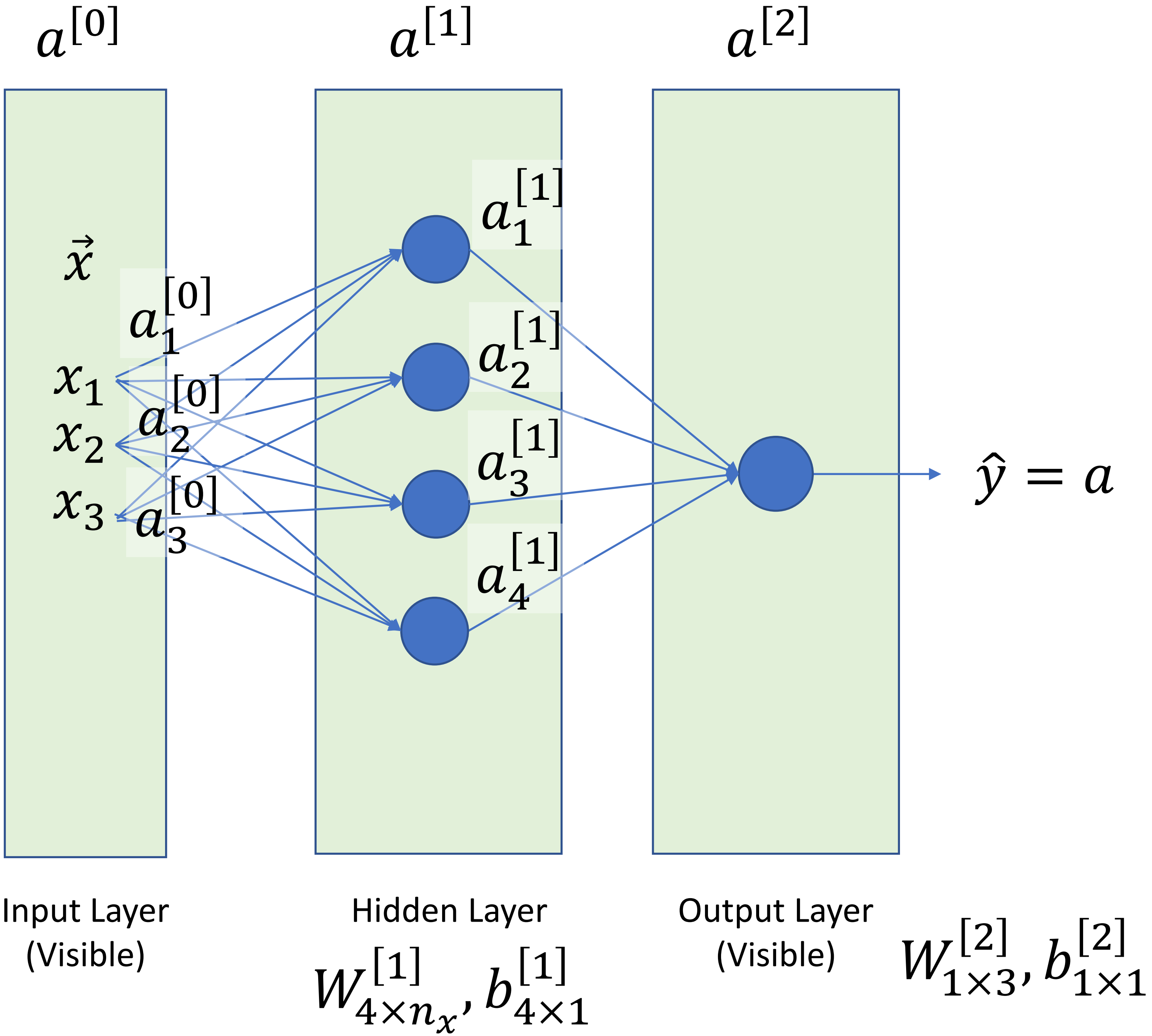
Each Neuron:



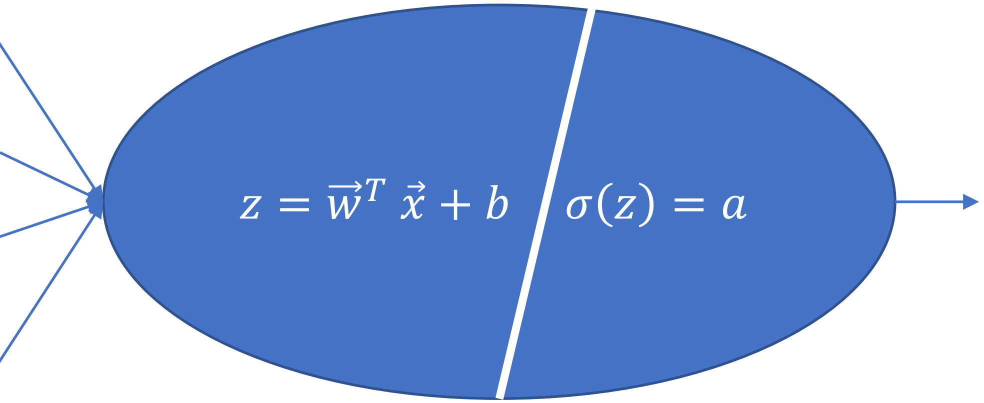
$$z_1^{[1]} = \vec{w}_1^{[1]T}\,\vec{x} + b_1^{[1]}; \qquad a_1^{[1]} = \sigma\left(z_1^{[1]}\right) + b_1^{[1]}$$
$$z_2^{[1]} = \vec{w}_2^{[1]T}\,\vec{x} + b_2^{[1]}; \qquad a_2^{[1]} = \sigma\left(z_2^{[1]}\right) + b_2^{[1]}$$
$$z_3^{[1]} = \vec{w}_3^{[1]T}\,\vec{x} + b_3^{[1]}; \qquad a_3^{[1]} = \sigma\left(z_3^{[1]}\right) + b_3^{[1]}$$
$$z_4^{[1]} = \vec{w}_4^{[1]T}\,\vec{x} + b_4^{[1]}; \qquad a_4^{[1]} = \sigma\left(z_4^{[1]}\right) + b_4^{[1]}$$

$$\begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \\ z_3^{[1]} \\ z_4^{[1]} \end{bmatrix} = \begin{bmatrix} -\vec{w}_1^{[1]T} - \\ -\vec{w}_2^{[1]T} - \\ -\vec{w}_3^{[1]T} - \\ -\vec{w}_4^{[1]T} - \end{bmatrix}_{4\times 3} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \\ b_3^{[1]} \\ b_4^{[1]} \end{bmatrix}$$

$$A_{4\times 1}^{[1]} = \sigma\left(Z_{4\times 1}^{[1]}\right)$$



Each Neuron:



$$z_1^{[1]} = \vec{w}_1^{[1]T} \vec{x} + b_1^{[1]}; \qquad a_1^{[1]} = \sigma\left(z_1^{[1]}\right) + b_1^{[1]}$$

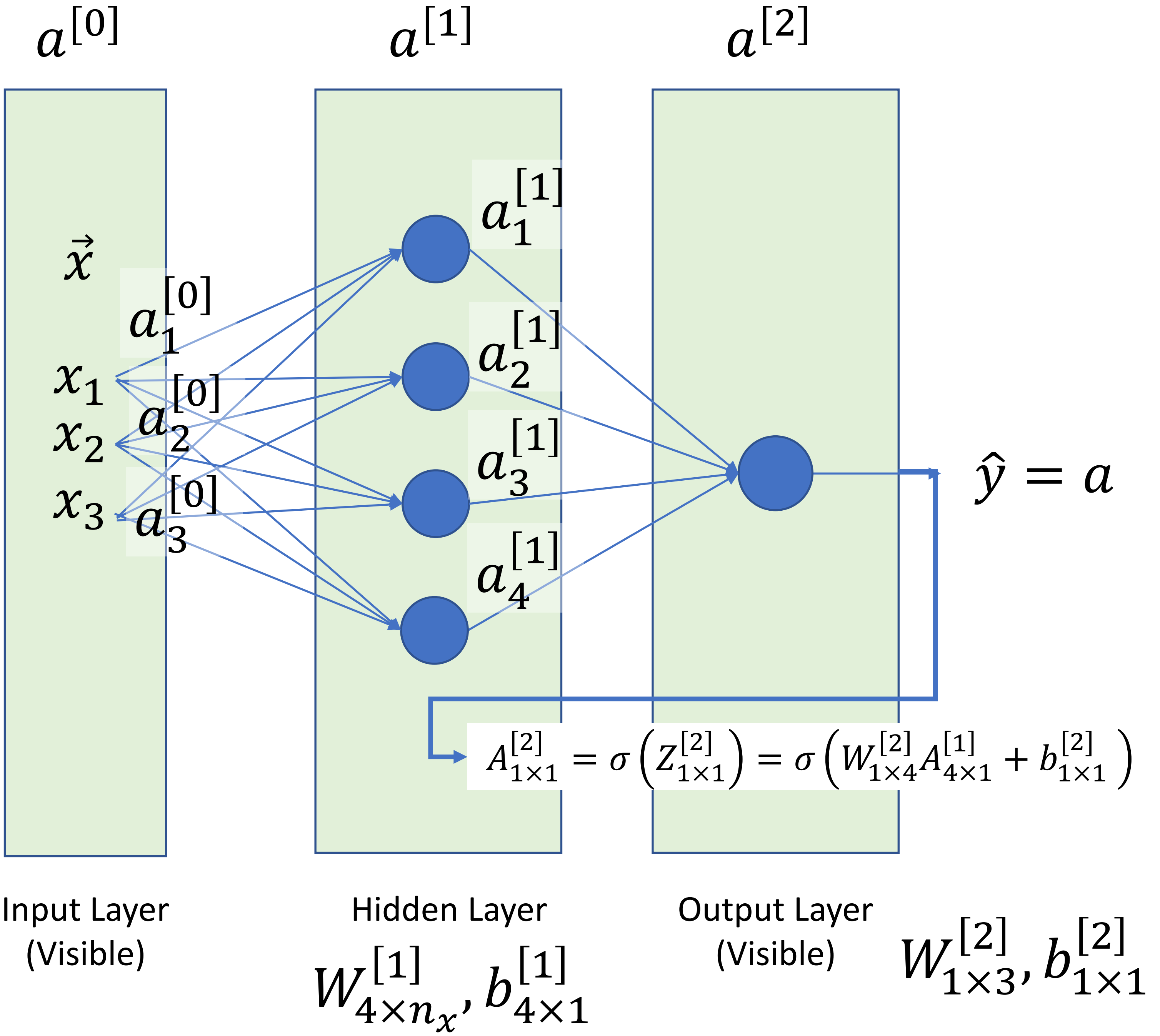
$$z_2^{[1]} = \vec{w}_2^{[1]T} \vec{x} + b_2^{[1]}; \qquad a_2^{[1]} = \sigma\left(z_2^{[1]}\right) + b_2^{[1]}$$

$$z_3^{[1]} = \vec{w}_3^{[1]T} \vec{x} + b_3^{[1]}; \qquad a_3^{[1]} = \sigma\left(z_3^{[1]}\right) + b_3^{[1]}$$

$$z_4^{[1]} = \vec{w}_4^{[1]T} \vec{x} + b_4^{[1]}; \qquad a_4^{[1]} = \sigma\left(z_4^{[1]}\right) + b_4^{[1]}$$

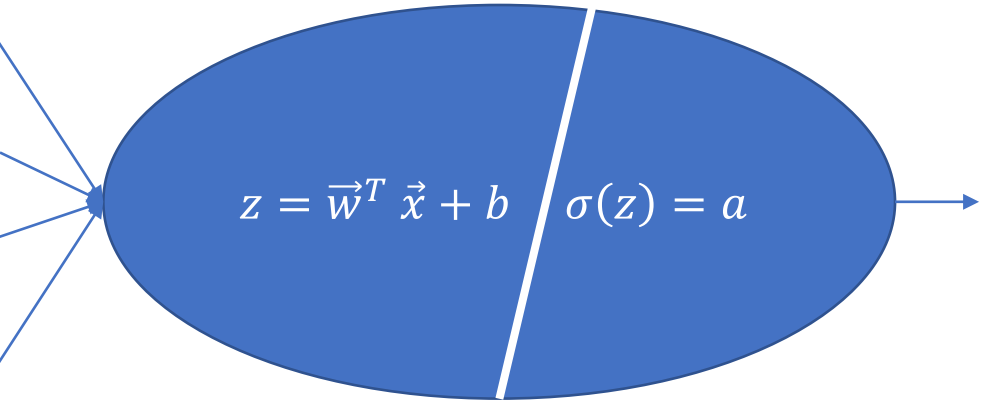
$$\begin{bmatrix} z_1^{[1]} \\ z_2^{[1]} \\ z_3^{[1]} \\ z_4^{[1]} \end{bmatrix} = \begin{bmatrix} -\vec{w}_1^{[1]T} - \\ -\vec{w}_2^{[1]T} - \\ -\vec{w}_3^{[1]T} - \\ -\vec{w}_4^{[1]T} - \end{bmatrix}_{4\times 3} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \\ b_3^{[1]} \\ b_4^{[1]} \end{bmatrix}$$

$$A_{4\times 1}^{[1]} = \sigma\left(Z_{4\times 1}^{[1]}\right) = \sigma\left(W_{4\times 3}^{[1]} A_{3\times 1}^{[0]} + b_{4\times 1}^{[1]}\right)$$



Gradient descent for NNs (one hidden layer)

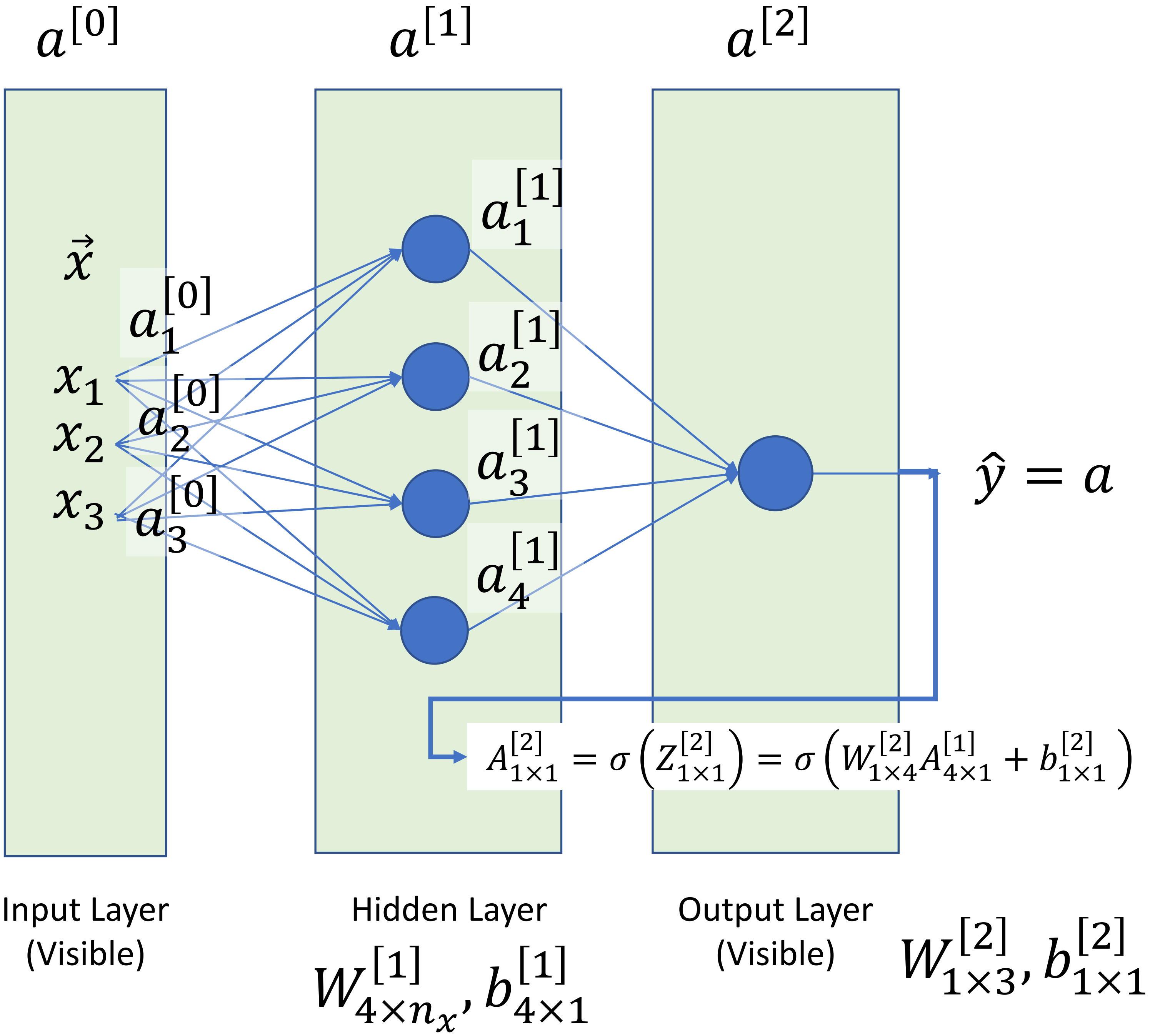
Each Neuron:



$$\begin{bmatrix} -\vec{w}_1^{[1]T} - \\ -\vec{w}_2^{[1]T} - \\ -\vec{w}_3^{[1]T} - \\ -\vec{w}_4^{[1]T} - \end{bmatrix}_{4\times 3} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \\ b_3^{[1]} \\ b_4^{[1]} \end{bmatrix}$$

Parameters:

- $W^{[1]}_{n^{[1]}\times n^{[0]}}, \vec{b}^{[1]}_{n^{[1]},1}, W^{[2]}_{n^{[2]},n^{[1]}}, \vec{b}^{[2]}_{n^{[2]},1}$
- $n^{[0]} = n_x$: Number of input features.
- $n^{[1]}$: Number of hidden units.
- $n^{[2]} = 1$: Number of output units.



Cost Function:

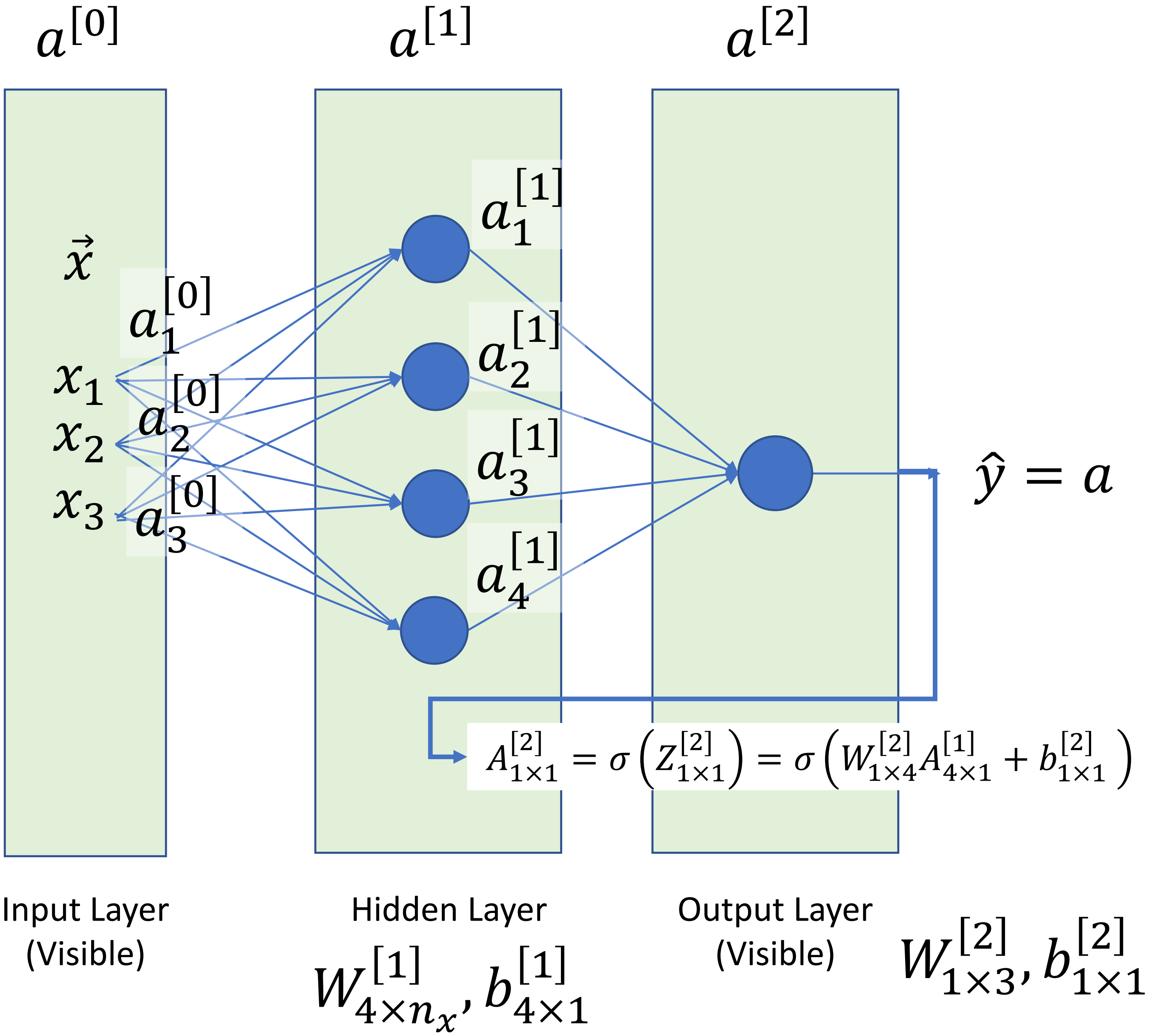
$$J\left(W^{[1]},\vec{b}^{[1]},W^{[2]},\vec{b}^{[2]}\right)=\frac{1}{m}\sum_{i=1}^{m}\mathcal{L}(\hat{y},y)$$

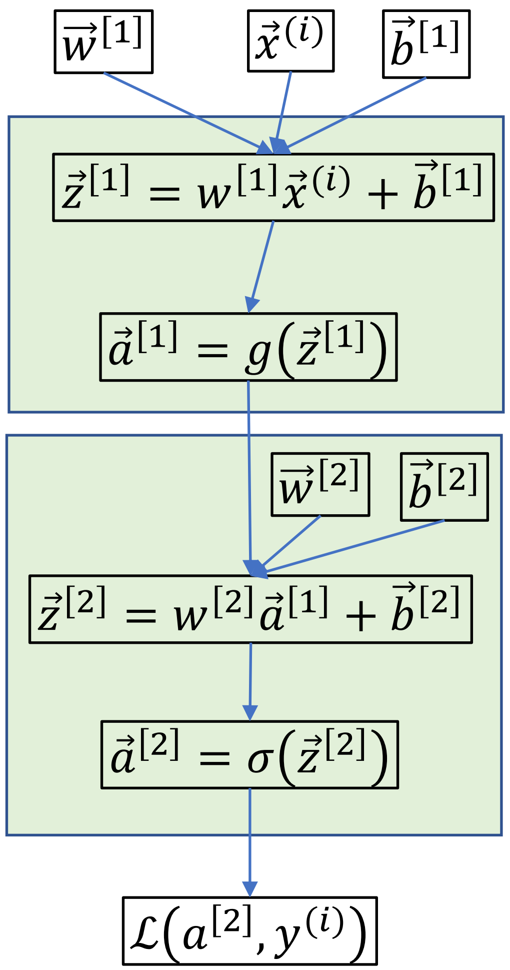
Gradient descent (to “learn” the parameters):

repeat {

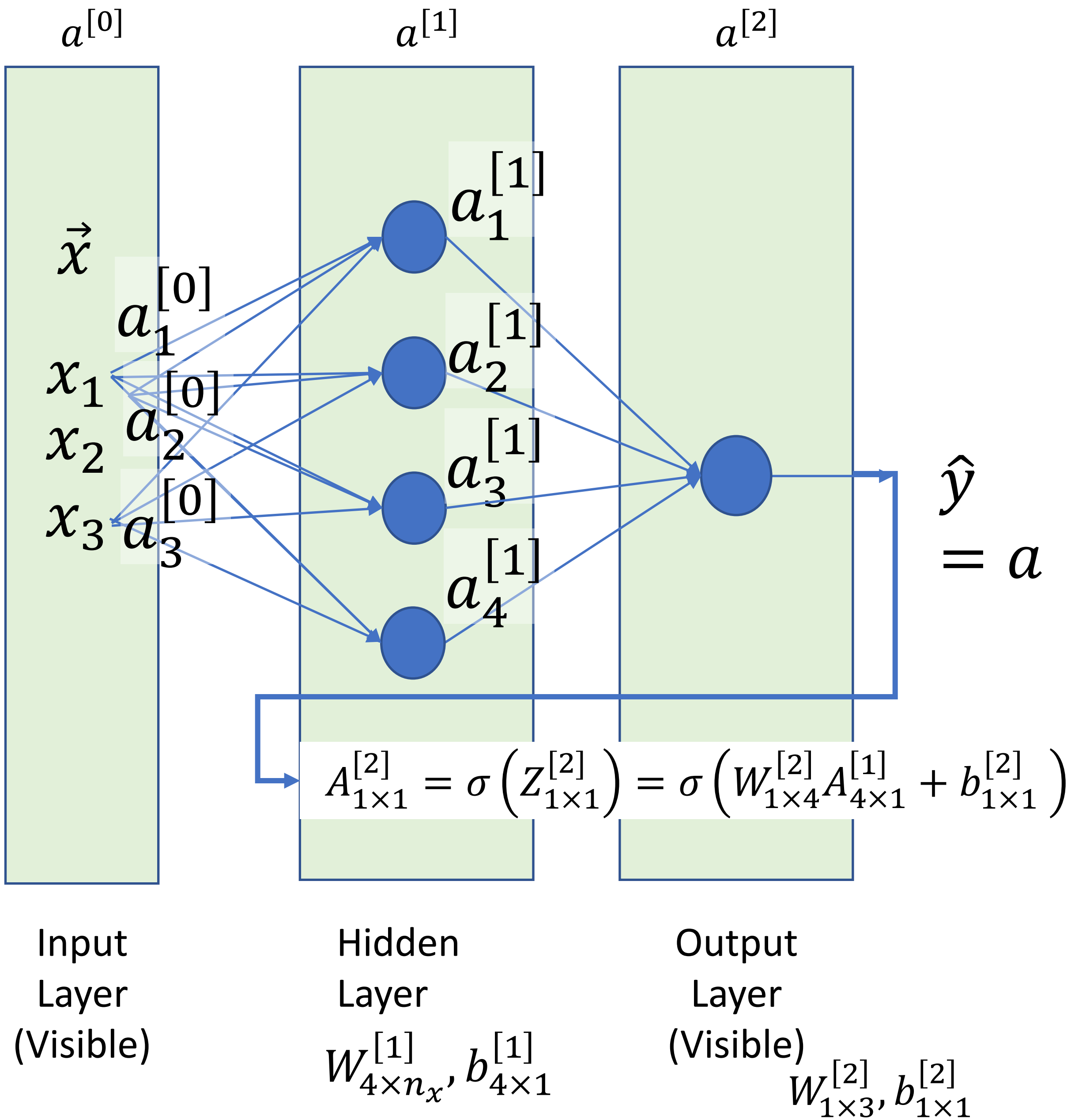
1. Compute predictions ($\hat{y}^{(i)}, i = 1, \dots, m$);
2. Compute derivatives
   - $dw^{[1]} = \frac{\partial J}{\partial w^{[1]}}$,
   - $\mathrm{db}^{[1]} = \frac{\partial J}{\partial b^{[1]}}$,
   - $dw^{[2]} = \frac{\partial J}{\partial w^{[2]}}$,
   - $db^{[2]} = \frac{\partial J}{\partial b^{[2]}}$.
3. Update parameters:
   - $w^{[1]} \leftarrow w^{[1]} - \alpha\, dw^{[1]}$
   - $b^{[1]} \leftarrow b^{[1]} + \alpha\, db^{[1]}$
   - $w^{[2]} \leftarrow w^{[2]} - \alpha\, dw^{[2]}$
   - $b^{[2]} \leftarrow b^{[2]} + \alpha\, db^{[2]}$

}

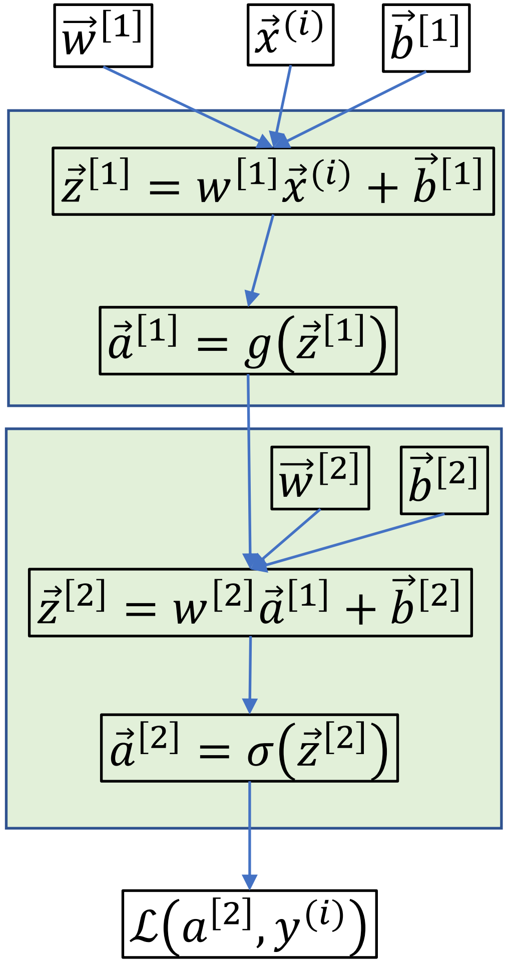




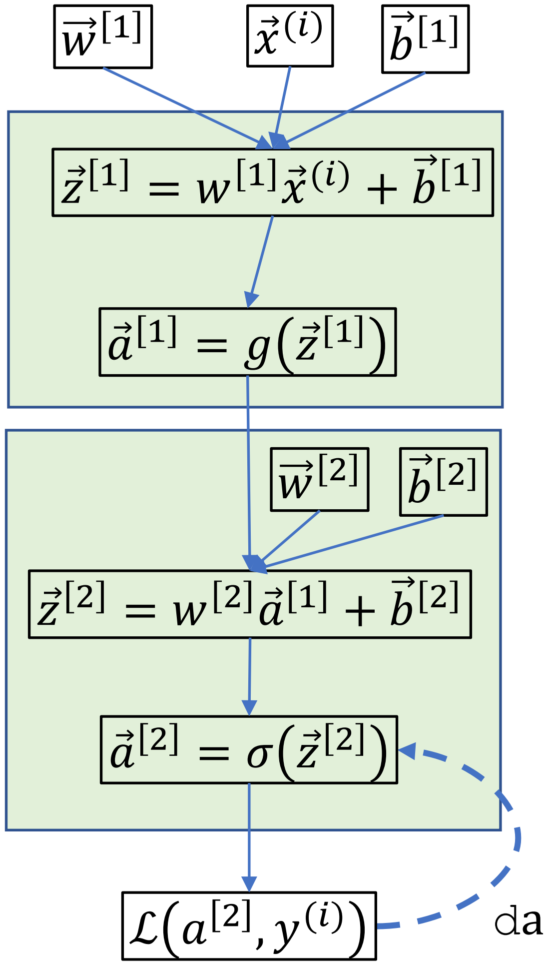
$\vec{w}^{[1]}$
$\vec{x}^{(i)}$
$\vec{b}^{[1]}$
$\vec{z}^{[1]} = w^{[1]}\vec{x}^{(i)} + \vec{b}^{[1]}$
$\vec{a}^{[1]} = g\left(\vec{z}^{[1]}\right)$
$\vec{w}^{[2]}$
$\vec{b}^{[2]}$
$\vec{z}^{[2]} = w^{[2]}\vec{a}^{[1]} + \vec{b}^{[2]}$
$\vec{a}^{[2]} = \sigma\left(\vec{z}^{[2]}\right)$
$\mathcal{L}\left(a^{[2]}, y^{(i)}\right)$



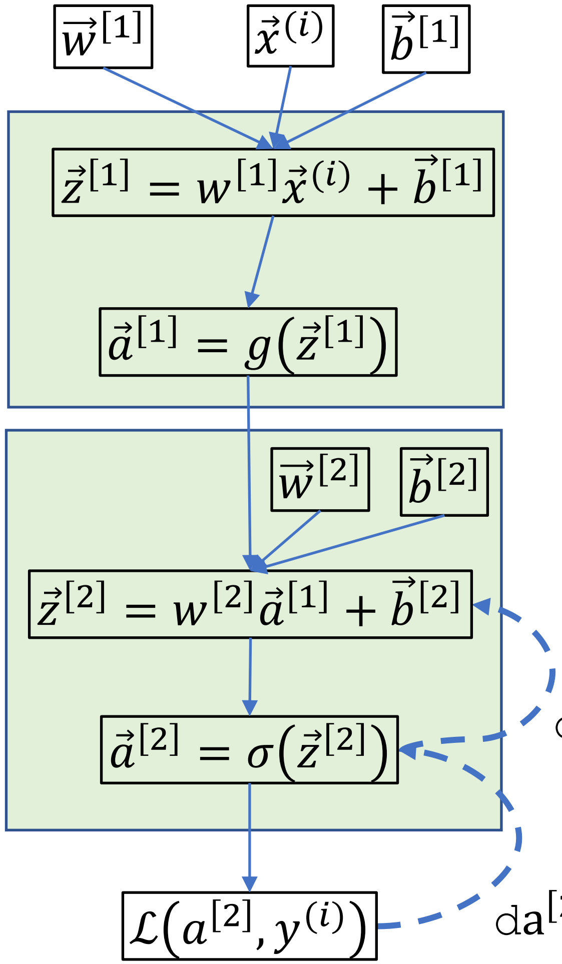
$a^{[0]}$
$a^{[1]}$
$a^{[2]}$
$\vec{x}$
x_1 $a_1^{[0]}$
x_2 $a_2^{[0]}$
x_3 $a_3^{[0]}$
$a_1^{[1]}$
$a_2^{[1]}$
$a_3^{[1]}$
$a_4^{[1]}$
$\hat{y} = a$
$A_{1\times1}^{[2]} = \sigma\left(Z_{1\times1}^{[2]}\right) = \sigma\left(W_{1\times4}^{[2]}A_{4\times1}^{[1]} + b_{1\times1}^{[2]}\right)$
Input Layer (Visible)
Hidden Layer
$W_{4\times n_x}^{[1]}, b_{4\times1}^{[1]}$
Output Layer (Visible)
$W_{1\times3}^{[2]}, b_{1\times1}^{[2]}$



$\vec{w}^{[1]}$
$\vec{x}^{(i)}$
$\vec{b}^{[1]}$
$\vec{z}^{[1]} = w^{[1]}\vec{x}^{(i)} + \vec{b}^{[1]}$
$\vec{a}^{[1]} = g\left(\vec{z}^{[1]}\right)$
$\vec{w}^{[2]}$
$\vec{b}^{[2]}$
$\vec{z}^{[2]} = w^{[2]}\vec{a}^{[1]} + \vec{b}^{[2]}$
$\vec{a}^{[2]} = \sigma\left(\vec{z}^{[2]}\right)$
$\mathcal{L}\left(a^{[2]}, y^{(i)}\right)$

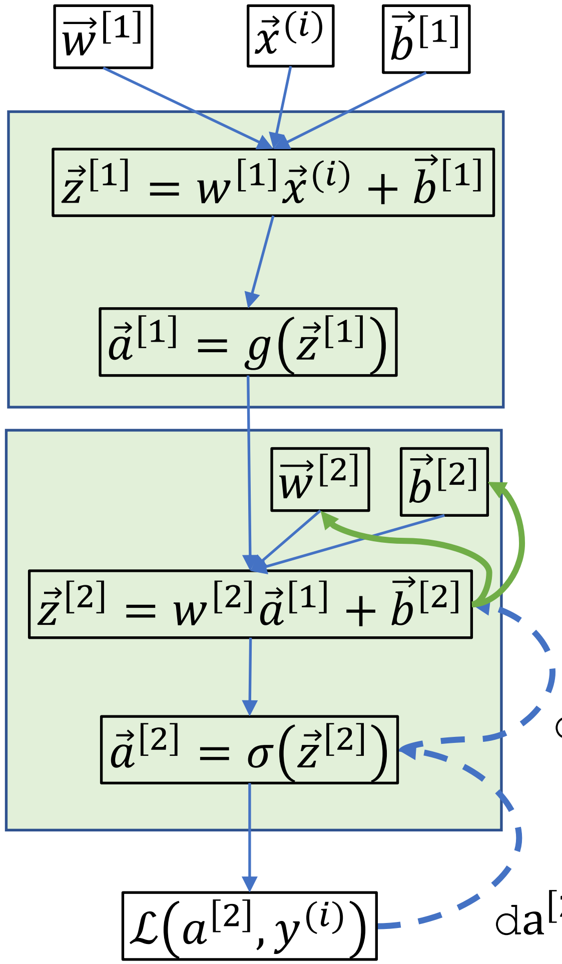


$$\mathrm{da}^{[2]} = \frac{\partial \mathcal{L}}{\partial a^{[2]}} = -\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}$$

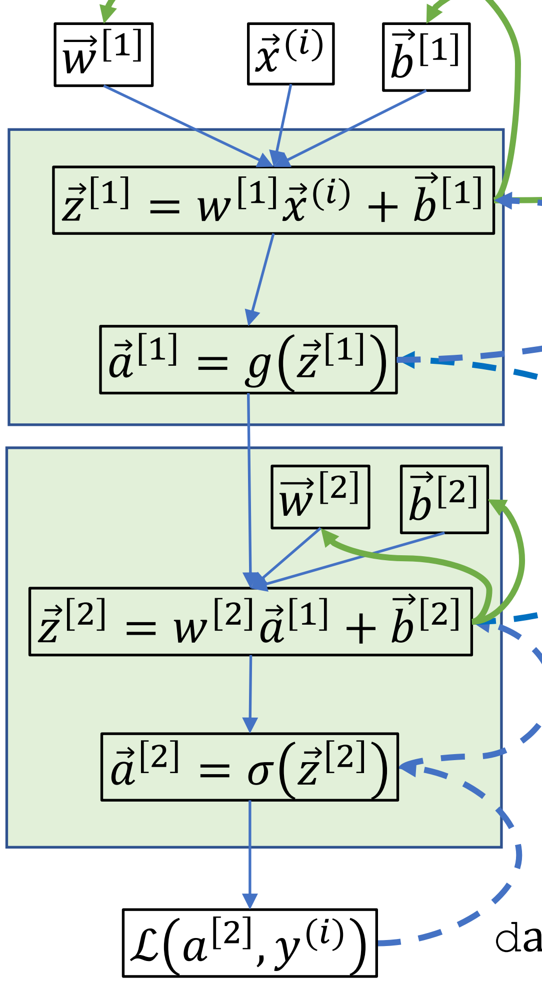


$$\mathrm{dz}^{[2]} = \frac{\partial \mathcal{L}}{\partial a^{[2]}}\frac{\partial a^{[2]}}{\partial z^{[2]}} = \left[-\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}\right]\left[a^{[2]}\left(1 - a^{[2]}\right)\right] = \left[a^{[2]} - y\right]$$

$$\mathrm{da}^{[2]} = \frac{\partial \mathcal{L}}{\partial a^{[2]}} = -\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}$$



$$\left.\begin{aligned}
&\mathrm{db}^{[2]} = \frac{\partial \mathcal{L}}{\partial b^{[2]}} = \left[\frac{\partial \mathcal{L}}{\partial a^{[2]}}\frac{\partial a^{[2]}}{\partial z^{[2]}}\right]\frac{\partial z^{[2]}}{\partial b^{[2]}} = \left[\boldsymbol{a}^{[2]} - \boldsymbol{y}\right] \\
&\mathrm{dw}^{[2]} = \frac{\partial \mathcal{L}}{\partial w^{[2]}} = \left[\frac{\partial \mathcal{L}}{\partial a^{[2]}}\frac{\partial a^{[2]}}{\partial z^{[2]}}\right]\frac{\partial z^{[2]}}{\partial w^{[2]}} = \left[\boldsymbol{a}^{[2]} - \boldsymbol{y}\right]\boldsymbol{a}^{[1]} \\
&\mathrm{dz}^{[2]} = \frac{\partial \mathcal{L}}{\partial a^{[2]}}\frac{\partial a^{[2]}}{\partial z^{[2]}} = \left[-\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}\right]\left[a^{[2]}\left(1 - a^{[2]}\right)\right] = \left[a^{[2]} - y\right] \\
&\mathrm{da}^{[2]} = \frac{\partial \mathcal{L}}{\partial a^{[2]}} = -\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}
\end{aligned}\right\} \text{Second Layer}$$



$$\mathrm{dw}^{[1]} = \frac{\partial\mathcal{L}}{\partial z^{[1]}}\frac{\partial z^{[1]}}{\partial w^{[1]}} = dz^{[1]}\vec{x}$$

$$\mathrm{db}^{[1]} = dz^{[1]}$$

$$\mathrm{dz}^{[1]} = \frac{\partial\mathcal{L}}{\partial z^{[1]}} = \left[\frac{\partial\mathcal{L}}{\partial a^{[2]}}\frac{\partial a^{[2]}}{\partial z^{[2]}}\right]\frac{\partial z^{[2]}}{\partial a^{[1]}}\frac{\partial a^{[1]}}{\partial z^{[1]}} = [a^{[2]} - y][w^{[2]}][g'(\vec{z}^{[1]})]$$

$$\mathrm{da}^{[1]} = \frac{\partial\mathcal{L}}{\partial a^{[1]}} = \left[\frac{\partial\mathcal{L}}{\partial a^{[2]}}\frac{\partial a^{[2]}}{\partial z^{[2]}}\right]\frac{\partial z^{[2]}}{\partial a^{[1]}} = [a^{[2]} - y][w^{[2]}]$$

$$\left.\begin{aligned}
\mathrm{db}^{[2]} &= \frac{\partial\mathcal{L}}{\partial b^{[2]}} = \left[\frac{\partial\mathcal{L}}{\partial a^{[2]}}\frac{\partial a^{[2]}}{\partial z^{[2]}}\right]\frac{\partial z^{[2]}}{\partial b^{[2]}} = \boldsymbol{[a^{[2]} - y]}\\
\mathrm{dw}^{[2]} &= \frac{\partial\mathcal{L}}{\partial w^{[2]}} = \left[\frac{\partial\mathcal{L}}{\partial a^{[2]}}\frac{\partial a^{[2]}}{\partial z^{[2]}}\right]\frac{\partial z^{[2]}}{\partial w^{[2]}} = \boldsymbol{[a^{[2]} - y]a^{[1]}}\\
\mathrm{dz}^{[2]} &= \frac{\partial\mathcal{L}}{\partial a^{[2]}}\frac{\partial a^{[2]}}{\partial z^{[2]}} = \left[-\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}\right][a^{[2]}(1 - a^{[2]})] = [a^{[2]} - y]\\
\mathrm{da}^{[2]} &= \frac{\partial\mathcal{L}}{\partial a^{[2]}} = -\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}
\end{aligned}\right\}\text{Second Layer}$$

Six key equations

1. $\mathrm{d}z^{[2]} = \frac{\partial \mathcal{L}}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}} = \left[-\frac{y}{a^{[2]}} + \frac{1-y}{1-a^{[2]}}\right]\left[a^{[2]}\left(1 - a^{[2]}\right)\right] = \left[\boldsymbol{a}^{[\mathbf{2}]} - \boldsymbol{y}\right]$

2. $\mathrm{dw}^{[2]} = \frac{\partial \mathcal{L}}{\partial w^{[2]}} = \left[\frac{\partial \mathcal{L}}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}}\right] \frac{\partial z^{[2]}}{\partial w^{[2]}} = \left[\boldsymbol{a}^{[\mathbf{2}]} - \boldsymbol{y}\right]\boldsymbol{a}^{[\mathbf{1}]}$

3. $\mathrm{db}^{[2]} = \frac{\partial \mathcal{L}}{\partial b^{[2]}} = \left[\frac{\partial \mathcal{L}}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}}\right] \frac{\partial z^{[2]}}{\partial b^{[2]}} = \left[\boldsymbol{a}^{[\mathbf{2}]} - \boldsymbol{y}\right]$

4. $\mathrm{d}z^{[1]} = \frac{\partial \mathcal{L}}{\partial z^{[1]}} = \left[\frac{\partial \mathcal{L}}{\partial a^{[2]}} \frac{\partial a^{[2]}}{\partial z^{[2]}}\right] \frac{\partial z^{[2]}}{\partial a^{[1]}} \frac{\partial a^{[1]}}{\partial z^{[1]}} = \left[\boldsymbol{a}^{[\mathbf{2}]} - \boldsymbol{y}\right]\left[\boldsymbol{w}^{[\mathbf{2}]}\right]\left[\boldsymbol{g}'\left(\vec{\boldsymbol{z}}^{[\mathbf{1}]}\right)\right]$

5. $\mathrm{dw}^{[1]} = \frac{\partial \mathcal{L}}{\partial z^{[1]}} \frac{\partial z^{[1]}}{\partial w^{[1]}} = dz^{[1]}\vec{x} = \left[\boldsymbol{a}^{[\mathbf{2}]} - \boldsymbol{y}\right]\left[\boldsymbol{w}^{[\mathbf{2}]}\right]\left[\boldsymbol{g}'\left(\vec{\boldsymbol{z}}^{[\mathbf{1}]}\right)\right]\vec{\boldsymbol{x}}$

6. $\mathrm{db}^{[1]} = dz^{[1]} = \left[\boldsymbol{a}^{[\mathbf{2}]} - \boldsymbol{y}\right]\left[\boldsymbol{w}^{[\mathbf{2}]}\right]\left[\boldsymbol{g}'\left(\vec{\boldsymbol{z}}^{[\mathbf{1}]}\right)\right]$

For m samples

$$\begin{bmatrix} | & | & & | \\ \vec{z}^{[l](1)} & \vec{z}^{[l](2)} & \cdots & \vec{z}^{[l](m)} \\ | & | & & | \end{bmatrix}_{n^{[l]}\times m} = \begin{bmatrix} - & \vec{w}_1^{[l]T} & - \\ - & \vec{w}_2^{[l]T} & - \\ & \vdots & \\ - & \vec{w}_{n^{[l]}}^{[l]T} & - \end{bmatrix}_{n^{[l]}\times n^{[l]-1}} \begin{bmatrix} | & | & & | \\ \vec{a}^{[l-1](1)} & \vec{a}^{[l-1](2)} & \cdots & \vec{a}^{[l-1](m)} \\ | & | & & | \end{bmatrix}_{n^{[l-1]}\times m} + \begin{bmatrix} b_1^{[l]} \\ b_2^{[l]} \\ \vdots \\ b_{n^{[l]}}^{[l]} \end{bmatrix}_{n^{[l]}\times m}$$

$$Z^{[l]} = W^{[l]T}X + B$$

$$A^{[l]} = g^{[l]}(Z) = g^{[l]}\left(W^{[l]T}X + B\right)$$

Deep Neural Networks

DL Building Blocks

Forward:

Input:

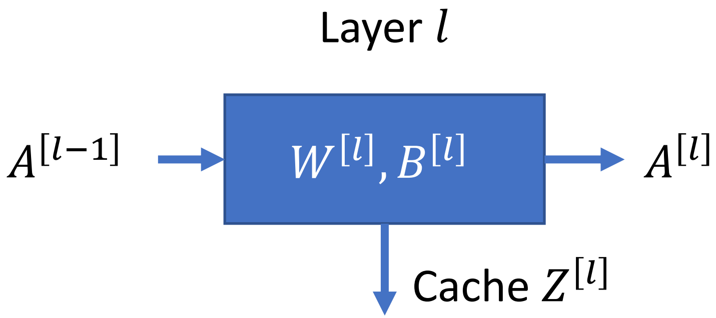
- $A^{[l-1]}$.

Compute:

- $Z^{[l]} = W^{[l]}A^{[l-1]} + B$.
- $A^{[l]} = g^{[l]}(Z^{[l]})$.

Output:

- $A^{[l]}$.
- Cache $Z^{[l]}$.



DL Building Blocks

Forward:

Input:

- $A^{[l-1]}$.

Compute:

- $Z^{[l]} = W^{[l]}A^{[l-1]} + B$.
- $A^{[l]} = g^{[l]}(Z^{[l]})$.

Output:

- $A^{[l]}$.
- Cache $Z^{[l]}$.

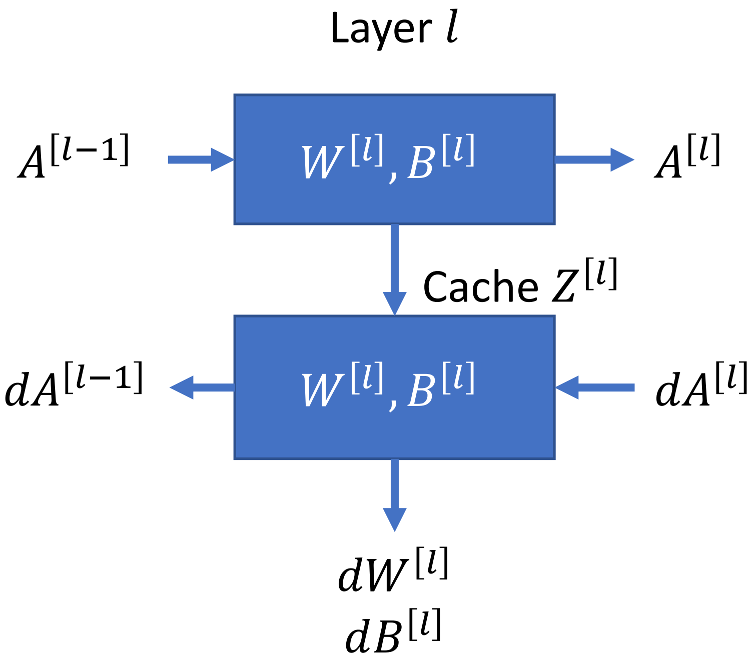
Backward:

Input:

- $dA^{[l]}$.
- Cache $Z^{[l]}$.

Output:

- $dA^{[l-1]}$.
- $dW^{[l]}$.
- $dB^{[l]}$.



DL Building Blocks

Forward:

Input:

- $A^{[l-1]}$.

Compute:

- $Z^{[l]} = W^{[l]}A^{[l-1]} + B$.
- $A^{[l]} = g^{[l]}(Z^{[l]})$.

Output:

- $A^{[l]}$.
- Cache $Z^{[l]}$.

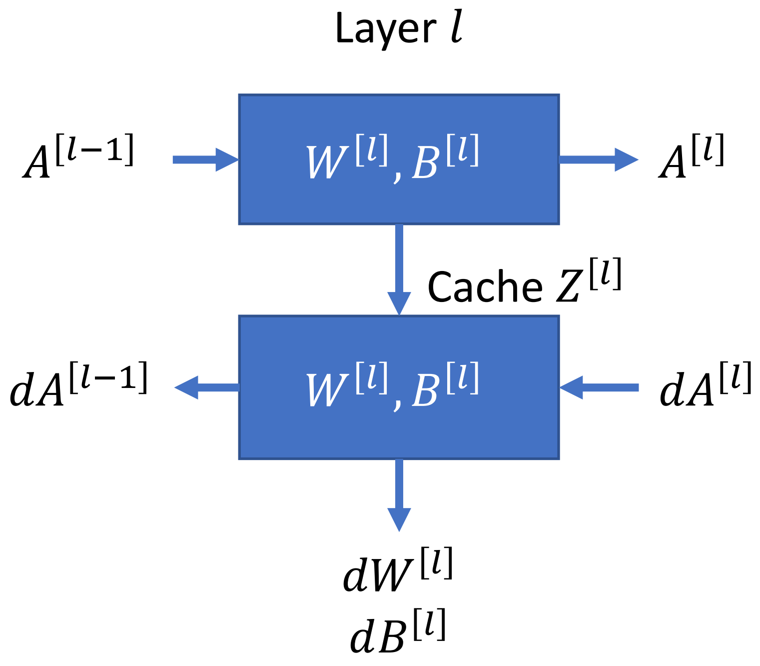
Backward:

Input:

- $dA^{[l]}$.
- Cache $Z^{[l]}$.

Output:

- $dA^{[l-1]}$.
- $dW^{[l]}$.
- $dB^{[l]}$.



$$W^{[l]} \leftarrow W^{[l]} - \alpha dW^{[l]}$$
$$B^{[l]} \leftarrow B^{[l]} - \alpha dB^{[l]}$$

Training adjustment

- Parameters:
  - $W^{[l]}, B^{[l]}$.
- Hyperparameters:
  - Learning rate α.
  - Number of iterations.
  - Number of hidden units $n^{[1]}, n^{[2]}, \ldots$.
  - Choice of activation function.

Thank you