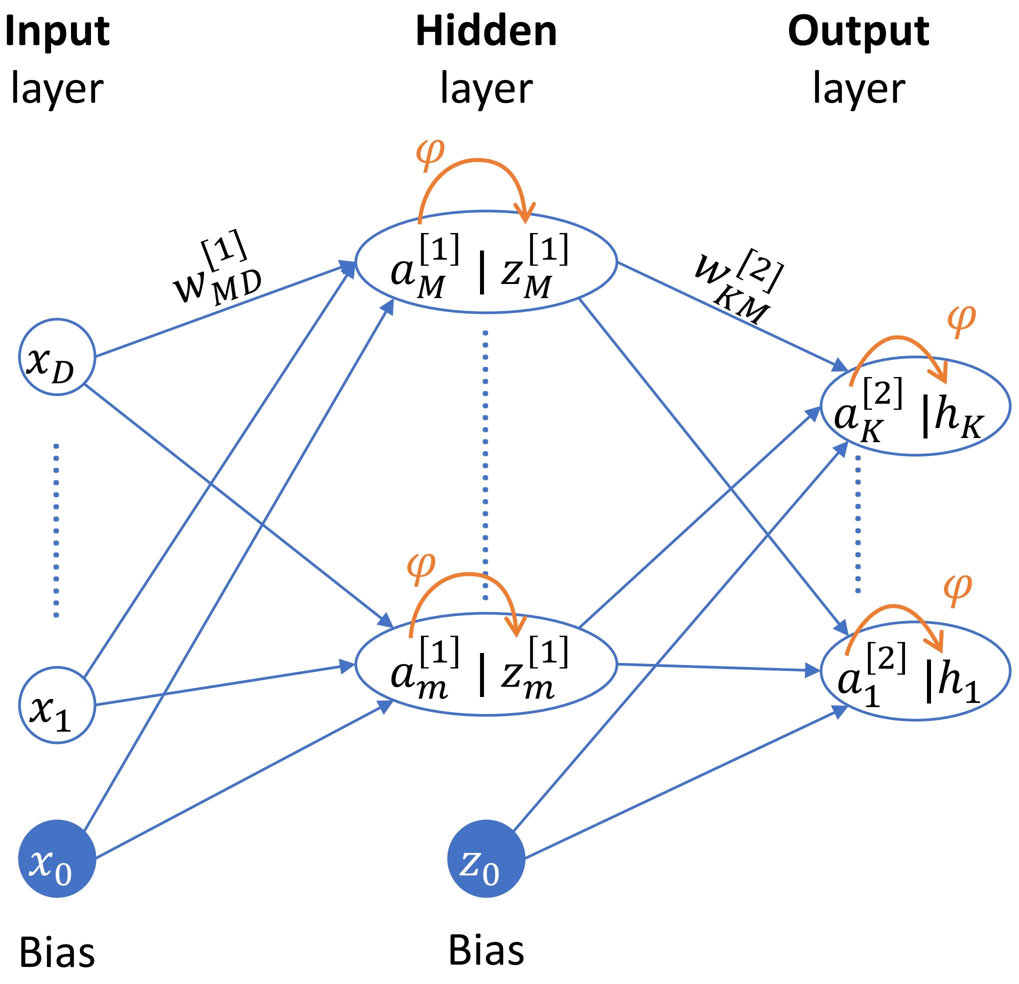
NEURAL NETWORKS REVIEW



- $h_k = \varphi\left(a_k^{[2]}\right)$, where $\varphi(\cdot)$ is an **activation function**

  $\varphi(a) = \sigma(a) = \frac{1}{1+\exp(-a)}$ logistic function

  $\varphi(a) = \tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$ hyperbolic tangent

  $\varphi(a) = \max\{0, a\}$ Rectified Linear Unit (ReLU)

- $a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m$

- $z_m = \varphi\left(a_m^{[1]}\right)$

- $a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$

EXERCICE 1

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(a) Draw the network from left (input) to right (output). Put also the units that represent the bias.

Input layer

Hidden layer

Output layer

EXERCICE 1

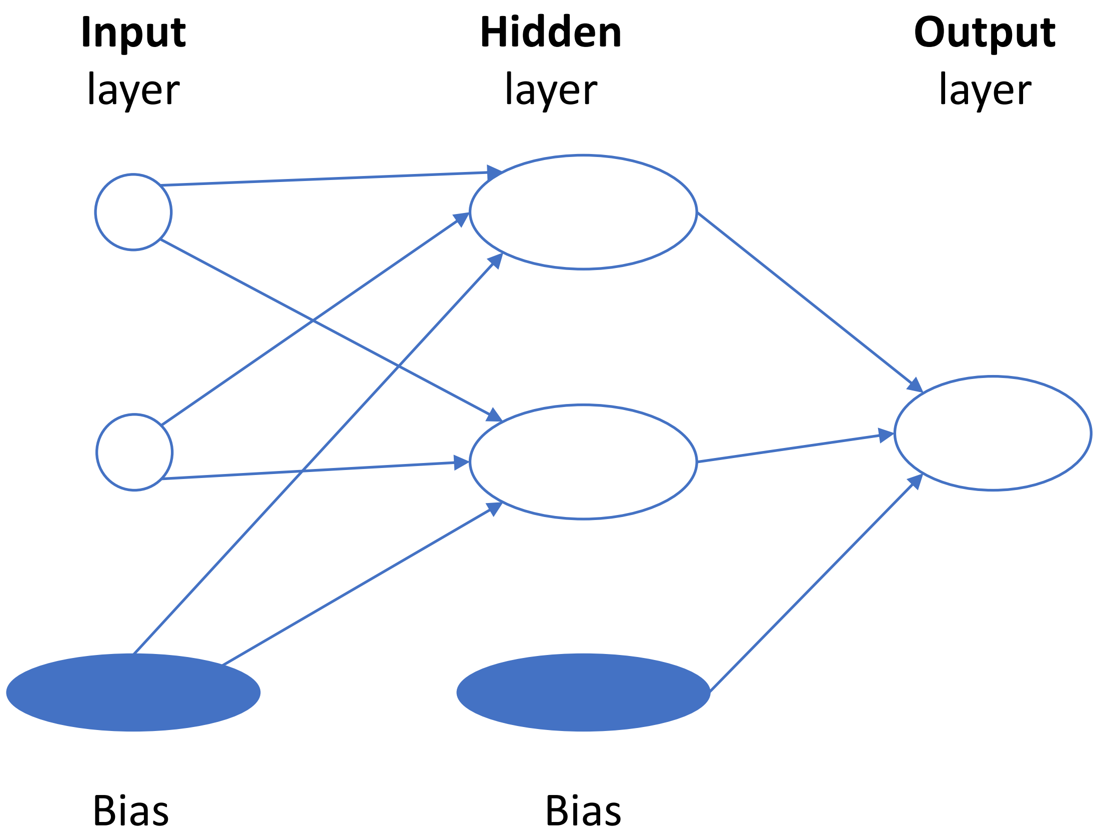
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(a) Draw the network from left (input) to right (output). Put also the units that represent the bias.



EXERCICE 1

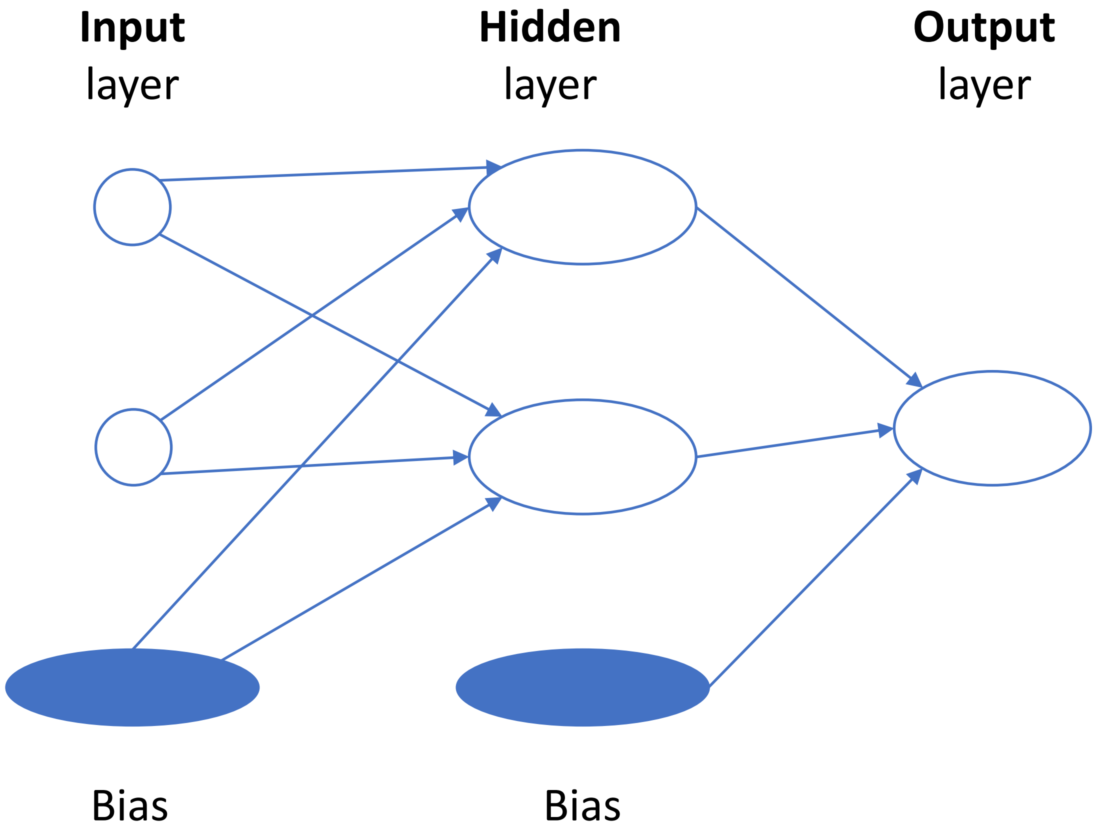
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(b) Put the name at each element of the graph and write, as a closed formula, the output in function of all the elements of the network.



EXERCICE 1

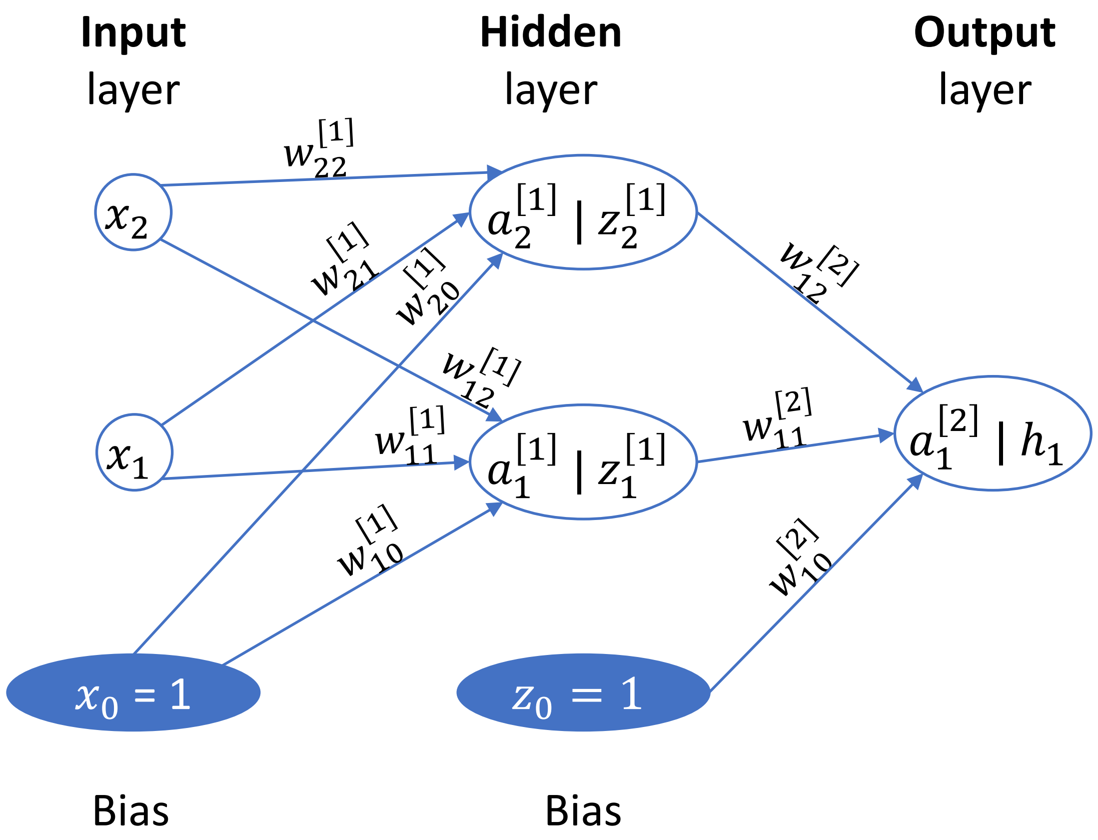
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(b) Put the name at each element of the graph and write, as a closed formula, the output in function of all the elements of the network.



EXERCICE 1

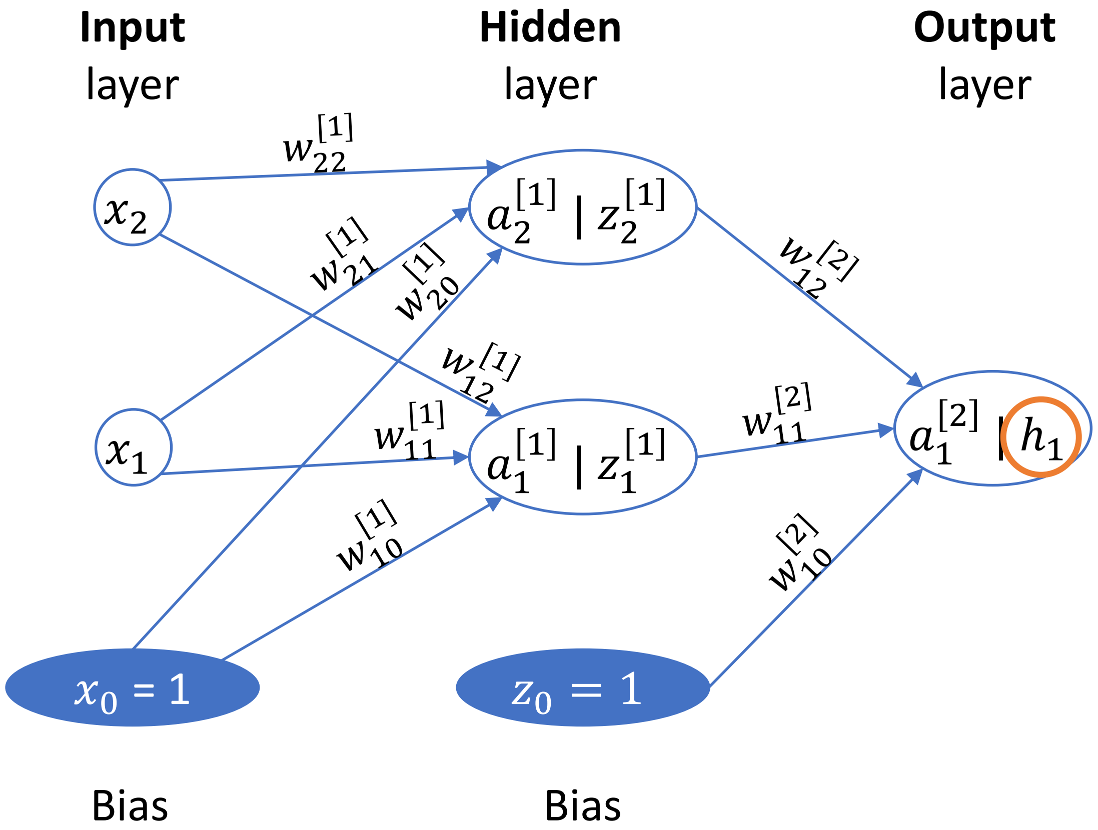
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(b) Put the name at each element of the graph and <u>write, as a closed formula, the output in function of all the elements of the network</u>.



$h_1 =$ $\boxed{?}$

EXERCICE 1

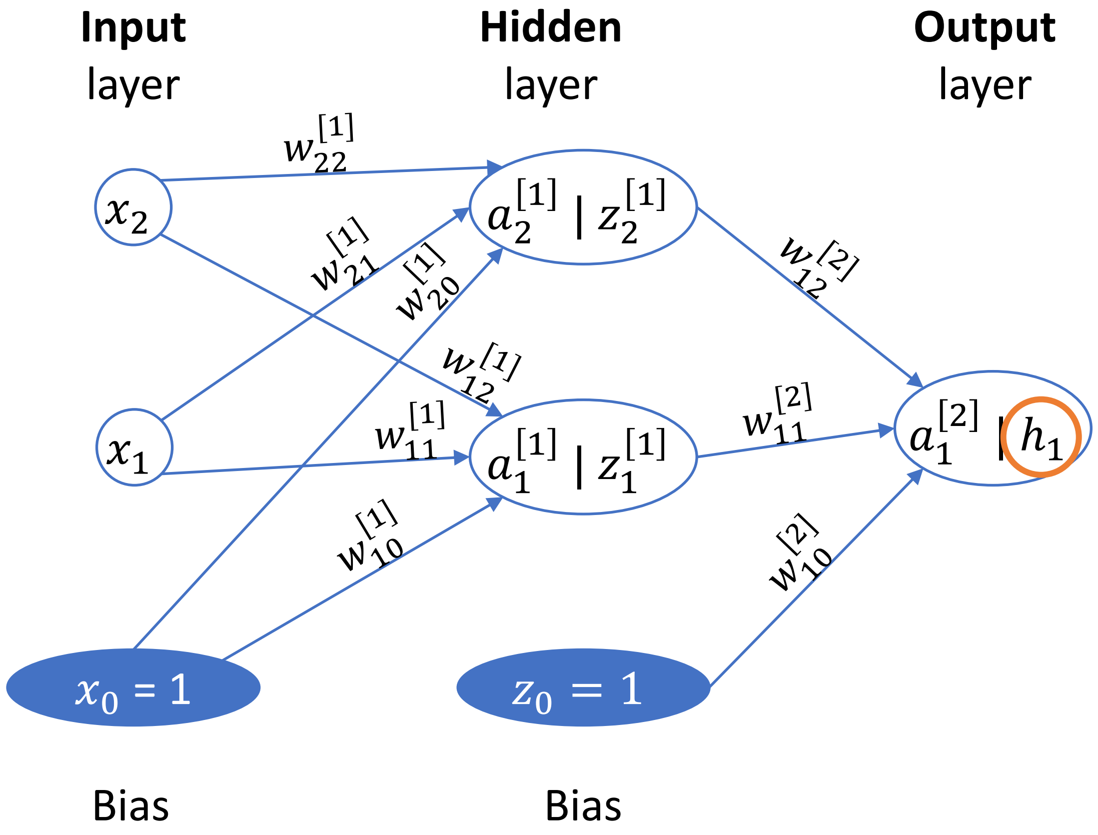
We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(b) Put the name at each element of the graph and write, as a closed formula, the output in function of all the elements of the network.



$$h_1 = \boxed{\varphi\left(a_1^{[2]}\right)}$$

Activation function

EXERCICE 1

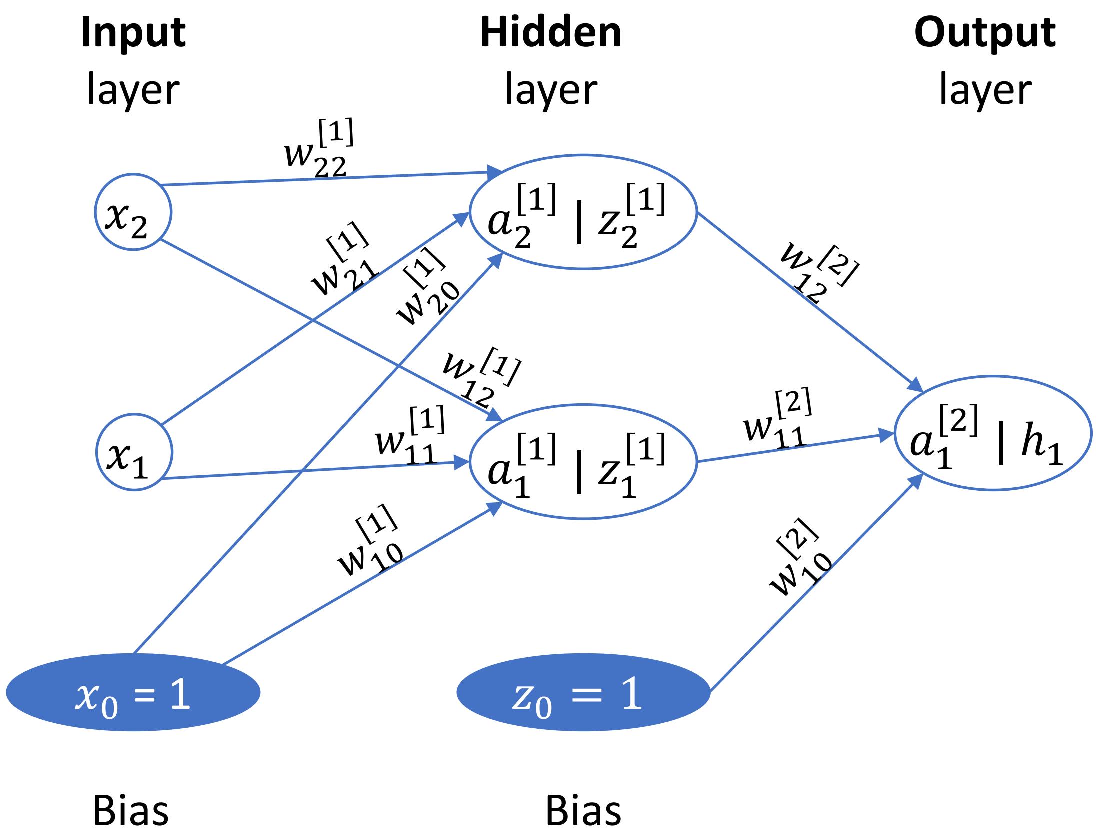
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(b) Put the name at each element of the graph and <u>write, as a closed formula, the output in function of all the elements of the network</u>.



$$h_1 = \varphi\left(a_1^{[2]}\right)$$

$$\| $$

$$?$$

EXERCICE 1

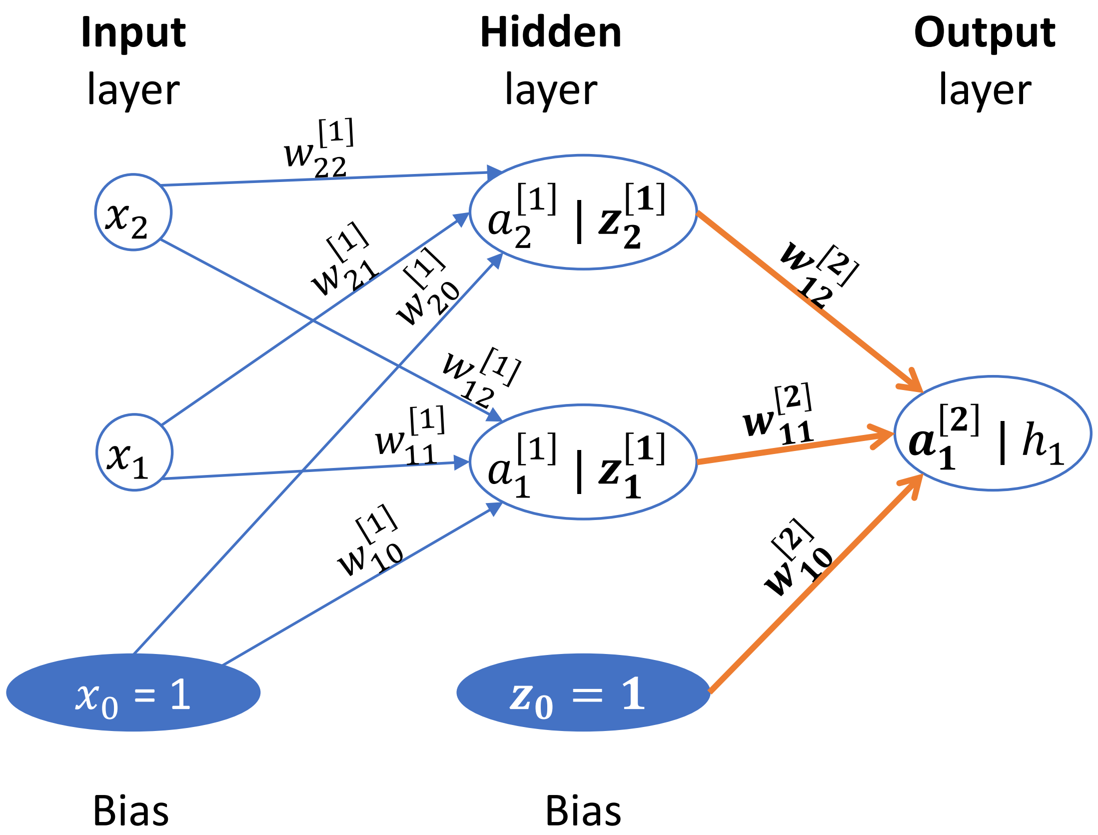
We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(b) Put the name at each element of the graph and <u>write, as a closed formula, the output in function of all the elements of the network</u>.



$$h_1 = \varphi\left(a_1^{[2]}\right)$$

$$a_1^{[2]} = \sum_{m=0}^{2} w_{1m}^{[2]} z_m^{[1]} = w_{10}^{[2]} z_0^{[1]} + w_{11}^{[2]} z_1^{[1]} + w_{12}^{[2]} z_2^{[1]}$$

EXERCICE 1

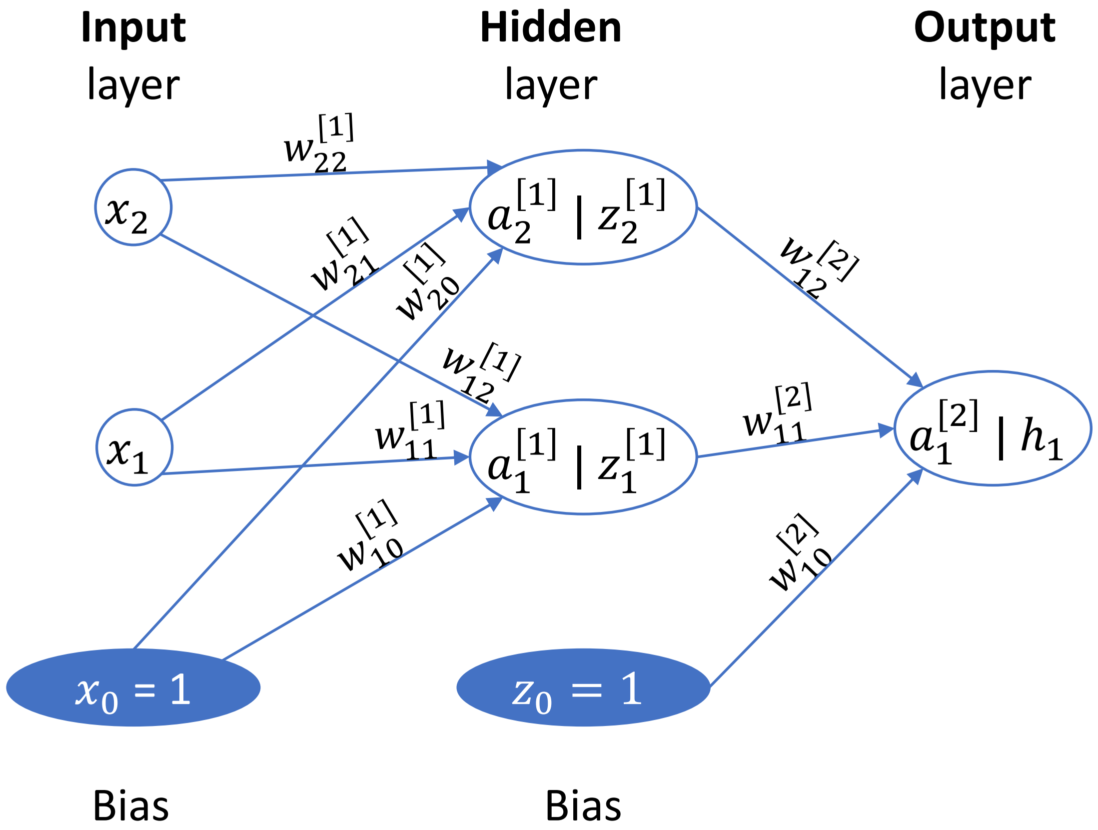
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(b) Put the name at each element of the graph and <u>write, as a closed formula, the output in function of all the elements of the network</u>.



$$h_1 = \varphi\left(a_1^{[2]}\right)$$

$$= \varphi\left(w_{10}^{[2]} \underset{?}{\underset{\|}{z_0^{[1]}}} + w_{11}^{[2]} \underset{?}{\underset{\|}{z_1^{[1]}}} + w_{12}^{[2]} \underset{?}{\underset{\|}{z_2^{[1]}}}\right)$$

EXERCICE 1

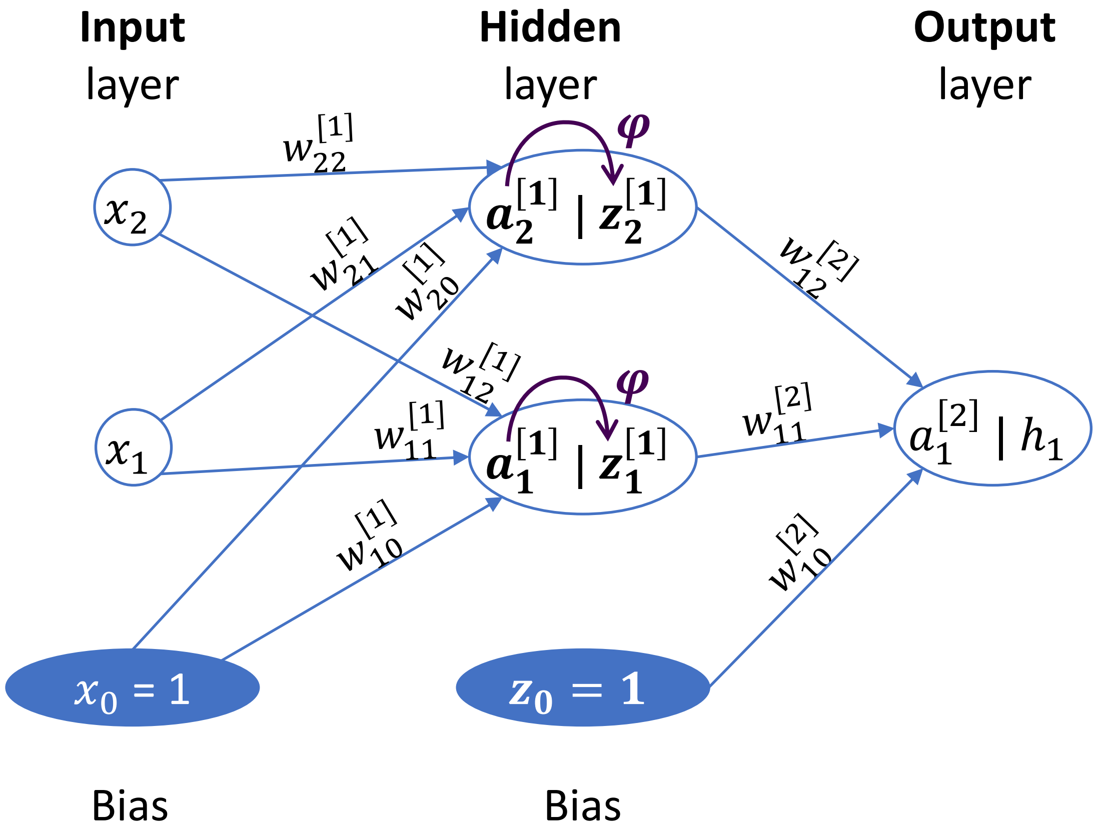
We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(b) Put the name at each element of the graph and <u>write, as a closed formula, the output in function of all the elements of the network</u>.



$$h_1 = \varphi\left(a_1^{[2]}\right)$$

$$= \varphi\left(w_{10}^{[2]} z_0^{[1]} + w_{11}^{[2]} z_1^{[1]} + w_{12}^{[2]} z_2^{[2]}\right)$$

$$z_0^{[1]} = \mathbf{1} \qquad z_1^{[1]} = \varphi\left(a_1^{[1]}\right) \qquad z_2^{[1]} = \varphi\left(a_2^{[1]}\right)$$

EXERCICE 1

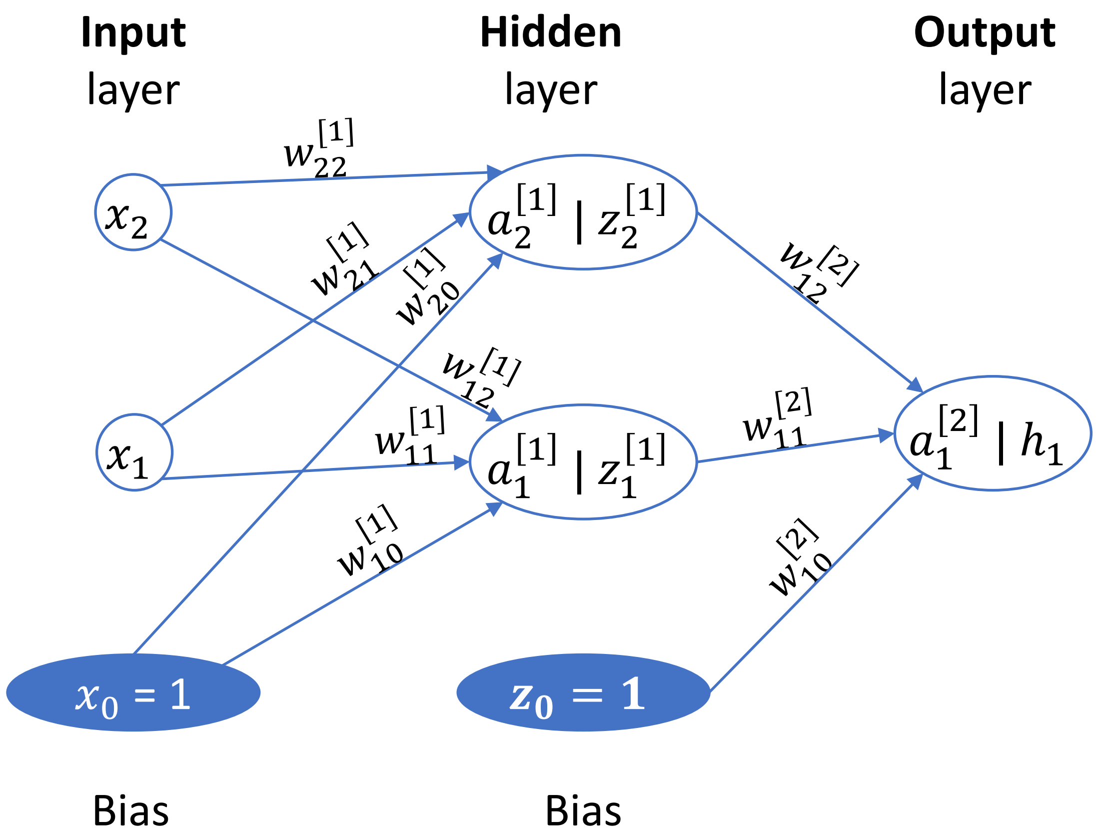
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(b) Put the name at each element of the graph and <u>write, as a closed formula, the output in function of all the elements of the network</u>.



$$\begin{aligned} h_1 &= \varphi\left(a_1^{[2]}\right) \\ &= \varphi\left(w_{10}^{[2]} z_0^{[1]} + w_{11}^{[2]} z_1^{[1]} + w_{12}^{[2]} z_2^{[1]}\right) \\ &= \varphi\left(w_{10}^{[2]} 1 + w_{11}^{[2]} \varphi\left(\underset{?}{\underset{\|}{a_1^{[1]}}}\right) + w_{12}^{[2]} \varphi\left(\underset{?}{\underset{\|}{a_2^{[1]}}}\right)\right) \end{aligned}$$

EXERCICE 1

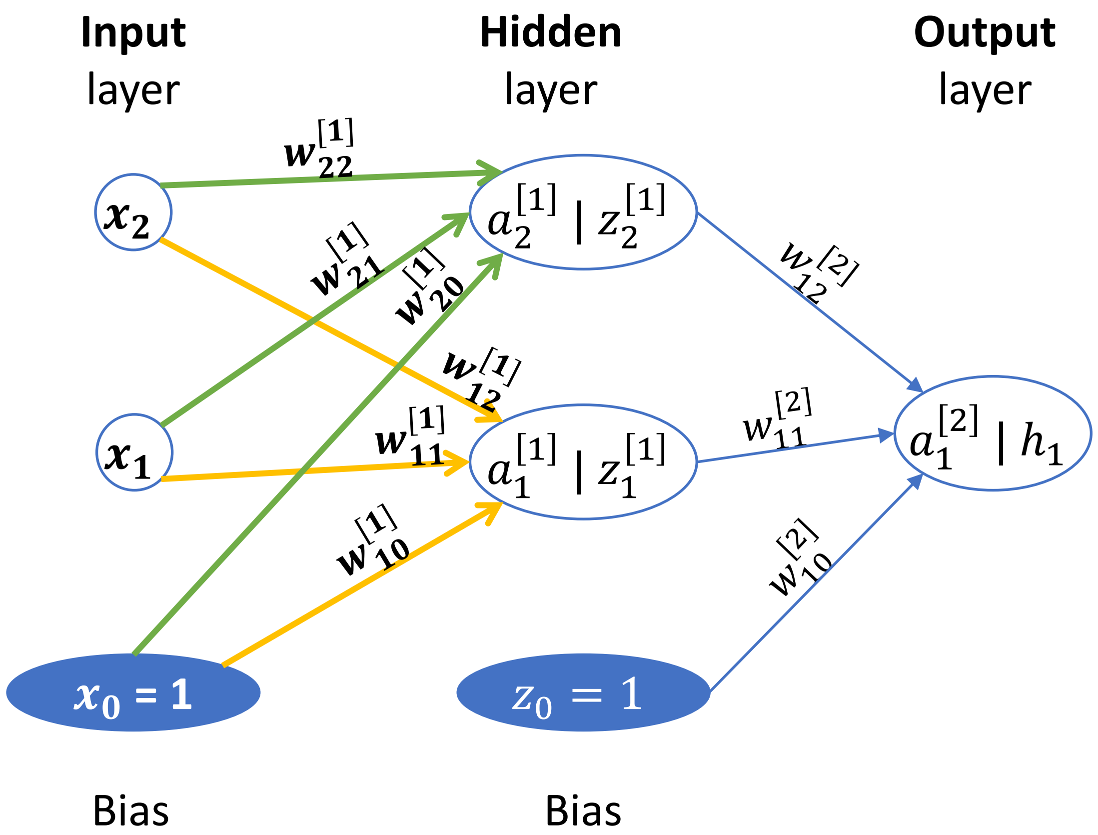
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(b) Put the name at each element of the graph and write, as a closed formula, the output in function of all the elements of the network.



$$h_1 = \varphi\left(a_1^{[2]}\right)$$

$$= \varphi\left(w_{10}^{[2]} z_0^{[1]} + w_{11}^{[2]} z_1^{[1]} + w_{12}^{[2]} z_2^{[1]}\right)$$

$$= \varphi\left(w_{10}^{[2]} 1 + w_{11}^{[2]} \varphi\left(a_1^{[1]}\right) + w_{12}^{[2]} \varphi\left(a_2^{[1]}\right)\right)$$

$$a_1^{[1]} = \sum_{d=0}^{D} w_{1d}^{[1]} x_d = w_{10}^{[1]} x_0 + w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2$$

$$a_2^{[1]} = \sum_{d=0}^{D} w_{2d}^{[1]} x_d = w_{20}^{[1]} x_0 + w_{21}^{[1]} x_1 + w_{22}^{[1]} x_2$$

EXERCICE 1

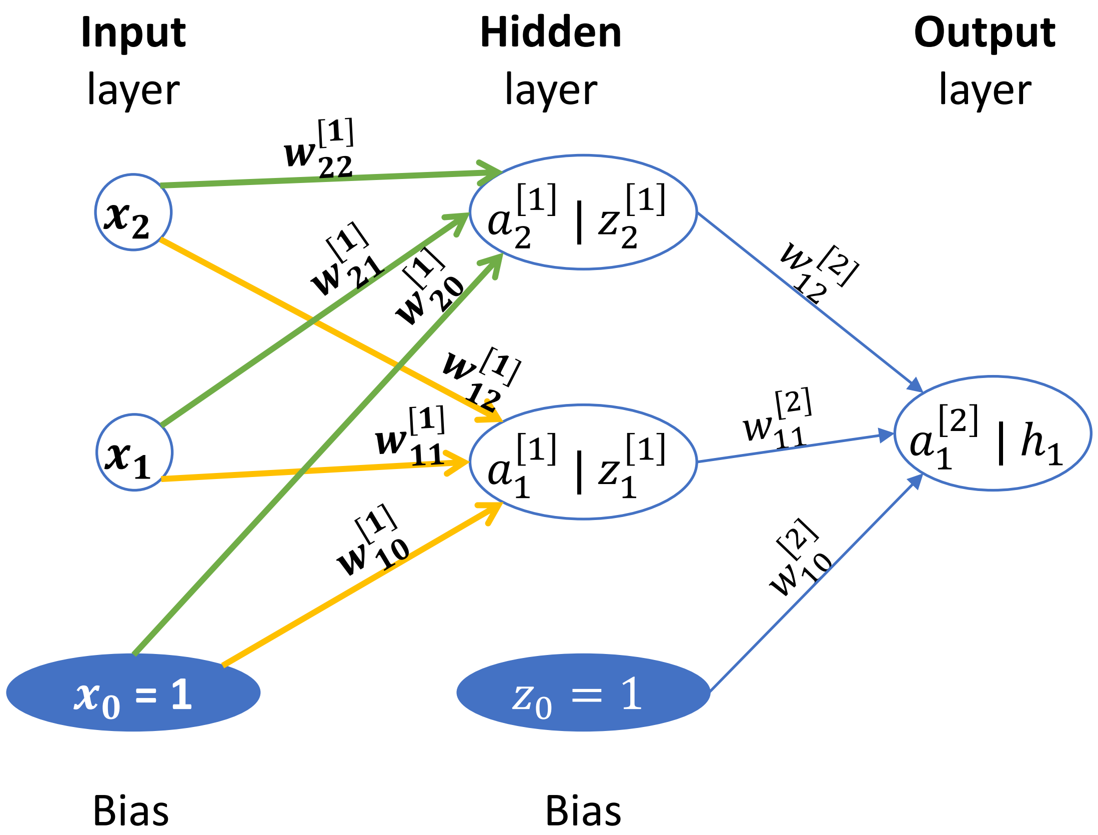
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(b) Put the name at each element of the graph and write, as a closed formula, the output in function of all the elements of the network.



$$h_1 = \varphi\left(a_1^{[2]}\right)$$

$$= \varphi\left(w_{10}^{[2]} z_0^{[1]} + w_{11}^{[2]} z_1^{[1]} + w_{12}^{[2]} z_2^{[1]}\right)$$

$$= \varphi\left(w_{10}^{[2]} 1 + w_{11}^{[2]} \varphi\left(a_1^{[1]}\right) + w_{12}^{[2]} \varphi\left(a_2^{[1]}\right)\right)$$

$$= \varphi\left(\begin{array}{l} w_{10}^{[2]} + w_{11}^{[2]} \varphi\left(\boxed{w_{10}^{[1]} \underbrace{x_0}_{=\,1} + w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2}\right) \\ \qquad + w_{12}^{[2]} \varphi\left(\boxed{w_{20}^{[1]} \underbrace{x_0}_{=\,1} + w_{21}^{[1]} x_1 + w_{22}^{[1]} x_2}\right) \end{array}\right)$$

EXERCICE 1

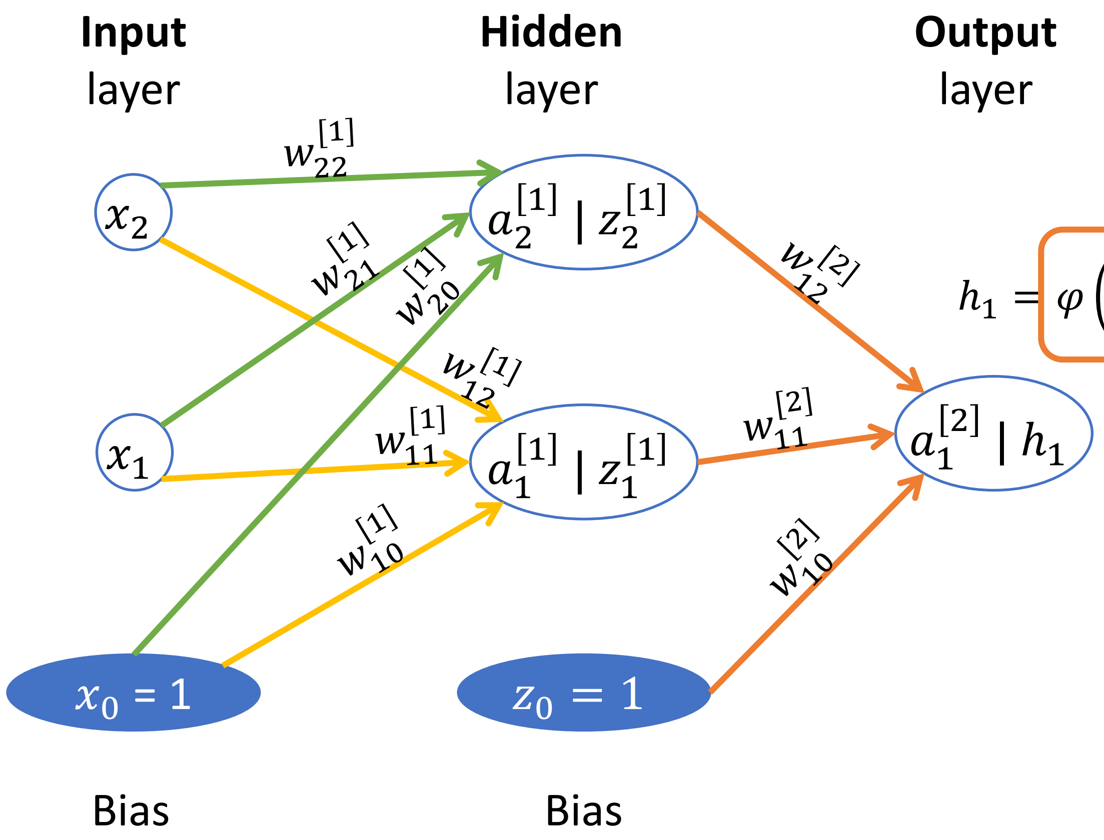
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

(b) Put the name at each element of the graph and <u>write, as a closed formula, the output in function of all the elements of the network</u>.



$$h_1 = \boxed{\varphi\left(\underbrace{w_{10}^{[2]} + w_{11}^{[2]} \boxed{\varphi\left(\overbrace{w_{10}^{[1]} + w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2}^{?}\right)} + w_{12}^{[2]} \boxed{\varphi\left(\overbrace{w_{20}^{[1]} + w_{21}^{[1]} x_1 + w_{22}^{[1]} x_2}^{?}\right)}}_{?}\right)}$$

EXERCICE 1

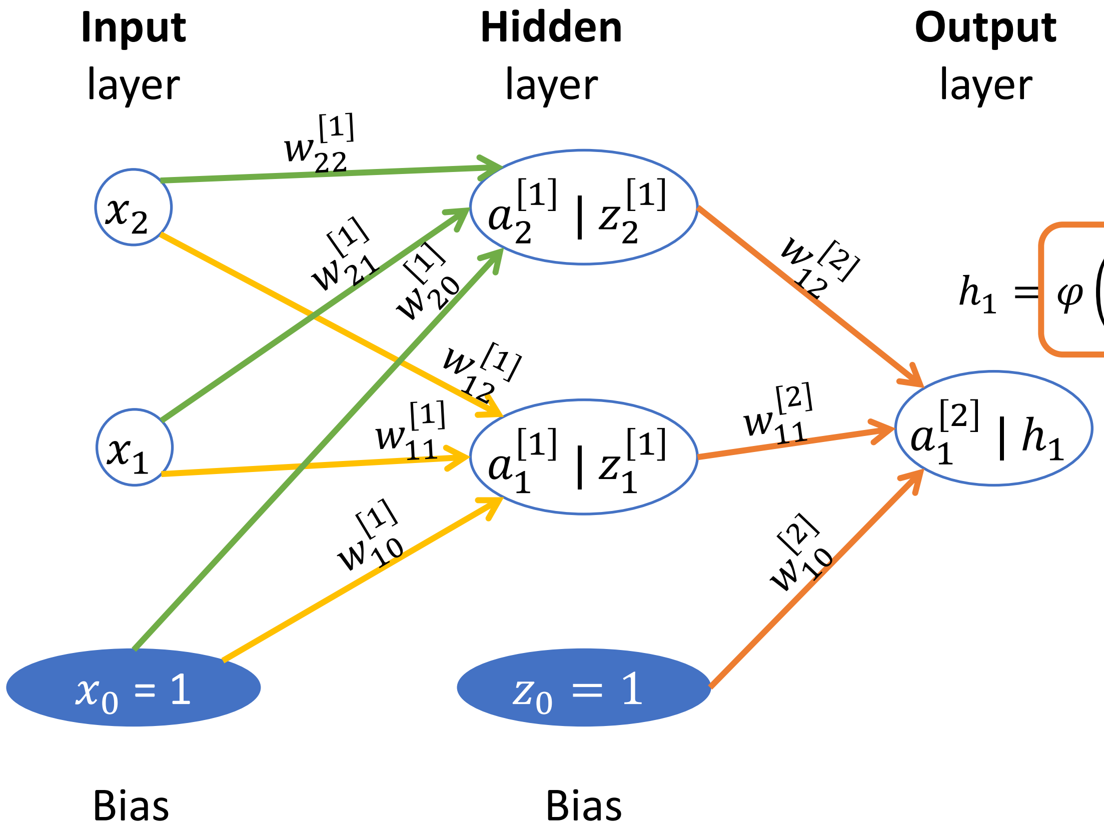
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

We have a fully connected network with the following characteristics:

- It takes 2D vectors as input.
- It has a hidden layer with two neurons.
- The output layer has one neuron.

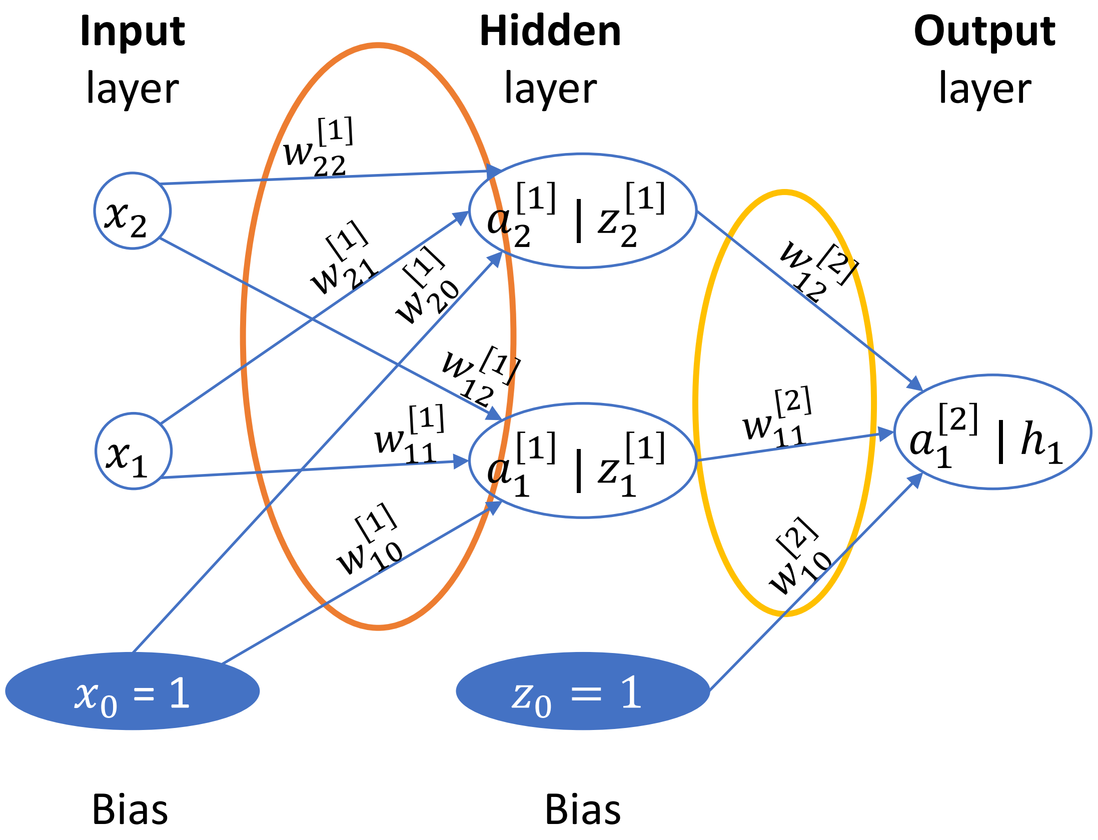
(b) Put the name at each element of the graph and <u>write, as a closed formula, the output in function of all the elements of the network</u>.



$$h_1 = \boxed{\varphi\left(\underbrace{w_{10}^{[2]} + w_{11}^{[2]}\,\boxed{\varphi\left(\overbrace{w_{10}^{[1]} + w_{11}^{[1]}x_1 + w_{12}^{[1]}x_2}^{\boldsymbol{a_1^{[1]}}}\right)} + w_{12}^{[2]}\,\boxed{\varphi\left(\overbrace{w_{20}^{[1]} + w_{21}^{[1]}x_1 + w_{22}^{[1]}x_2}^{\boldsymbol{a_2^{[1]}}}\right)}}_{\boldsymbol{a_1^{[2]}}}\right)}$$

EXERCICE 1

(c) Initialize the weights of the 1st layer to 0.1 and of the 2nd layer to 0.2. Take the point $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1$. Calculate $h\left(\mathbf{x}^{(1)}\right)$ when all the activation function are sigmoids except in the output layer, where $h(a)=a$.



NEW INFORMATION

$\boxed{w_{md}^{[1]}=0.1}\ \forall m,d$

$\boxed{w_{1m}^{[2]}=0.2}\ \forall m$

$$h_1=\varphi\left(a_1^{[2]}\right)=a_1^{[2]}$$

$$z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)=\frac{1}{1+\exp\left(-a_m^{[1]}\right)}$$

with $\sigma'(a)=\sigma(a)(1-\sigma(a))$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

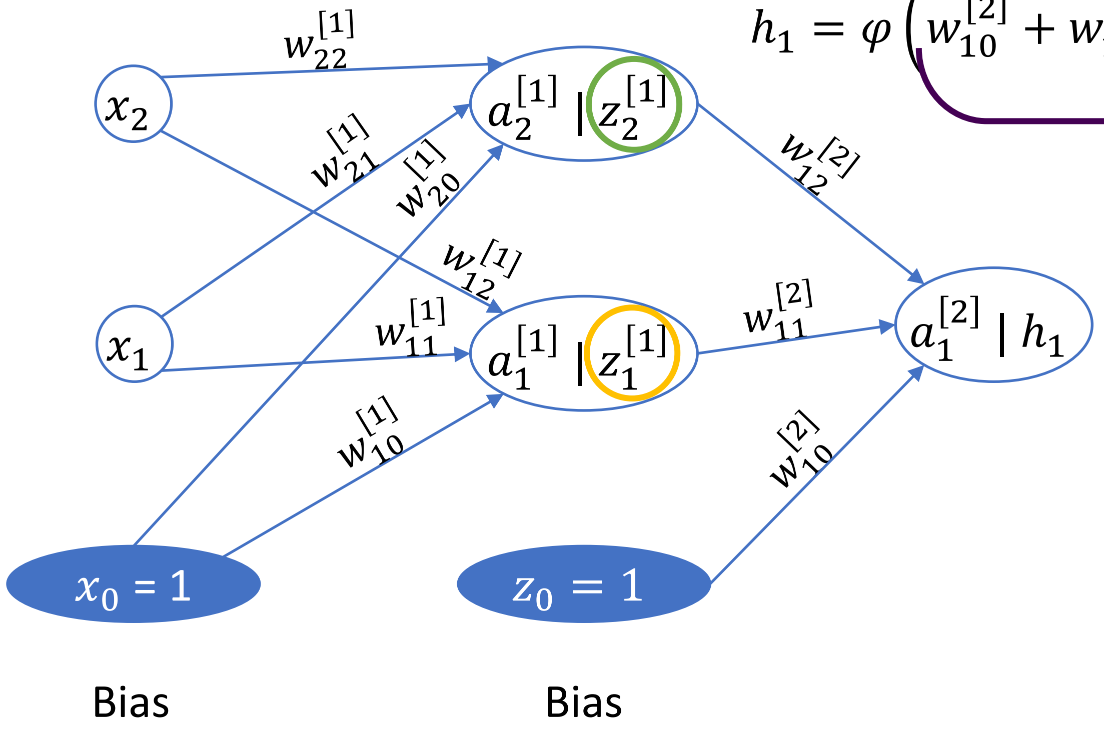
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = ?$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

REMEMBER FROM (b)



$$h_1 = \varphi\Bigg(\underbrace{w_{10}^{[2]} + w_{11}^{[2]}\,\overset{z_1^{[1]}\,\Vert}{\boxed{\varphi\Big(\overbrace{w_{10}^{[1]} + w_{11}^{[1]}x_1 + w_{12}^{[1]}x_2}^{a_1^{[1]}}\Big)}} + w_{12}^{[2]}\,\overset{z_2^{[1]}\,\Vert}{\boxed{\varphi\Big(\overbrace{w_{20}^{[1]} + w_{21}^{[1]}x_1 + w_{22}^{[1]}x_2}^{a_2^{[1]}}\Big)}}}_{a_1^{[2]}}\Bigg)$$

FORWARD PASS

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. Compute $a_1^{[2]}$.
4. Compute h_1

EXERCICE 1

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1 \longrightarrow h(\mathbf{x}^{(1)})=?$

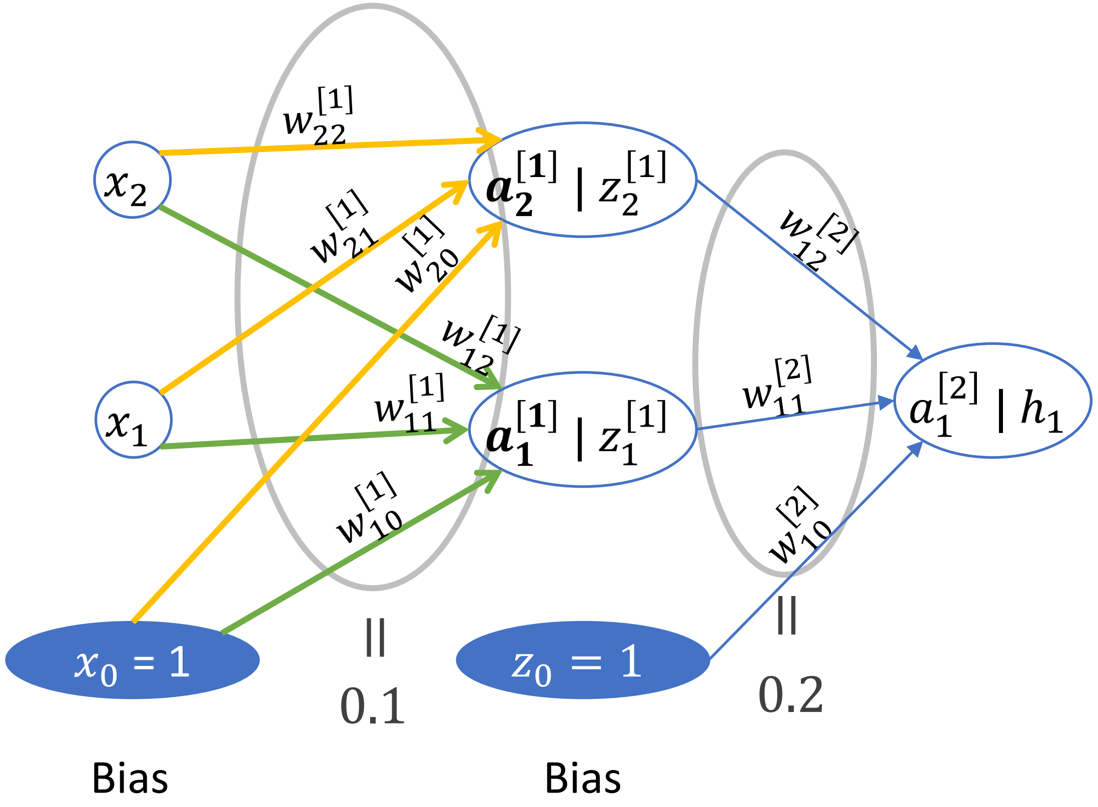
with $h_1\left(a_1^{[2]}\right)=a_1^{[2]}$ and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

$$\sigma(a)=\frac{1}{1+\exp(-a)}$$

$$\sigma'(a)=\sigma(a)(1-\sigma(a))$$

$$h_k=\varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]}=\sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]}=\varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]}=\sum_{d=0}^{D} w_{md}^{[1]} x_d$$



1. **Compute $a_1^{[1]}$ and $a_2^{[1]}$.**
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. Compute $a_1^{[2]}$.
4. Compute h_1

$$\boxed{a_1^{[1]}}=\sum_{d=0}^{2} w_{1d}^{[1]} x_d = w_{10}^{[1]} x_0 + w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2 = ?$$

$$\boxed{a_2^{[1]}}=\sum_{d=0}^{2} w_{2d}^{[1]} x_d = w_{20}^{[1]} x_0 + w_{21}^{[1]} x_1 + w_{22}^{[1]} x_2 = ?$$

EXERCICE 1

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1 \longrightarrow h(\mathbf{x}^{(1)})=?$

with $h_1\left(a_1^{[2]}\right)=a_1^{[2]}$ and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

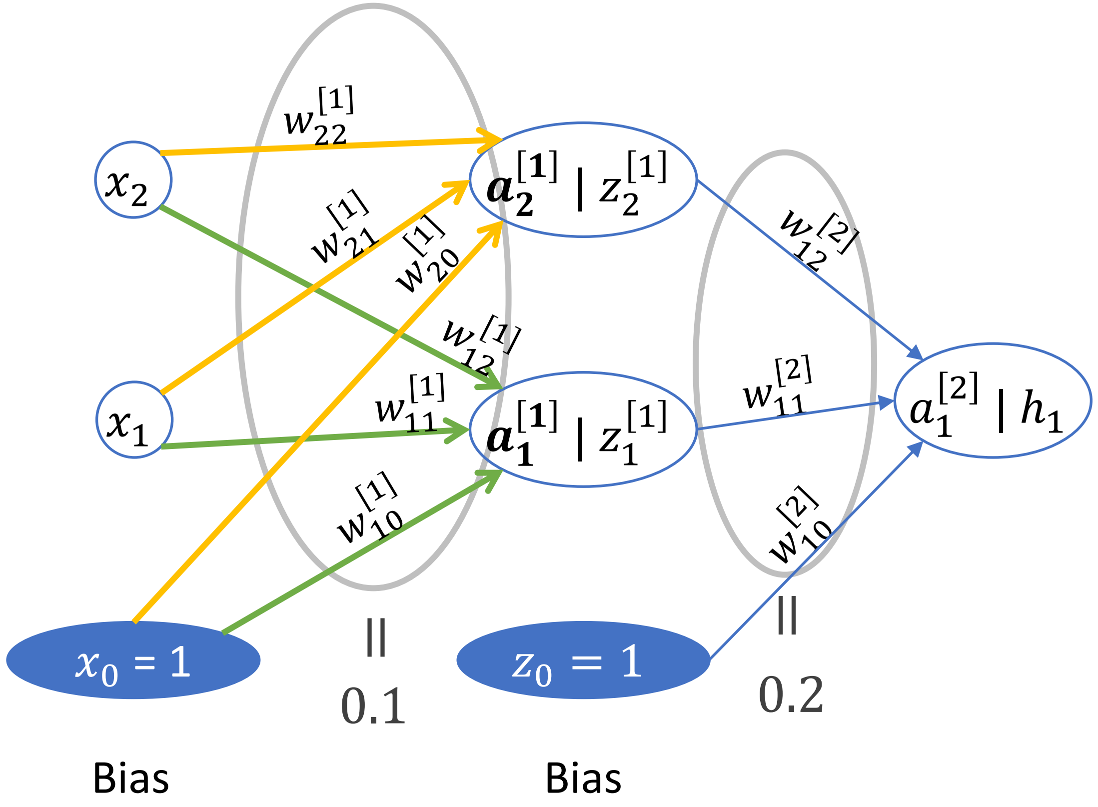
$$\sigma(a)=\frac{1}{1+\exp(-a)}$$

$$\sigma'(a)=\sigma(a)(1-\sigma(a))$$

$$h_k=\varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]}=\sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]}=\varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]}=\sum_{d=0}^{D} w_{md}^{[1]} x_d$$

1. **Compute $a_1^{[1]}$ and $a_2^{[1]}$.**
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. Compute $a_1^{[2]}$.
4. Compute h_1



$$\boxed{\boldsymbol{a_1^{[1]}}}=\sum_{d=0}^{2} w_{1d}^{[1]} x_d = w_{10}^{[1]} x_0 + w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2$$
$$=0.1\cdot 1+\ 0.1\cdot 0.3+\ 0.1\cdot 0.5=0.18$$

$$\boxed{\boldsymbol{a_2^{[1]}}}=\sum_{d=0}^{2} w_{2d}^{[1]} x_d = w_{20}^{[1]} x_0 + w_{21}^{[1]} x_1 + w_{22}^{[1]} x_2$$
$$=0.1\cdot 1+\ 0.1\cdot 0.3+\ 0.1\cdot 0.5=0.18$$

EXERCICE 1

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1 \longrightarrow h\left(\mathbf{x}^{(1)}\right)=?$

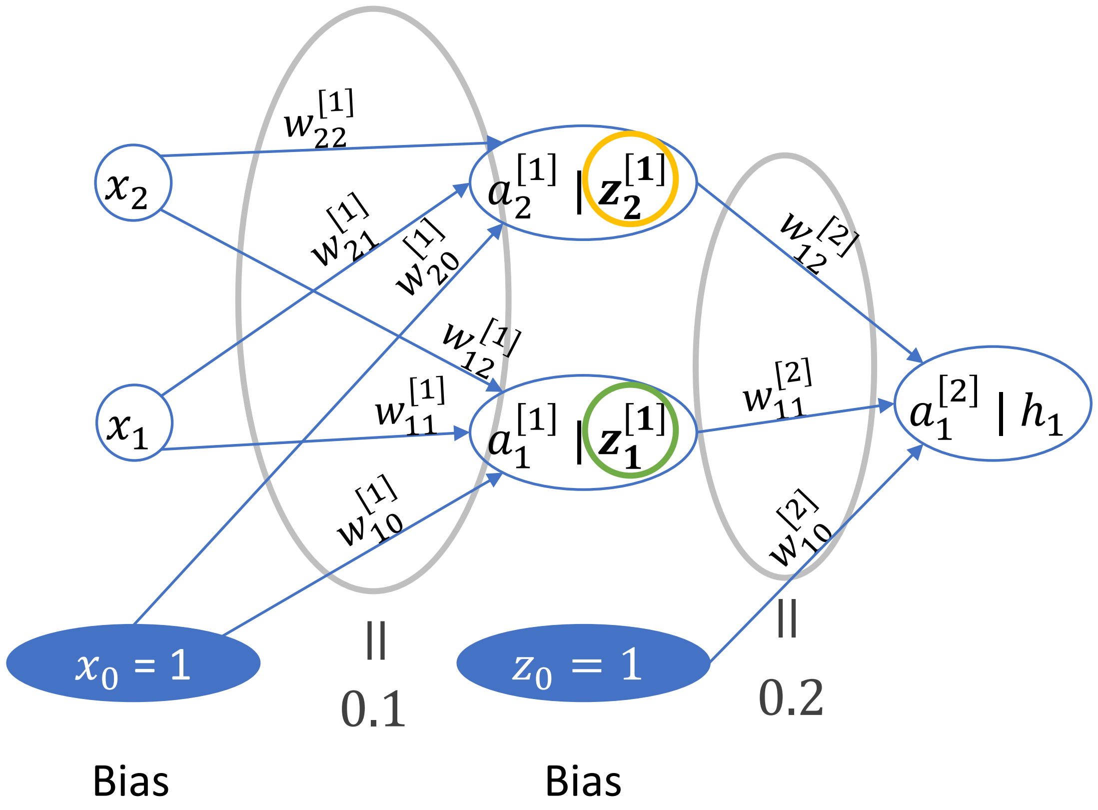
with $h_1\left(a_1^{[2]}\right)=a_1^{[2]}$ and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

$$\sigma(a)=\frac{1}{1+\exp(-a)}$$

$$\sigma'(a)=\sigma(a)(1-\sigma(a))$$

$$h_k=\varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]}=\sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]}=\varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]}=\sum_{d=0}^{D} w_{md}^{[1]} x_d$$



1. $a_1^{[1]}=0.18,\ a_2^{[1]}=0.18$

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. **Compute $z_1^{[1]}$ and $z_2^{[1]}$.**
3. Compute $a_1^{[2]}$.
4. Compute h_1

$\boxed{z_1^{[1]}}=\varphi\left(a_1^{[1]}\right)=\sigma\left(a_1^{[1]}\right)=?$

$\boxed{z_2^{[1]}}=\varphi\left(a_2^{[1]}\right)=\sigma\left(a_2^{[1]}\right)=?$

EXERCICE 1

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h(\mathbf{x}^{(1)}) = ?$

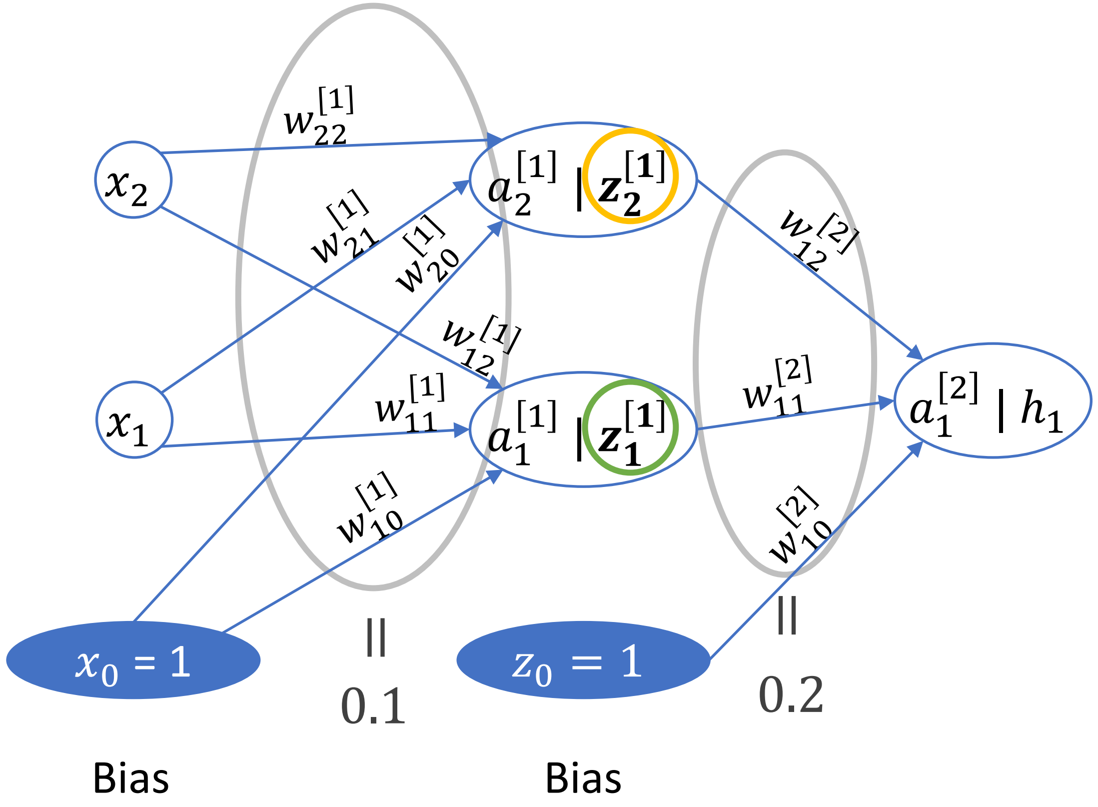
with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$



1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. **Compute $\mathbf{z_1^{[1]}}$ and $\mathbf{z_2^{[1]}}$.**
3. Compute $a_1^{[2]}$.
4. Compute h_1

$$\boxed{\mathbf{z_1^{[1]}}} = \varphi\left(a_1^{[1]}\right) = \sigma\left(a_1^{[1]}\right)$$
$$= \sigma(0.18) = \frac{1}{1+\exp(-0.18)} = 0.54$$

$$\boxed{\mathbf{z_2^{[1]}}} = \varphi\left(a_2^{[1]}\right) = \sigma\left(a_2^{[1]}\right)$$
$$= \sigma(0.18) = \frac{1}{1+\exp(-0.18)} = 0.54$$

EXERCICE 1

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = ?$

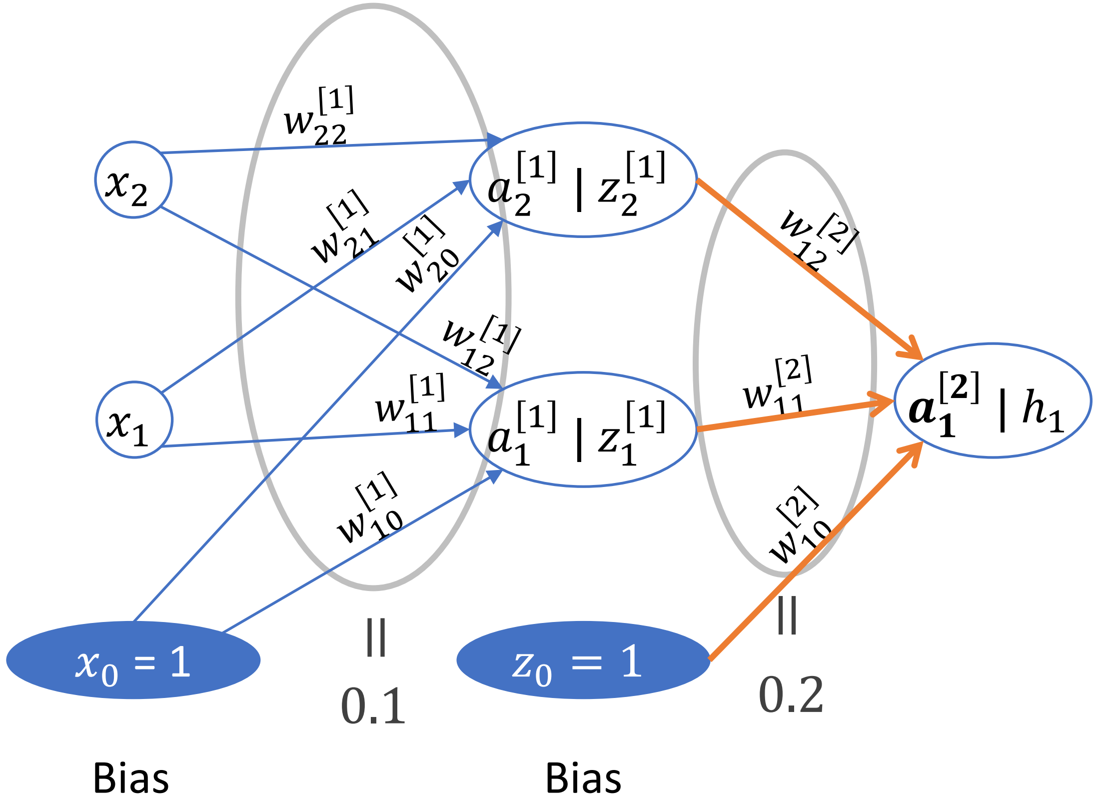
with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1 - \sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$



1. $a_1^{[1]} = 0.18,\ a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54,\ z_2^{[1]} = 0.54$

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. **Compute $\boldsymbol{a_1^{[2]}}$.**
4. Compute h_1

$$\boxed{\boldsymbol{a_1^{[2]}}} = \sum_{m=0}^{2} w_{1m}^{[2]} z_m^{[1]} = w_{10}^{[2]} z_0^{[1]} + w_{11}^{[2]} z_1^{[1]} + w_{12}^{[2]} z_2^{[1]} = ?$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

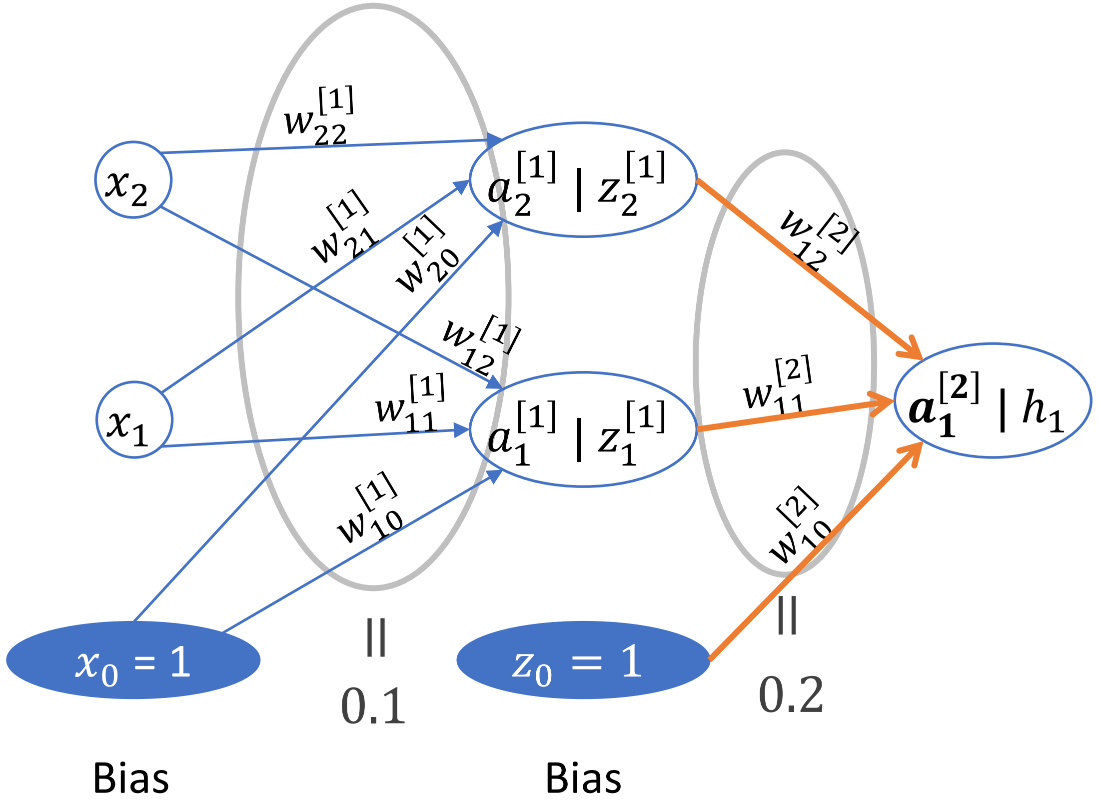
$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = ?$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$



1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54, z_2^{[1]} = 0.54$

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. **Compute $\boldsymbol{a_1^{[2]}}$.**
4. Compute h_1

$$\boxed{a_1^{[2]}} = \sum_{m=0}^{2} w_{1m}^{[2]} z_m^{[1]} = w_{10}^{[2]} z_0^{[1]} + w_{11}^{[2]} z_1^{[1]} + w_{12}^{[2]} z_2^{[1]}$$

$$= 0.2 \cdot 1 + 0.2 \cdot 0.54 + 0.2 \cdot 0.54 = 0.42$$

EXERCICE 1

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1 \longrightarrow h\left(\mathbf{x}^{(1)}\right)=?$

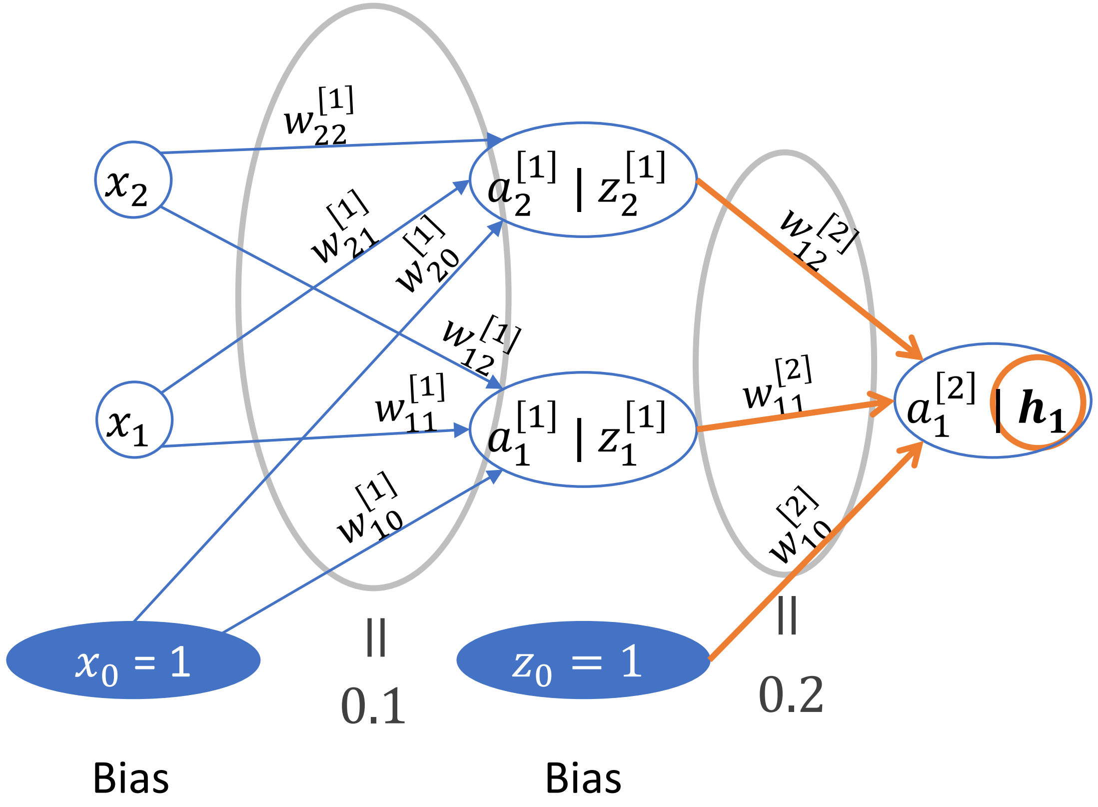
with $h_1\left(a_1^{[2]}\right)=a_1^{[2]}$ and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

$$\sigma(a)=\frac{1}{1+\exp(-a)}$$

$$\sigma'(a)=\sigma(a)(1-\sigma(a))$$

$$h_k=\varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]}=\sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]}=\varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]}=\sum_{d=0}^{D} w_{md}^{[1]} x_d$$



1. $a_1^{[1]}=0.18$, $a_2^{[1]}=0.18$
2. $z_1^{[1]}=0.54$, $z_2^{[1]}=0.54$
3. $a_1^{[2]}=0.42$

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. Compute $a_1^{[2]}$.
4. **Compute h_1**

$$\boxed{h_1}=\varphi\left(a_1^{[2]}\right)=h_1\left(a_1^{[2]}\right)=?$$

EXERCICE 1

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1 \longrightarrow h\left(\mathbf{x}^{(1)}\right)=?$

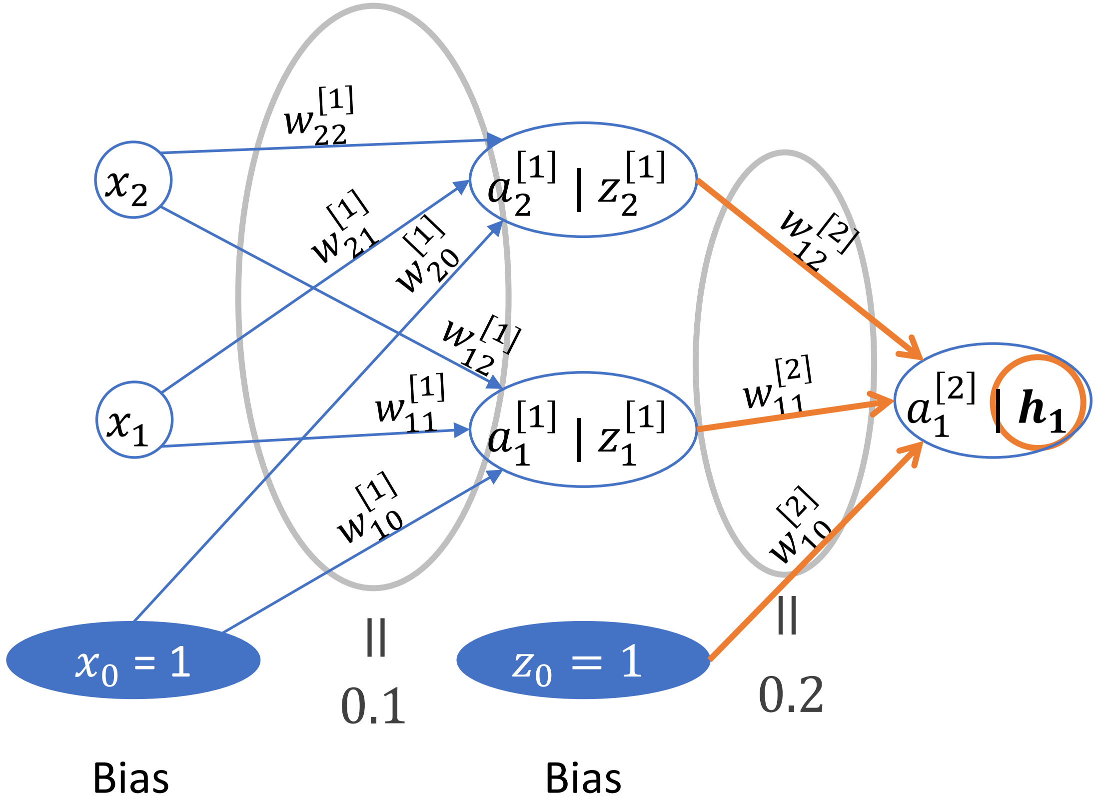
with $h_1\left(a_1^{[2]}\right)=a_1^{[2]}$ and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

$$\sigma(a)=\frac{1}{1+\exp(-a)}$$

$$\sigma'(a)=\sigma(a)(1-\sigma(a))$$

$$h_k=\varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]}=\sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]}=\varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]}=\sum_{d=0}^{D} w_{md}^{[1]} x_d$$



1. $a_1^{[1]}=0.18$, $a_2^{[1]}=0.18$
2. $z_1^{[1]}=0.54$, $z_2^{[1]}=0.54$
3. $a_1^{[2]}=0.42$

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. Compute $a_1^{[2]}$.
4. **Compute $\boldsymbol{h_1}$**

$$\boxed{\boldsymbol{h_1}}=\varphi\left(a_1^{[2]}\right)=h_1\left(a_1^{[2]}\right)=$$
$$=a_1^{[2]}=0.42$$

EXERCICE 1

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = ?$

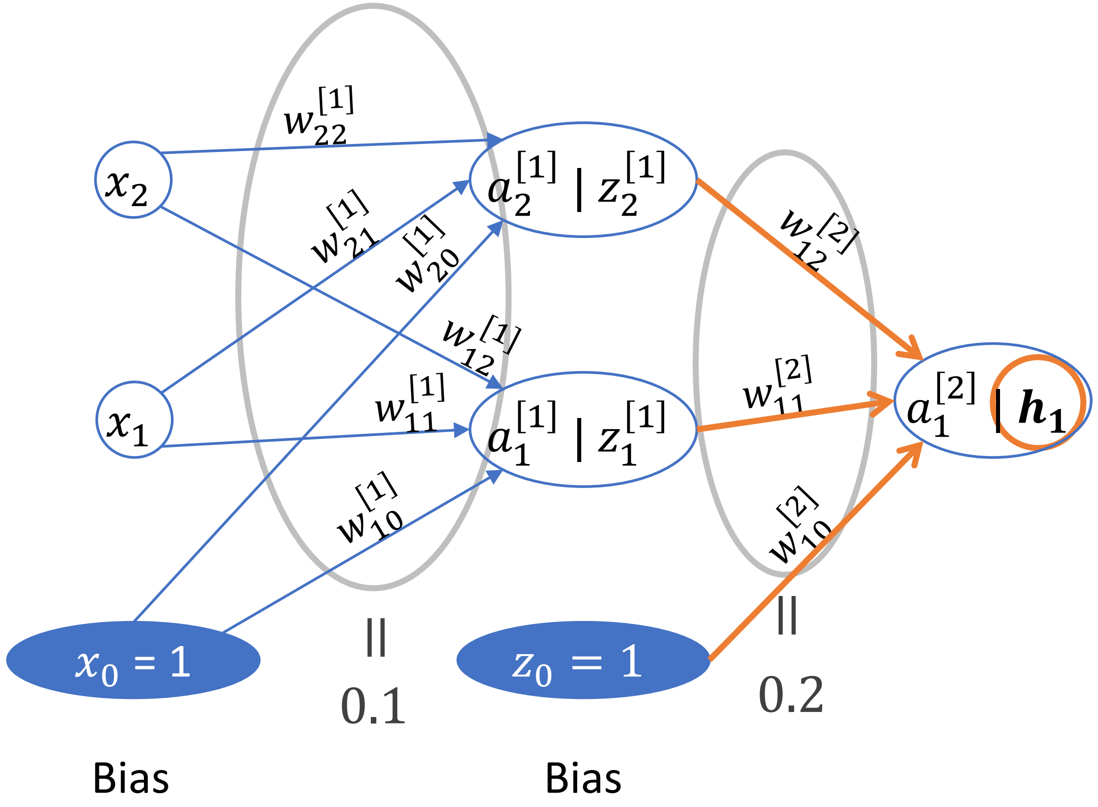
with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$



1. $a_1^{[1]} = 0.18$, $a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54$, $z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. Compute $a_1^{[2]}$.
4. **Compute $\boldsymbol{h_1}$**

$$\boxed{\boldsymbol{h_1}} = \varphi\left(a_1^{[2]}\right) = h_1\left(a_1^{[2]}\right) = $$
$$= a_1^{[2]} = 0.42$$

$$\boxed{\Longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42}$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

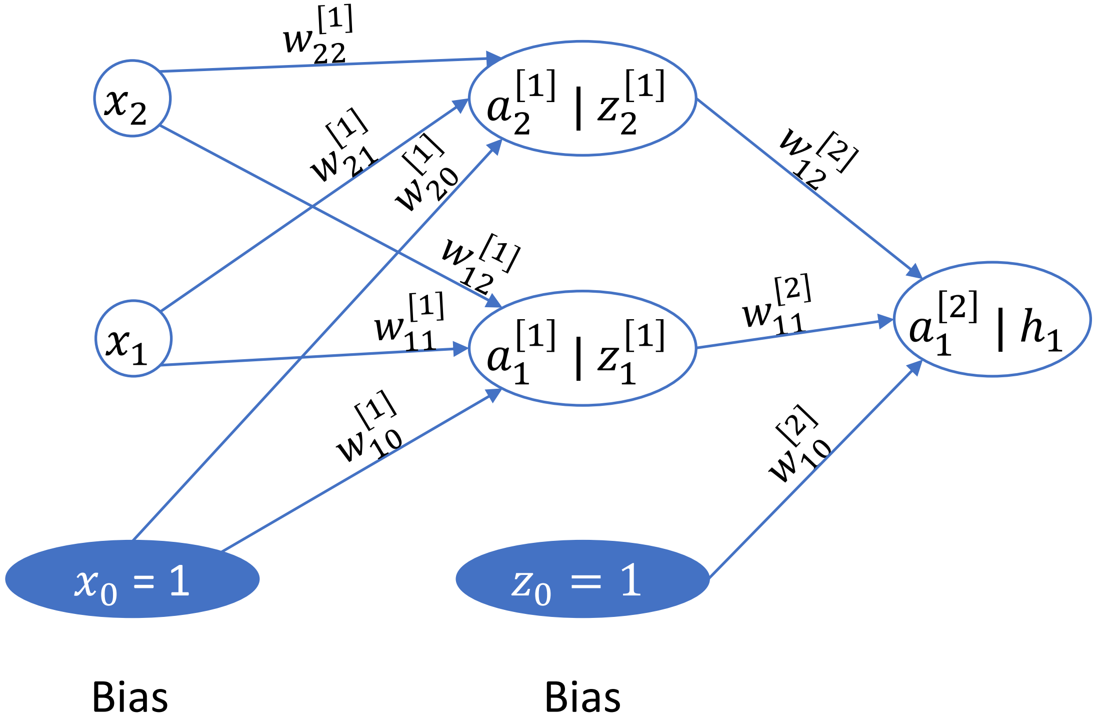
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow \boxed{h\left(\mathbf{x}^{(1)}\right) = 0.42}$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(i) Calculate the error using LS Error.



<u>LEAST SQUARES ERROR</u>

$$\mathbb{E}\left(\mathbf{x}^{(1)}\right) = ?$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

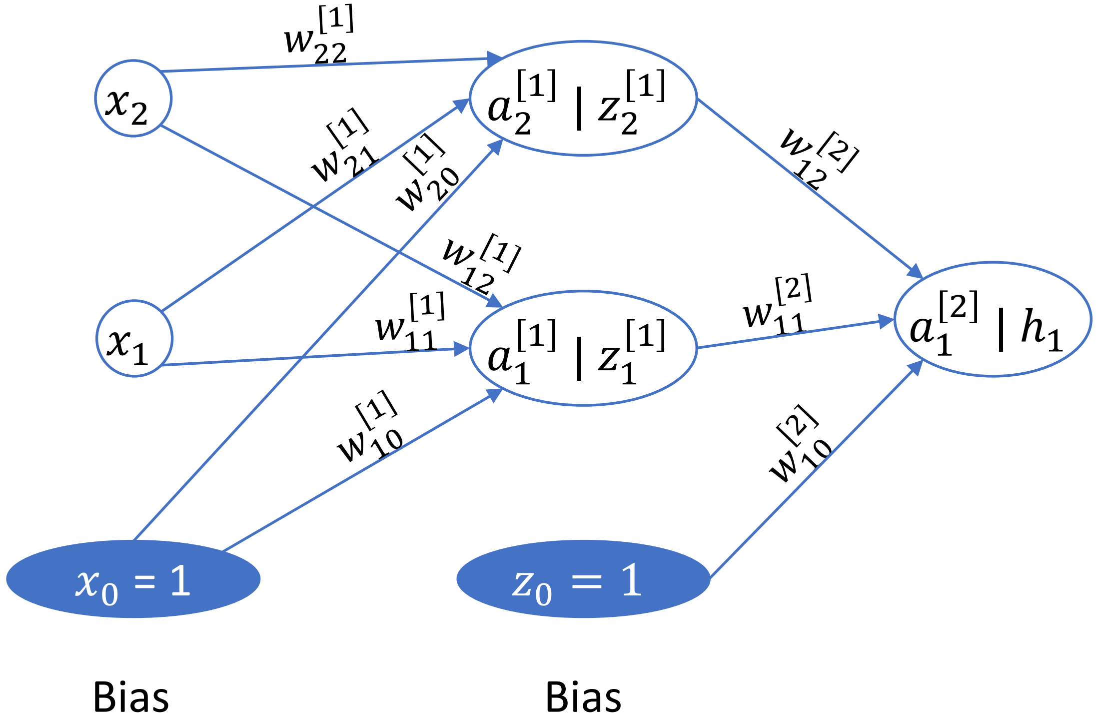
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow \boxed{h\left(\mathbf{x}^{(1)}\right) = 0.42}$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(i) Calculate the error using LS Error.



LEAST SQUARES ERROR

$$\mathbb{E}\left(\mathbf{x}^{(1)}\right) = \frac{1}{2}\sum_{n=1}^{N}\left(h\left(\mathbf{x}^{(n)}\right) - y^{(n)}\right)^2$$

$$= \frac{1}{2}\left(h\left(\mathbf{x}^{(1)}\right) - y^{(1)}\right)^2 = \frac{1}{2}(0.42-1)^2 = 0.17$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

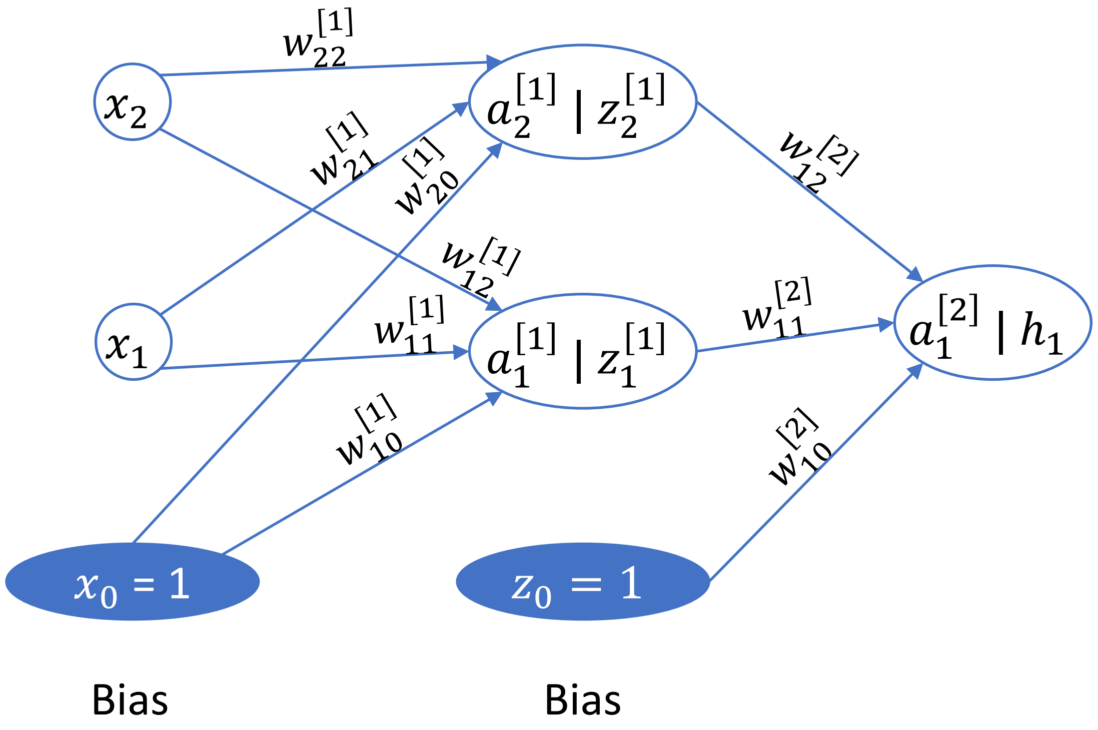
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$. Draw only the part of the graph and the relations involved. What is the local error? How could the local error be used to calculate the derivatives of the error with respect to the weights of the output layer?



EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

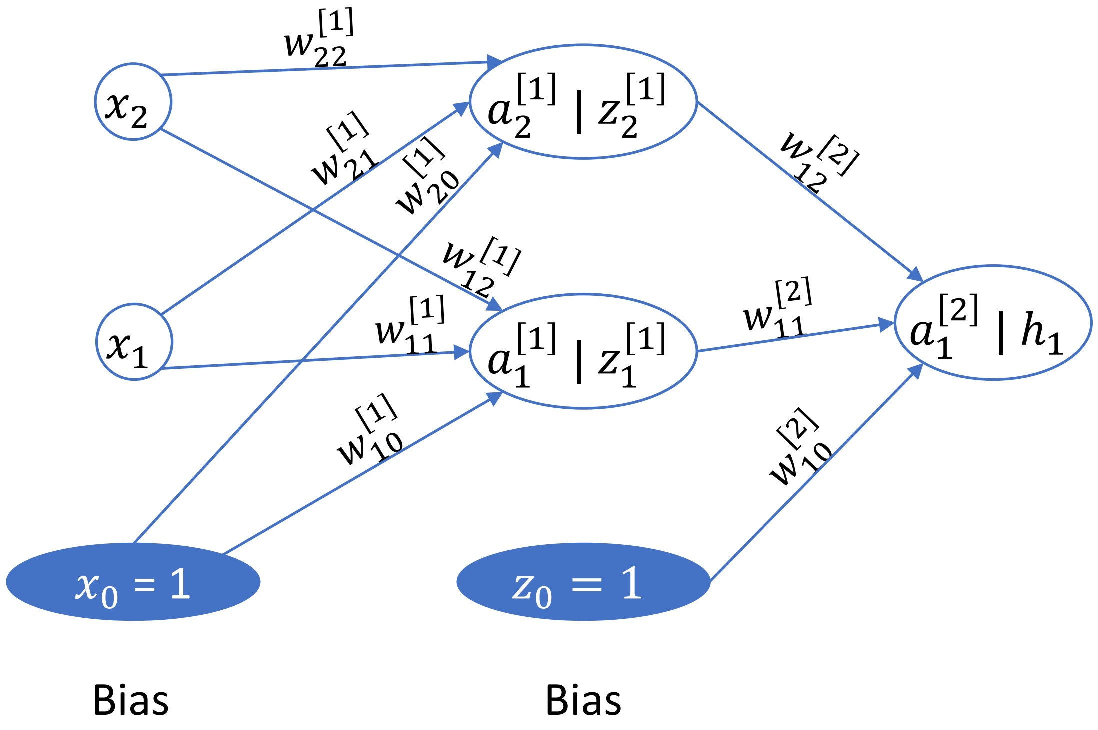
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

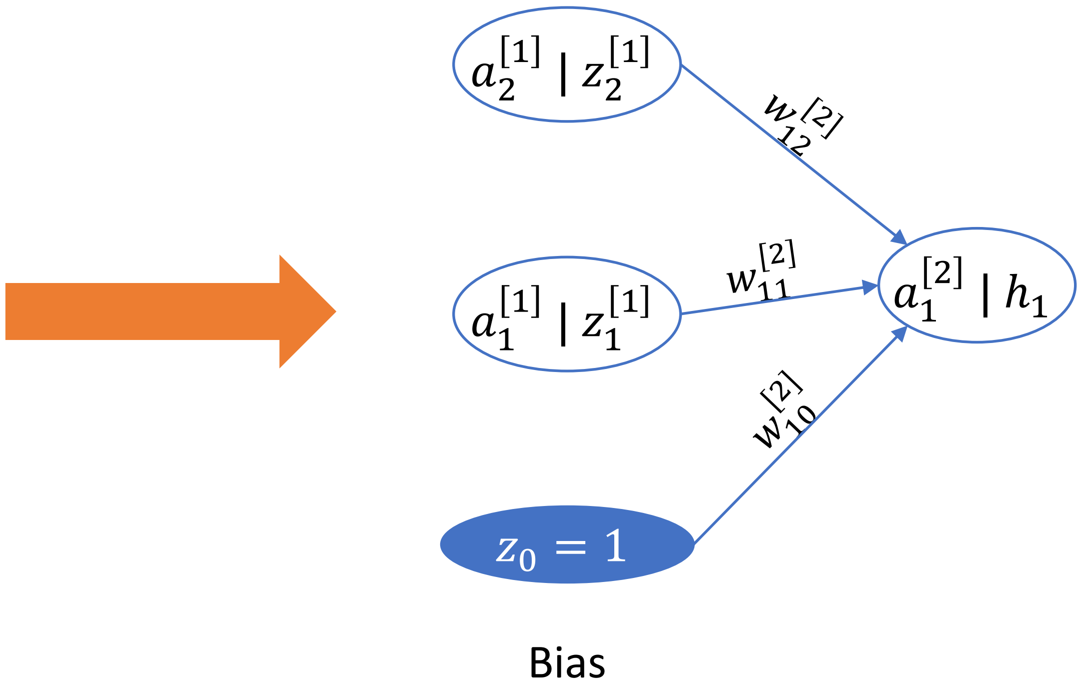
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$. Draw only the part of the graph and the relations involved. What is the local error? How could the local error be used to calculate the derivatives of the error with respect to the weights of the output layer?





EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

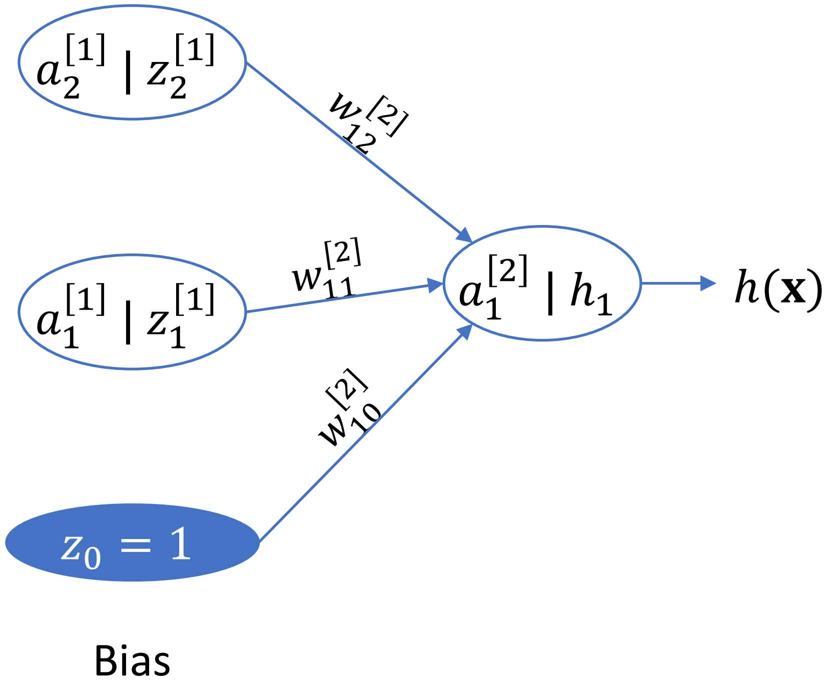
BACK-PROPAGATION

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.



$\mathbb{E}_1 = ?$

Error for input point $\mathbf{x}^{(1)}$.

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$
$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

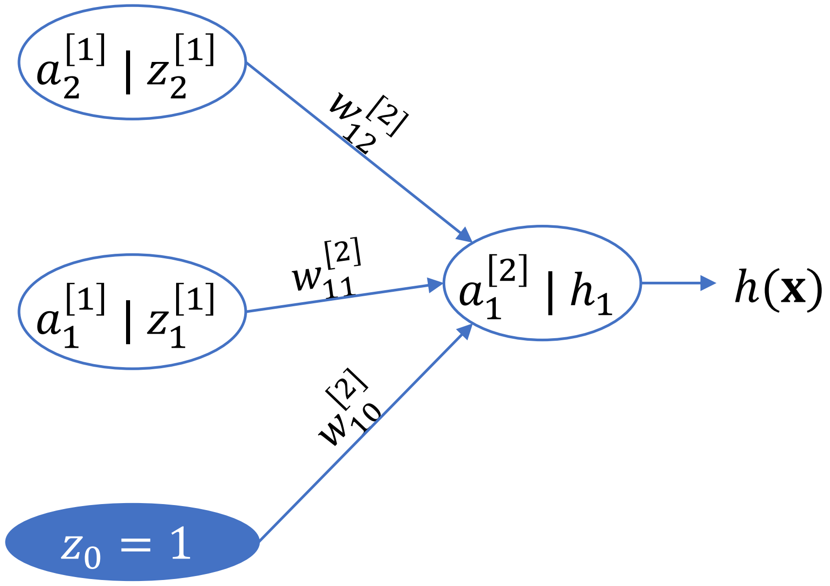
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.



Bias

The **output of the network** for input point $\mathbf{x}^{(1)}$.

$$\mathbb{E}_1 = \frac{1}{2}\left(h\left(\mathbf{x}^{(1)}\right) - y^{(1)}\right)^2 = \frac{1}{2}\left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)^2$$

$h\left(\mathbf{x}^{(1)}\right)$ is a function of $a_1^{[2]}$. Which in turn is a function of $w_{1m}^{[2]}$ and $z_m^{[1]}$, for $m = 1, 2$.

$$h\left(\mathbf{x}^{(1)}\right) = h_1\left(a_1^{[2]}\right) = h_1\left(\sum_{m=0}^{2} w_{1m}^{[2]} z_m^{[1]}\right)$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

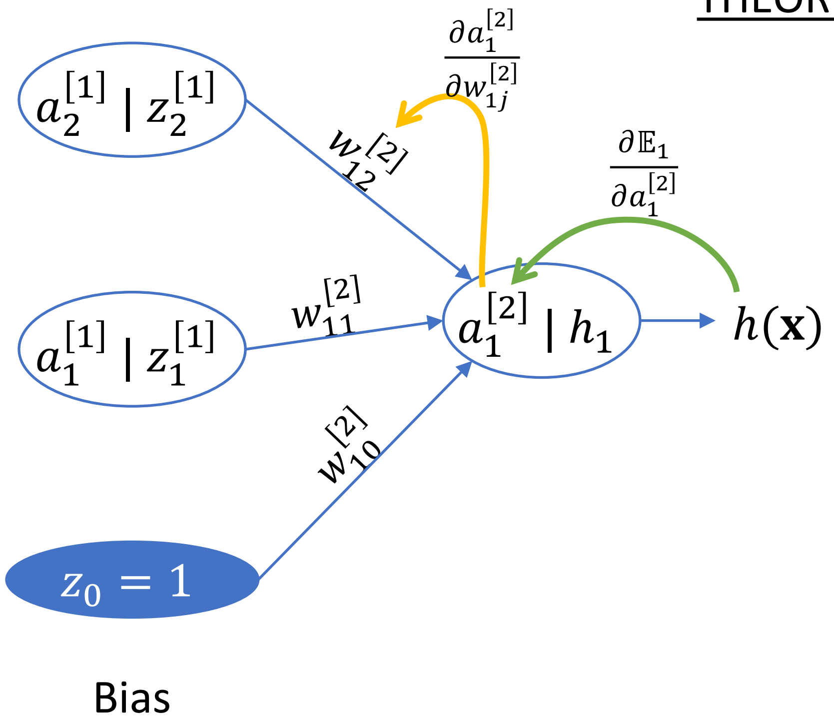
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.



<u>THEORETICALLY</u>:

$$\frac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}}$$

CHAIN RULE

$$\mathbb{E}_1 = \frac{1}{2}\left(h\left(\mathbf{x}^{(1)}\right) - y^{(1)}\right)^2 = \frac{1}{2}\left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)^2$$

LOCAL ERROR $\partial_1^{[2]}$ $\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}} = ?$

$$\boxed{\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}} = ?$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

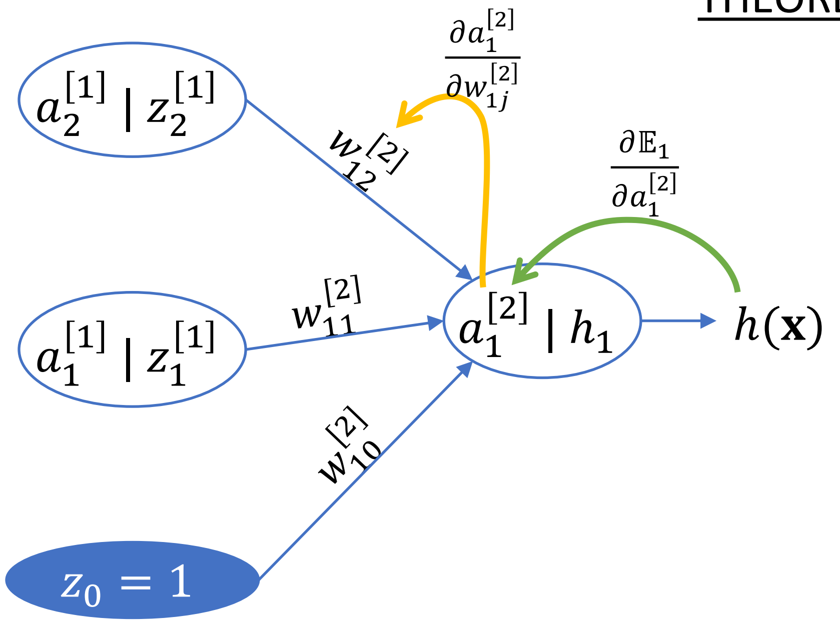
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.



Bias

<u>THEORETICALLY</u>:

$$\frac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}}$$

CHAIN RULE

$$\mathbb{E}_1 = \frac{1}{2}\left(h\left(\mathbf{x}^{(1)}\right) - y^{(1)}\right)^2 = \frac{1}{2}\left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)^2$$

LOCAL ERROR $\partial_1^{[2]}$

$$\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}} = \left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right) \cdot h_1'\left(a_1^{[2]}\right) = h_1\left(a_1^{[2]}\right) - y^{(1)}$$

$h_1(a) = a$

$$\boxed{\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}} = \frac{\partial}{\partial w_{1j}^{[2]}}\underbrace{\left\{w_{10}^{[2]} + w_{11}^{[2]} z_1^{[1]} + w_{12}^{[2]} z_2^{[1]}\right\}}_{a_1^{[2]}} = z_j^{[1]}$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

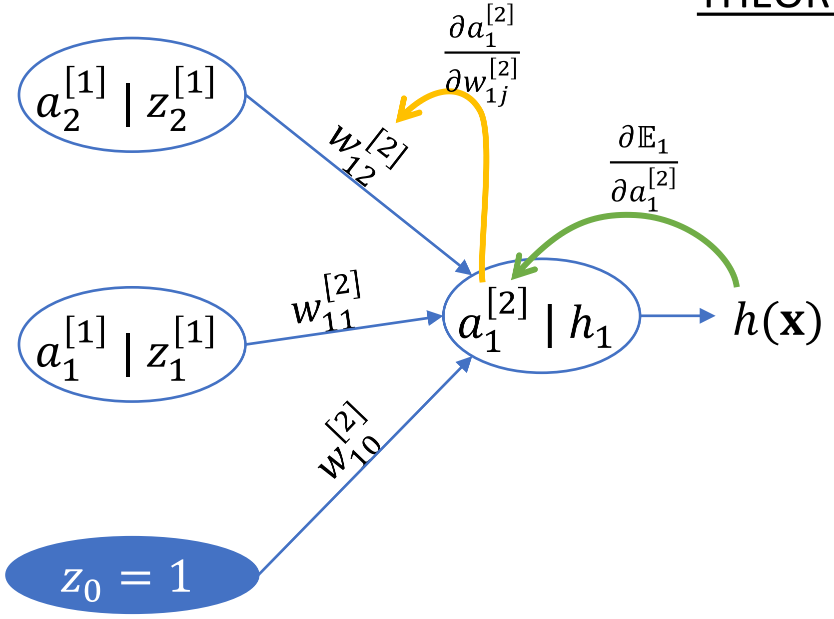
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.



Bias

<u>THEORETICALLY</u>:

$$\frac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}} = \underbrace{\boxed{\left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)}}_{\substack{\textit{LOCAL ERROR} \\ \partial_1^{[2]}}} \boxed{z_j^{[1]}}$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

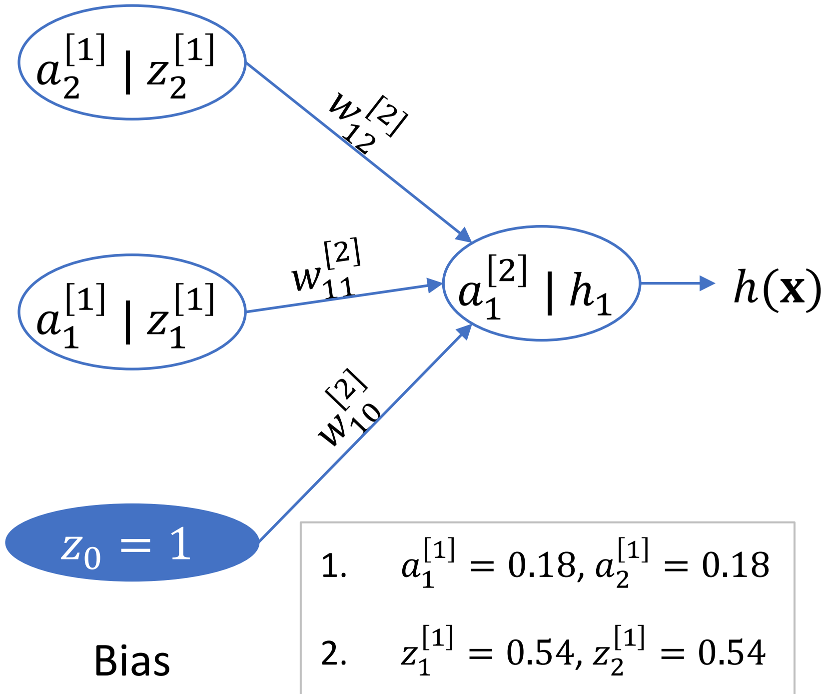
(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1$ $\longrightarrow$ $h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ $\quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\boldsymbol{\partial_1^{[2]}}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.

<u>THEORETICALLY</u>: $\quad \frac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}} = \left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right) \cdot z_j^{[1]}$



LOCAL ERROR $\quad \partial_1^{[2]} = \left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right) = ?$

1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54, z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

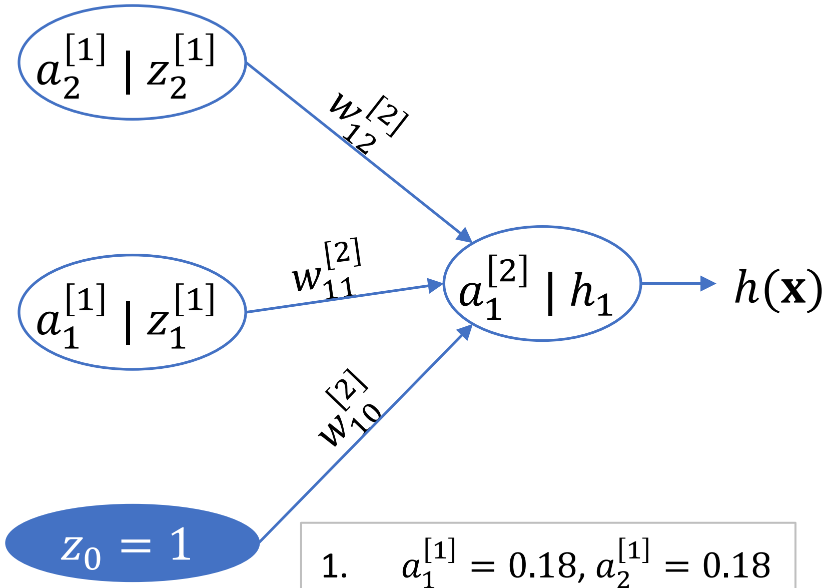
with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\boldsymbol{\partial_1^{[2]}}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.

<ins>THEORETICALLY</ins>:

$$\frac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}} = \left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right) \cdot z_j^{[1]}$$



Bias

LOCAL ERROR

$$\partial_1^{[2]} = \left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)$$
$$= a_1^{[2]} - y^{(1)} = 0.42 - 1 = -0.58$$

1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54, z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

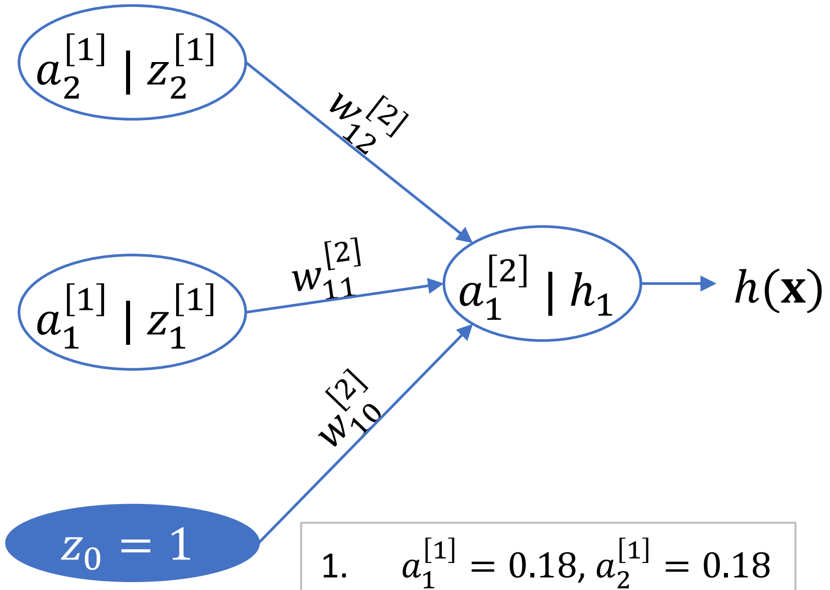
with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\boldsymbol{\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}}$ for all the parameters involved.

<u>THEORETICALLY</u>:

$$\frac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}} = \overbrace{\left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)}^{\partial_1^{[2]} = -0.58} \cdot z_j^{[1]}$$



Bias

1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54, z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

$$\frac{\partial \mathbb{E}_1}{\partial w_{10}^{[2]}} = ?$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{11}^{[2]}} = ?$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{12}^{[2]}} = ?$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

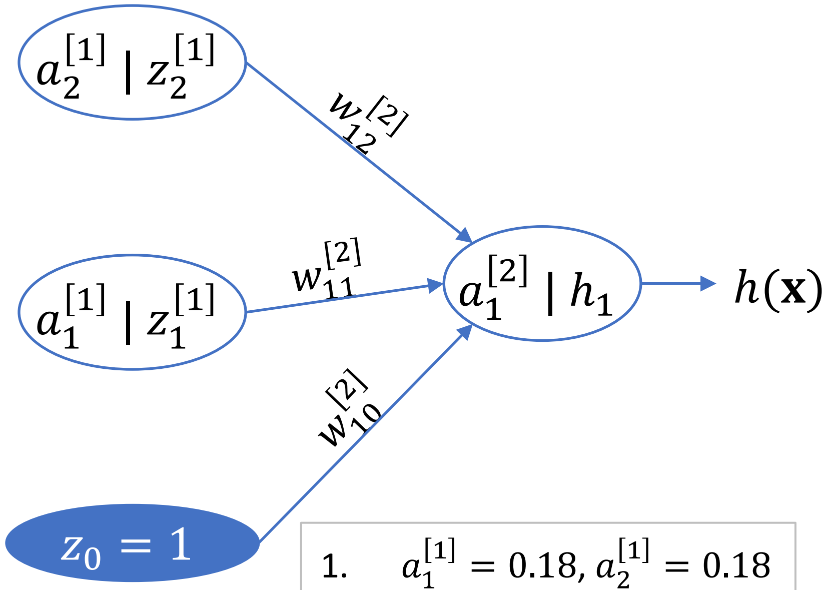
with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\boldsymbol{\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}}$ for all the parameters involved.

<u>THEORETICALLY</u>:

$$\frac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}} = \overbrace{\left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)}^{-0.58} \cdot z_j^{[1]}$$



Bias

1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54, z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

$$\frac{\partial \mathbb{E}_1}{\partial w_{10}^{[2]}} = -0.58 \cdot z_0^{[1]} = -0.58 \cdot 1 = -0.58$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{11}^{[2]}} = -0.58 \cdot z_1^{[1]} = -0.58 \cdot 0.54 = -0.31$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{12}^{[2]}} = -0.58 \cdot z_2^{[1]} = -0.58 \cdot 0.54 = -0.31$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

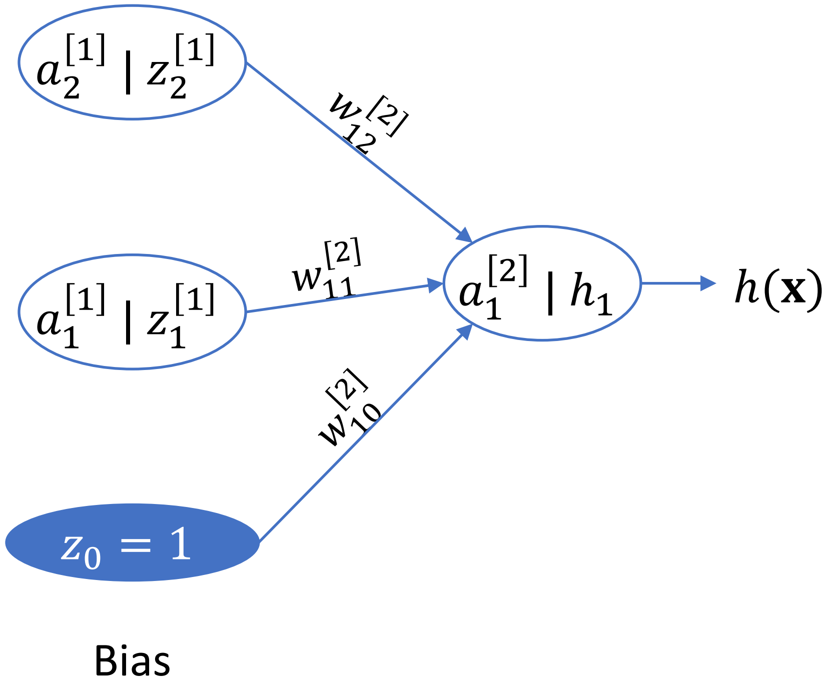
(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\boldsymbol{\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}}$ for all the parameters involved.

$$\frac{\partial \mathbb{E}_1}{\partial w_{10}^{[2]}} = -0.58 \qquad \frac{\partial \mathbb{E}_1}{\partial w_{11}^{[2]}} = -0.31 \qquad \frac{\partial \mathbb{E}_1}{\partial w_{12}^{[2]}} = -0.31$$



Gradient descent for the weights of the output layer, with $\alpha = 0.5$?

$$w_{10}^{[2]t+1} = w_{10}^{[2]t} - 0.5 \cdot (-0.58) = 0.2 + 0.29 = 0.49$$

$$w_{11}^{[2]t+1} = w_{11}^{[2]t} - 0.5 \cdot (-0.31) = 0.2 + 0.15 = 0.35$$

$$w_{12}^{[2]t+1} = w_{12}^{[2]t} - 0.5 \cdot (-0.31) = 0.2 + 0.15 = 0.35$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

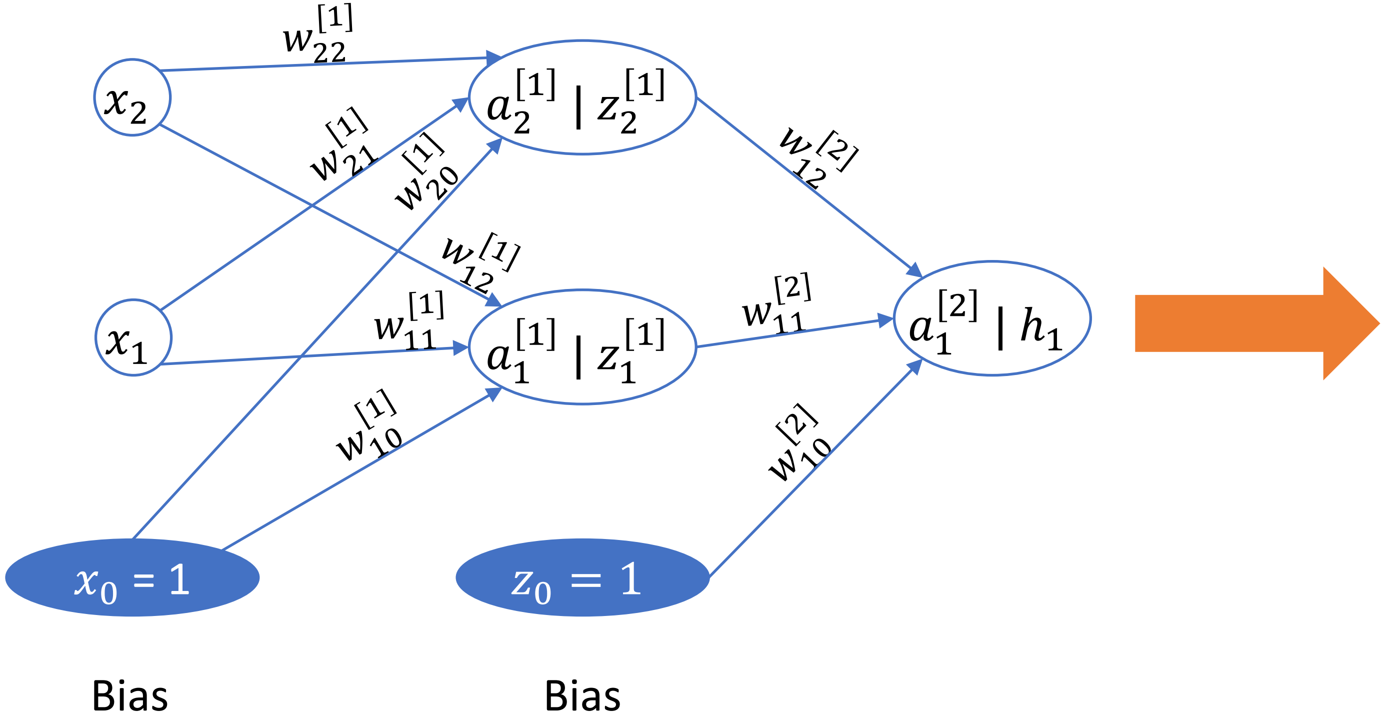
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$. Draw only the part of the graph and the relations involved.

Compute $\partial_1^{[1]}, \partial_2^{[1]}$and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.



EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

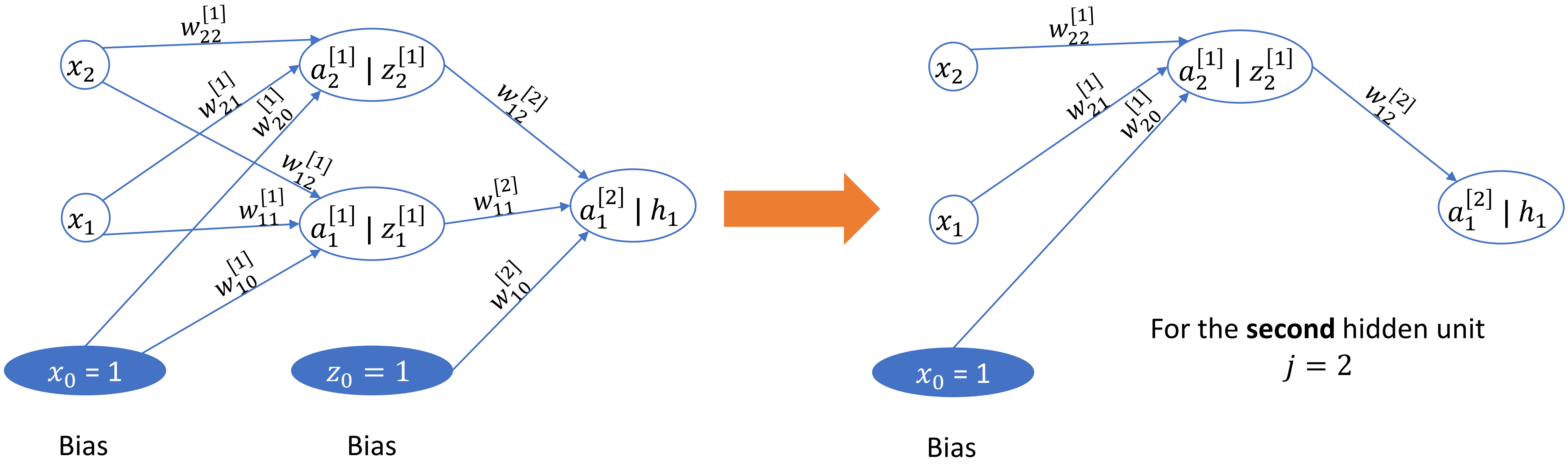
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ $\quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$. <u>Draw only the part of the graph and the relations involved</u>.

Compute $\partial_1^{[1]}, \partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.



EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

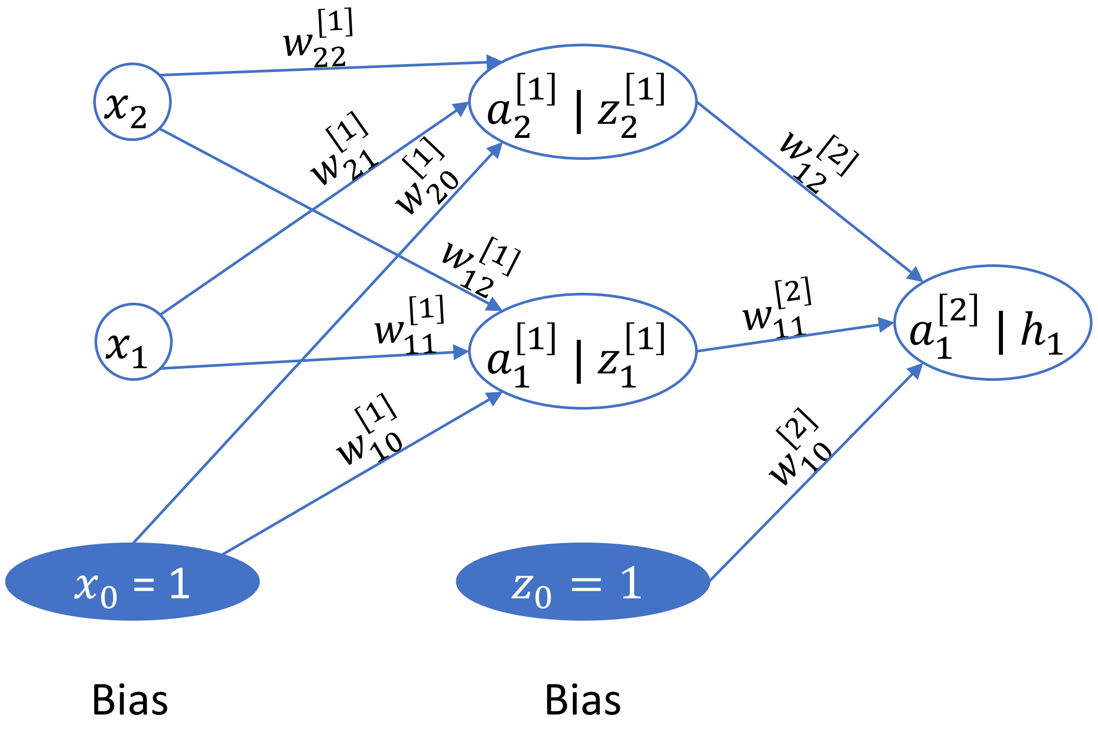
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

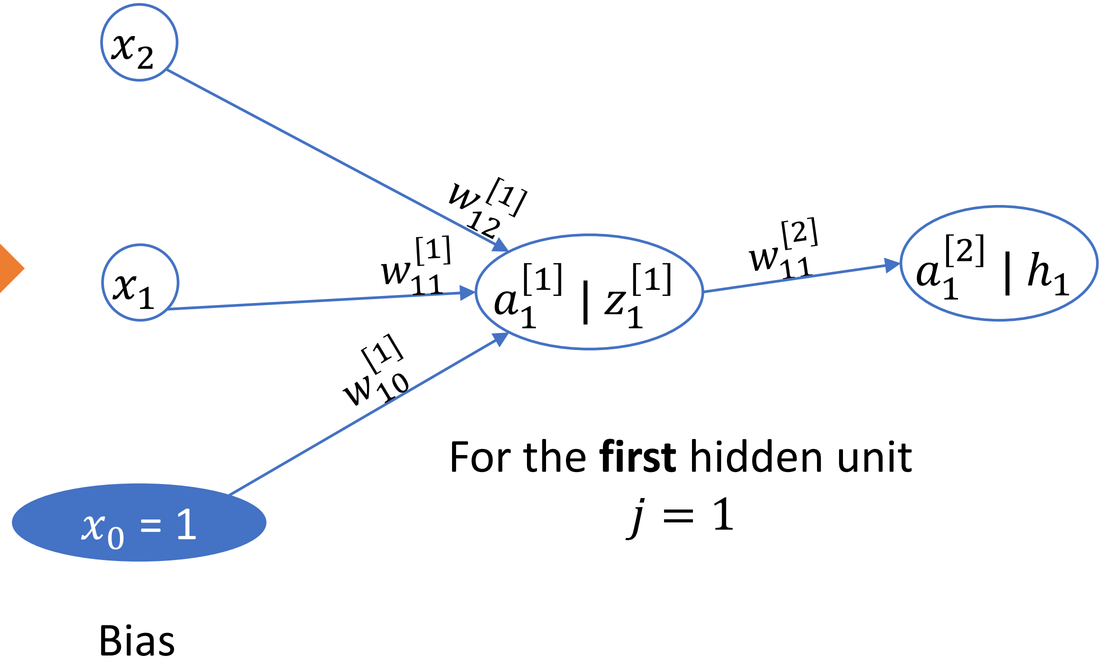
(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$. Draw only the part of the graph and the relations involved.

Compute $\partial_1^{[1]}, \partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.





For the **first** hidden unit
$j = 1$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

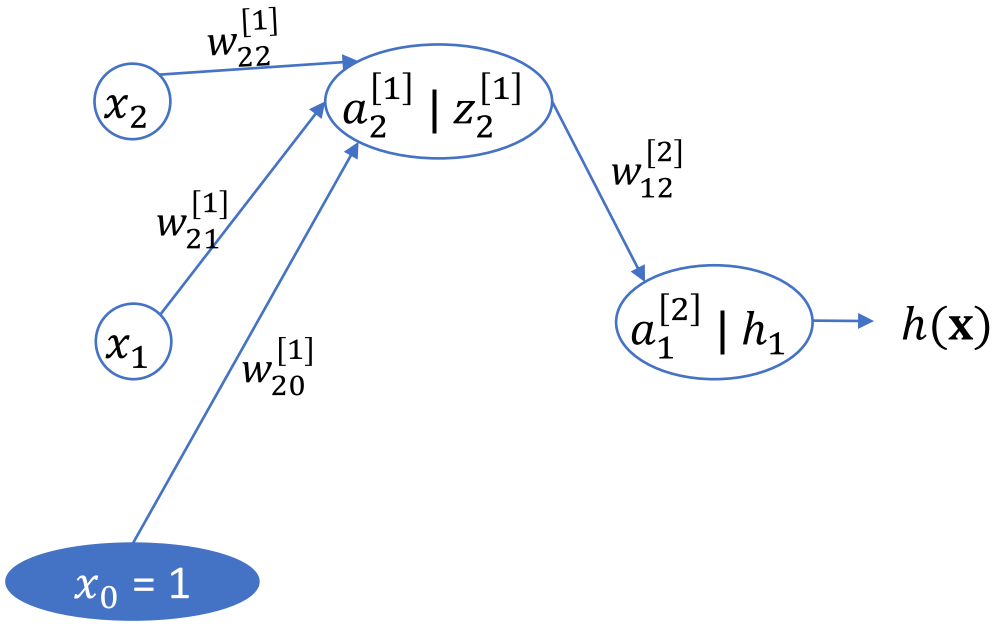
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$. <u>Draw only the part of the graph and the relations involved</u>.

Compute $\partial_1^{[1]}, \partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.



We will proceed for the **second** hidden unit.

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

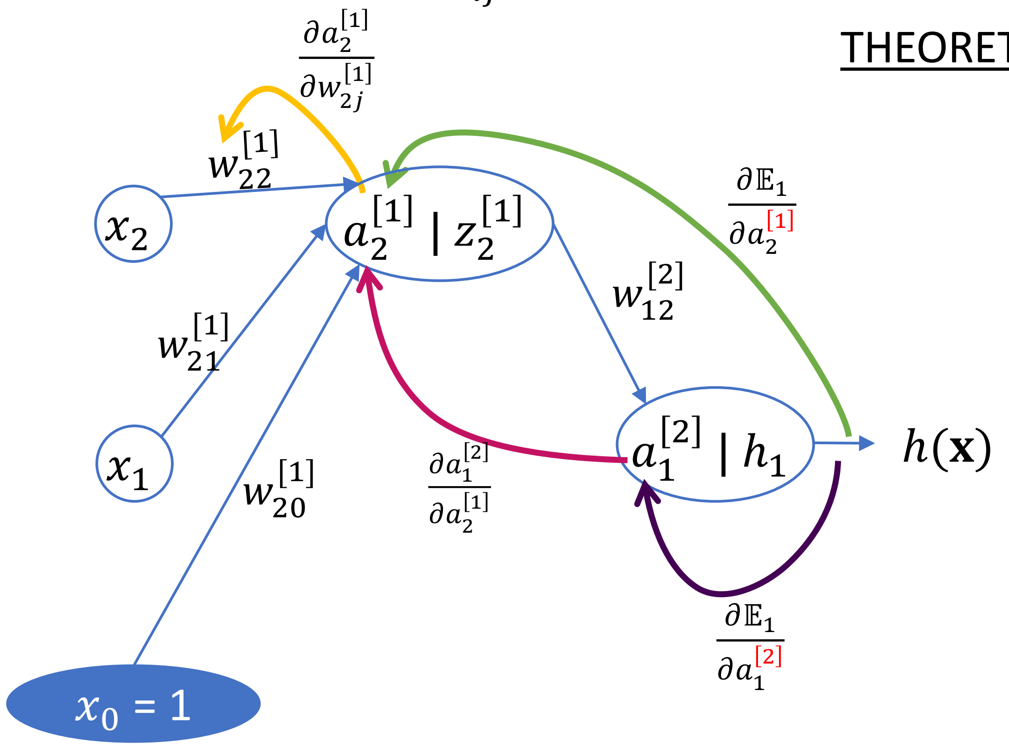
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$.

Compute $\partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.



<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}}}\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}}$$

CHAIN RULE

LOCAL ERROR $\partial_2^{[1]}$

$$a_1^{[2]} = \sum_{m=0}^{2} w_{1m}^{[2]} z_m^{[1]} = \sum_{m=0}^{2} w_{1m}^{[2]} \varphi\left(a_m^{[1]}\right)$$

$$\mathbb{E}_1 = \frac{1}{2}\left(h\left(\mathbf{x}^{(1)}\right) - y^{(1)}\right)^2 = \frac{1}{2}\left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)^2$$

$$\boxed{\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}}} = \underbrace{\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}}_{\partial_1^{[2]}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}}$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

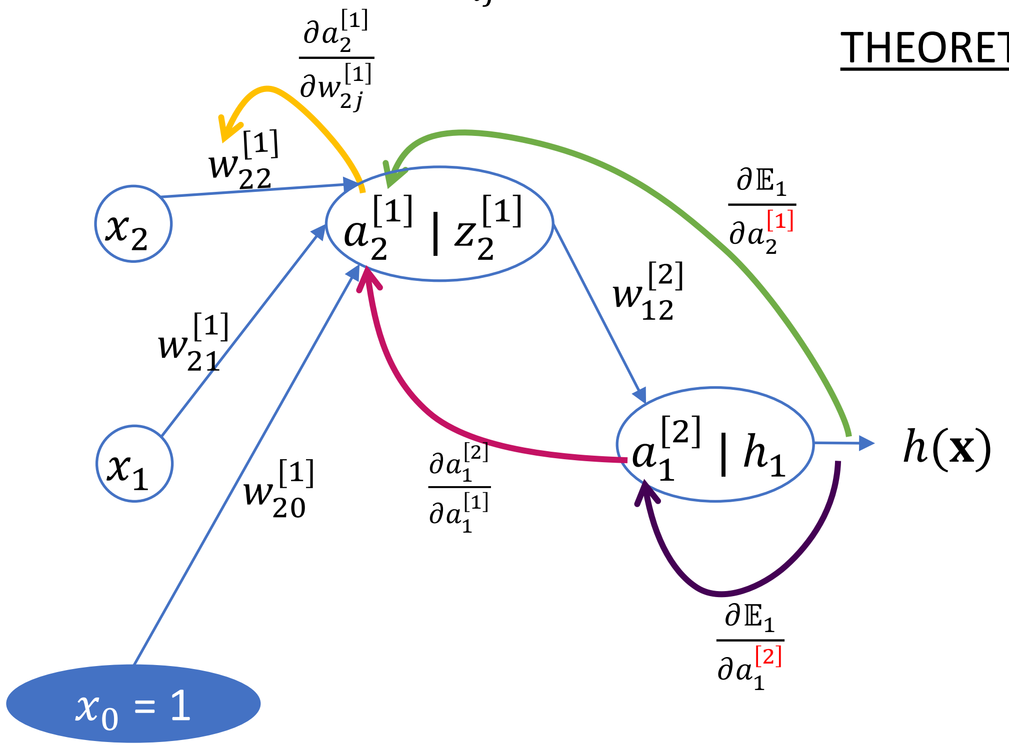
(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1$ $\longrightarrow$ $h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$.

Compute $\partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.

$$\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \partial_2^{[1]} \quad \textit{LOCAL ERROR}$$



<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}}}\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}} \qquad \textit{CHAIN RULE}$$

$$\boxed{\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}}} \left\{ \begin{array}{l} \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}} = ? \\ \\ \boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}} = ? \end{array} \right.$$

$$\mathbb{E}_1 = \frac{1}{2}\left(h\left(\mathbf{x}^{(1)}\right) - y^{(1)}\right)^2 = \frac{1}{2}\left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)^2$$

EXERCICE 1

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

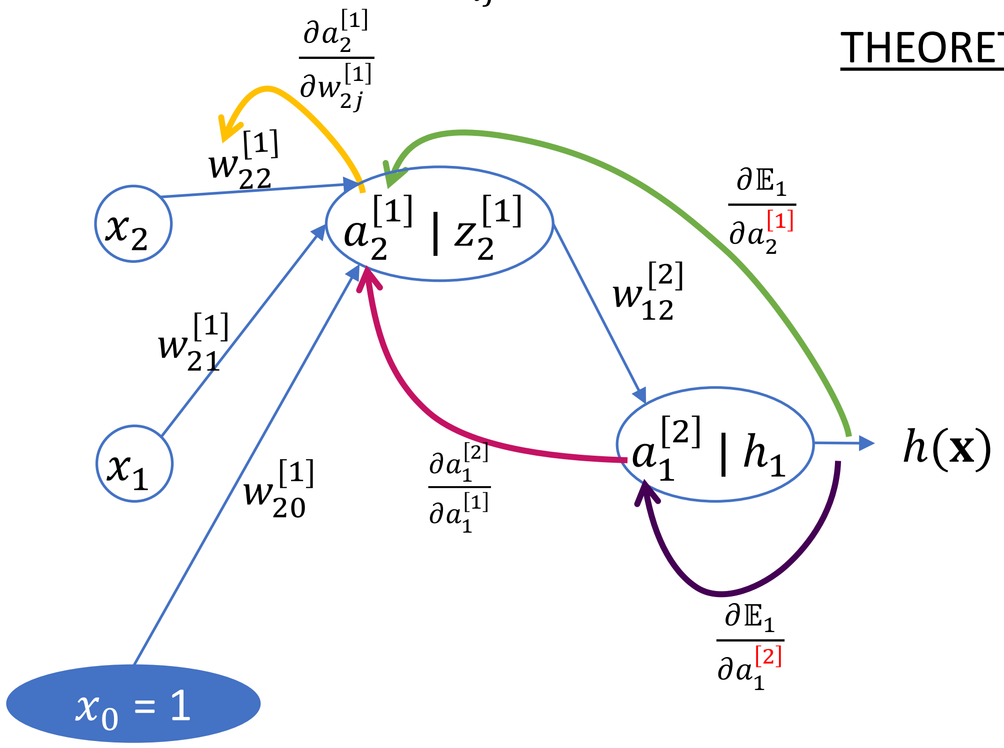
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$.

Compute $\partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.

$$\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \partial_2^{[1]}$$ *LOCAL ERROR*



<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}}}\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}}$$ *CHAIN RULE*

$$\boxed{\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}}} \begin{cases} \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}} = \left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right) \cdot h_1'\left(a_1^{[2]}\right) = h_1\left(a_1^{[2]}\right) - y^{(1)} \\ \mathbb{E}_1 = \frac{1}{2}\left(h\left(\mathbf{x}^{(1)}\right) - y^{(1)}\right)^2 = \frac{1}{2}\left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right)^2 \\ \boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}} = ? \end{cases}$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

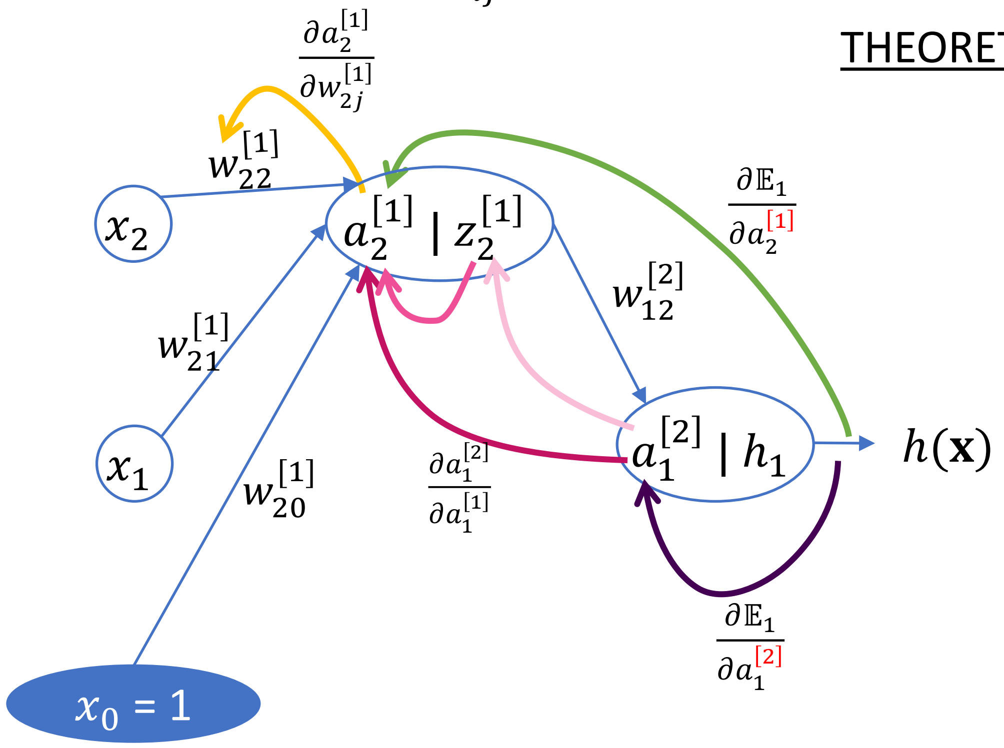
(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ $\quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$.

Compute $\partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.

$$\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \partial_2^{[1]}$$ *LOCAL ERROR*



<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}}}\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}}$$

CHAIN RULE

$$\boxed{\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}}}\left\{\begin{array}{l} \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}} = h_1\left(a_1^{[2]}\right) - y^{(1)} = \partial_1^{[2]} \\ a_1^{[2]} = \sum_{m=0}^{2} w_{1m}^{[2]} z_m^{[1]} = \sum_{m=0}^{2} w_{1m}^{[2]} \varphi\left(a_m^{[1]}\right) \\ \boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}} = \boxed{\frac{\partial a_1^{[2]}}{\partial z_2^{[1]}}}\boxed{\frac{\partial z_2^{[1]}}{\partial a_2^{[1]}}} = ? \end{array}\right.$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

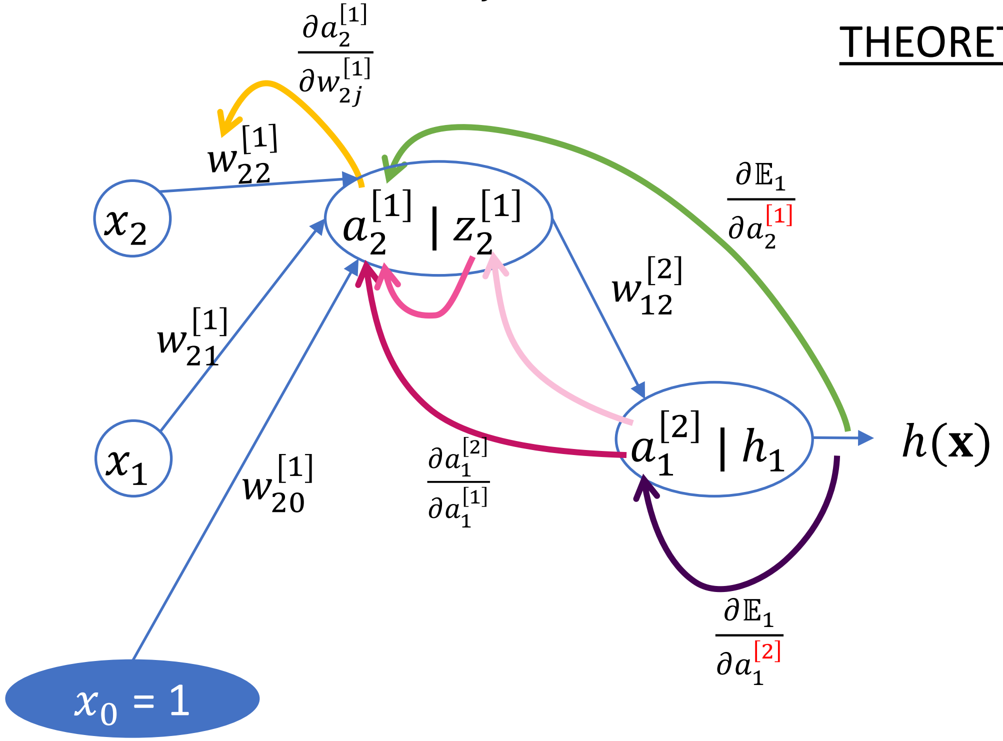
(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1$ $\longrightarrow$ $h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$.

Compute $\partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.

$$\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \partial_2^{[1]}$$ *LOCAL ERROR*



<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}} \frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}$$ *CHAIN RULE*

$$\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} \left\{ \begin{array}{l} \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} = h_1\left(a_1^{[2]}\right) - y^{(1)} = \partial_1^{[2]} \\[2ex] \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}} = \frac{\partial a_1^{[2]}}{\partial z_2^{[1]}} \frac{\partial z_2^{[1]}}{\partial a_2^{[1]}} \\[2ex] = \frac{\partial}{\partial z_2^{[1]}} \left\{ w_{10}^{[2]} z_0^{[1]} + w_{11}^{[2]} z_1^{[1]} + w_{12}^{[2]} z_2^{[1]} \right\} \frac{\partial}{\partial a_2^{[1]}} \varphi\left(a_2^{[1]}\right) \end{array} \right.$$

Activation function

σ

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

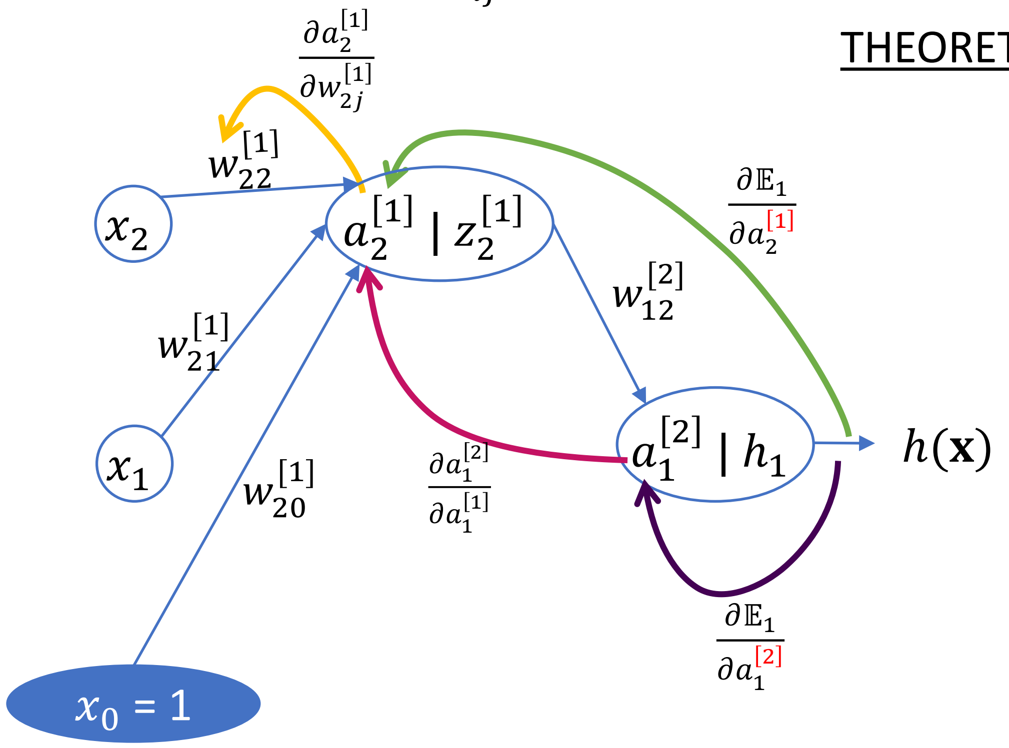
(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$.

Compute $\partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.

$$\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \partial_2^{[1]} \quad \textit{LOCAL ERROR}$$



<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}}}\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}} \qquad \textit{CHAIN RULE}$$

$$\boxed{\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}}} \begin{cases} \boxed{\dfrac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}} = h_1\left(a_1^{[2]}\right) - y^{(1)} = \partial_1^{[2]} \\[2em] \boxed{\dfrac{\partial a_1^{[2]}}{\partial a_2^{[1]}}} = \dfrac{\partial a_1^{[2]}}{\partial z_2^{[1]}} \dfrac{\partial z_2^{[1]}}{\partial a_2^{[1]}} \\ \qquad = w_{12}^{[2]} \sigma'\left(a_2^{[1]}\right) \end{cases}$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

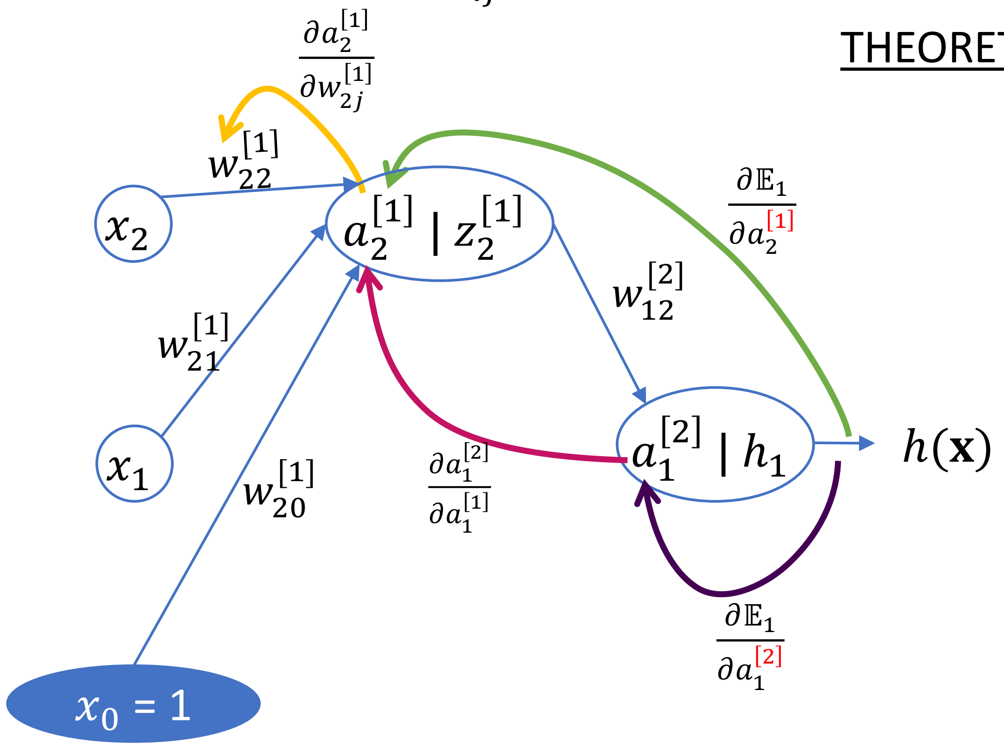
(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ $\quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$.

Compute $\partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.

$$\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \partial_2^{[1]} \quad \textit{LOCAL ERROR}$$



<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}}}\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}} \qquad \textit{CHAIN RULE}$$

LOCAL ERROR $\partial_2^{[1]}$

$$\boxed{\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}} = \partial_1^{[2]} w_{12}^{[2]} \sigma'\left(a_2^{[1]}\right)$$

$$\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}} = ?$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

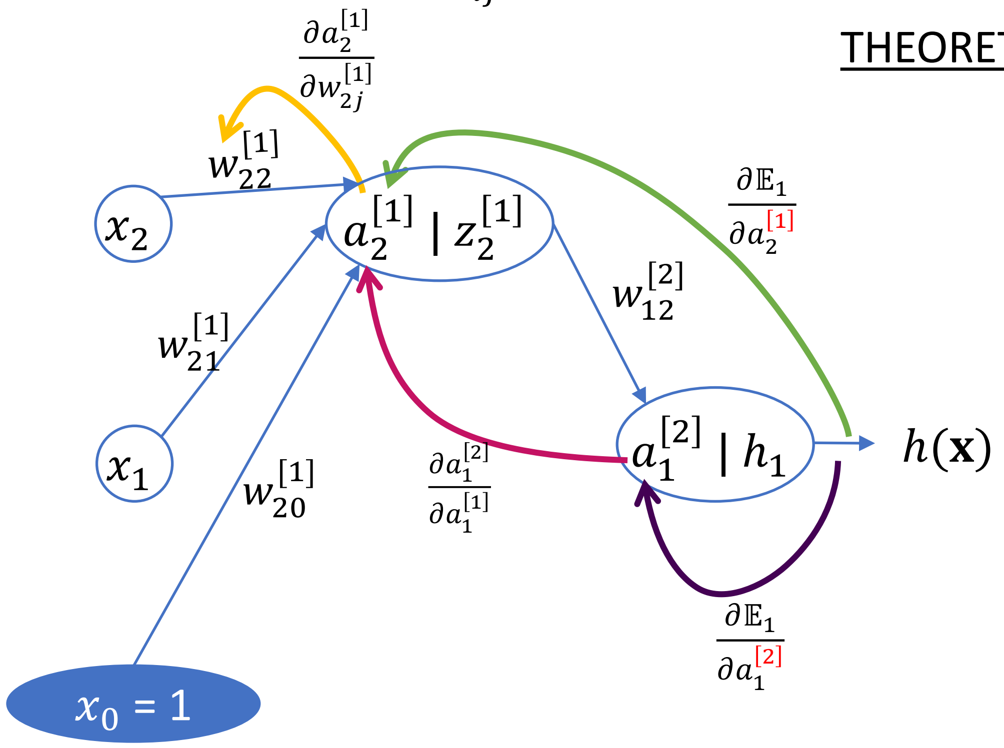
(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$.

Compute $\partial_2^{[1]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}}$ for all the parameters involved.

$$\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \partial_2^{[1]}$$ *LOCAL ERROR*



<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}}}\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}}$$ *CHAIN RULE*

LOCAL ERROR $\partial_2^{[1]}$

$$\boxed{\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}} = \partial_1^{[2]} w_{12}^{[2]} \sigma'\left(a_2^{[1]}\right)$$

$$\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}} = \frac{\partial}{\partial w_{2j}^{[1]}}\left\{w_{10}^{[1]} x_0 + w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2\right\} = x_j$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1$ $\longrightarrow$ $h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

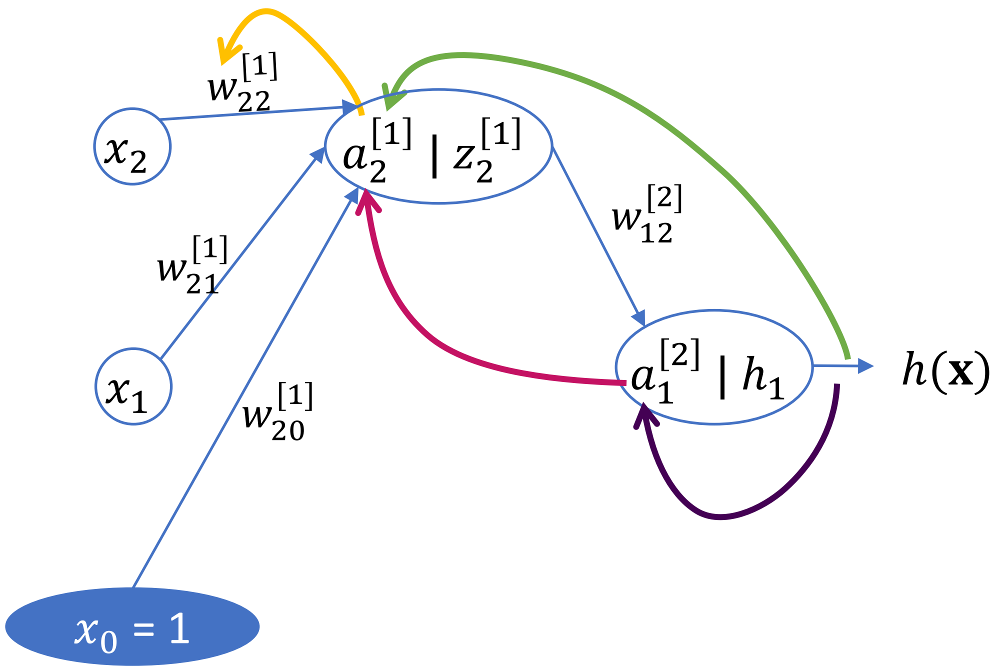
Compute $\partial_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.

$$\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \partial_2^{[1]}$$ *LOCAL ERROR*

<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}}}\boxed{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}} =$$

$$= \boxed{\boxed{\partial_1^{[2]}}\boxed{w_{12}^{[2]}\sigma'\left(a_2^{[1]}\right)}}\boxed{x_j}$$



EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

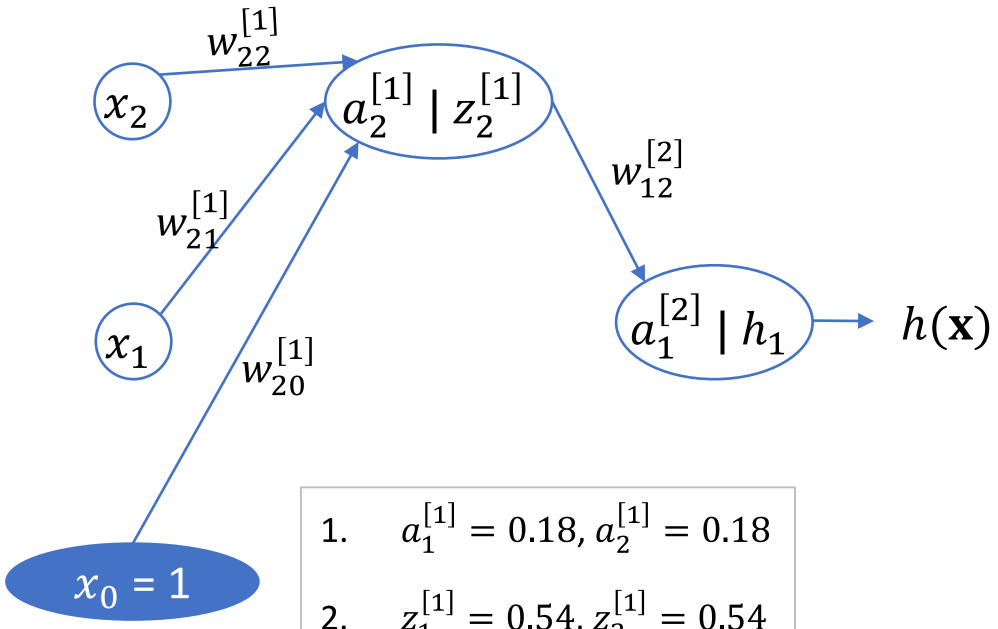
LOCAL ERROR

Compute $\boldsymbol{\partial}_{\mathbf{1}}^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.

<u>THEORETICALLY</u>:

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}} \frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}} =$$

$$= \boxed{\partial_1^{[2]} \; w_{12}^{[2]} \sigma'\left(a_2^{[1]}\right)} \; x_j$$



$$\boxed{\partial_2^{[1]}} = ?$$

1. $a_1^{[1]} = 0.18,\ a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54,\ z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

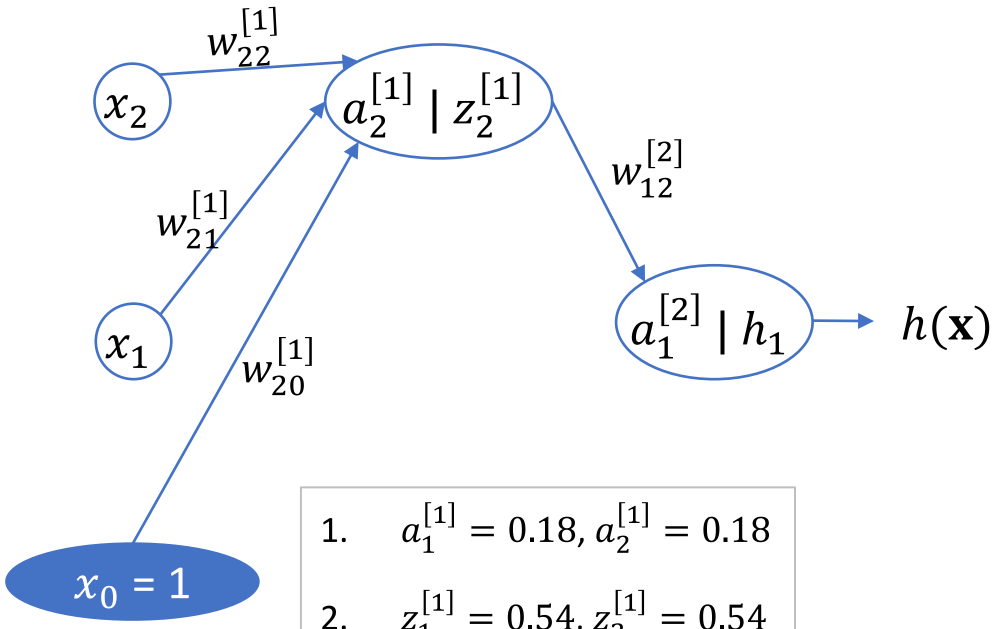
Compute $\boldsymbol{\partial}_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.

LOCAL ERROR

THEORETICALLY:

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}} \frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}} =$$

$$= \boxed{\partial_1^{[2]} \, w_{12}^{[2]} \sigma'\left(a_2^{[1]}\right)} \; x_j$$



1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54, z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

$$\boxed{\partial_2^{[1]}} = \frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}} = \partial_1^{[2]} \, w_{12}^{[2]} \sigma'\left(a_2^{[1]}\right)$$

$$= \left(h_1\left(a_1^{[2]}\right) - y^{(1)}\right) w_{12}^{[2]} \sigma\left(a_2^{[1]}\right)\left(1 - \sigma\left(a_2^{[2]}\right)^2\right)$$

$$= (0.42 - 1)0.2\sigma(0.18)(1 - \sigma(0.18)) = -0.03$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.42$

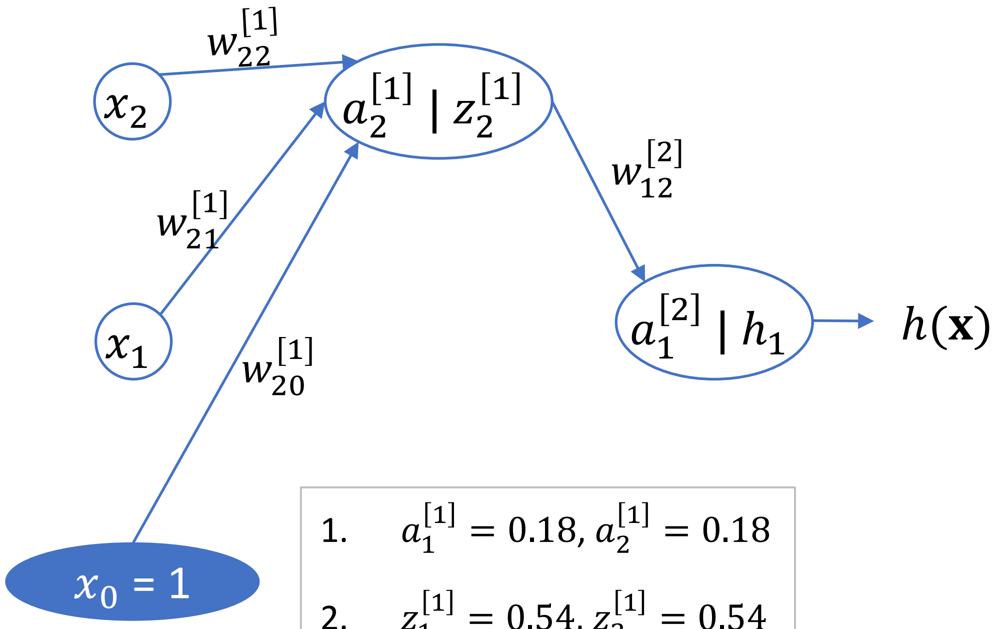
with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.

<u>THEORETICALLY</u>:

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}} \frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}} = \overbrace{\partial_1^{[2]} \; w_{12}^{[2]} \sigma'\left(a_2^{[1]}\right)}^{\partial_2^{[1]} = -0.03} \; x_j$$



1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54, z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

$$\frac{\partial \mathbb{E}_1}{\partial w_{20}^{[1]}} = ?$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{21}^{[1]}} = ?$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{22}^{[1]}} = ?$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1$ $\longrightarrow$ $h\left(\mathbf{x}^{(1)}\right) = 0.42$

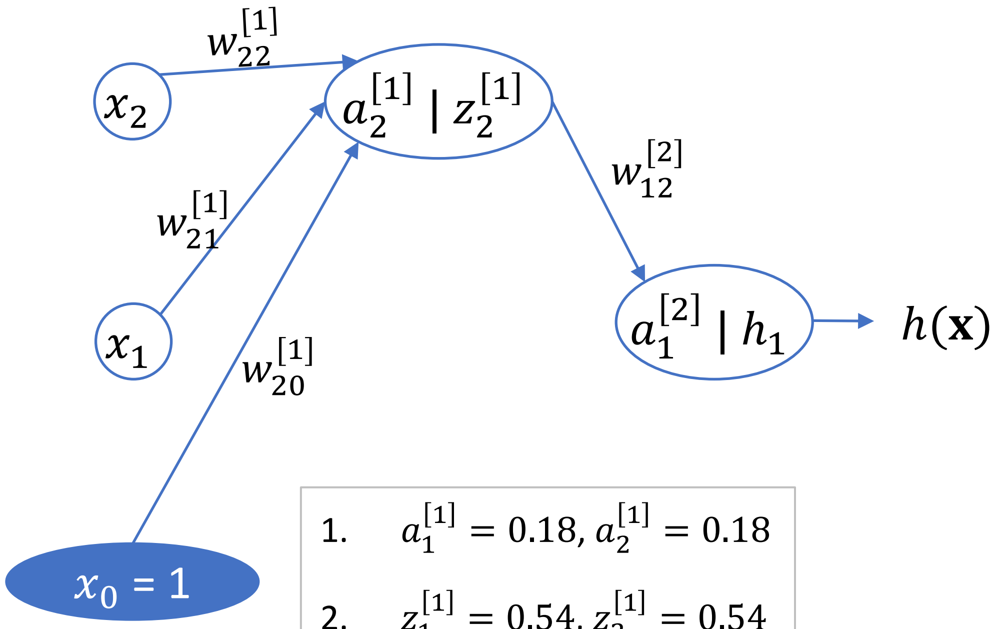
with $\quad h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$.

Compute $\partial_1^{[2]}$ and $\boldsymbol{\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}}$ for all the parameters involved.

<u>THEORETICALLY</u>:

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}} \frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}} = \overbrace{\partial_1^{[2]} \; w_{12}^{[2]} \sigma'\left(a_2^{[1]}\right)}^{\partial_2^{[1]} = -0.03} \quad x_j$$



1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54, z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

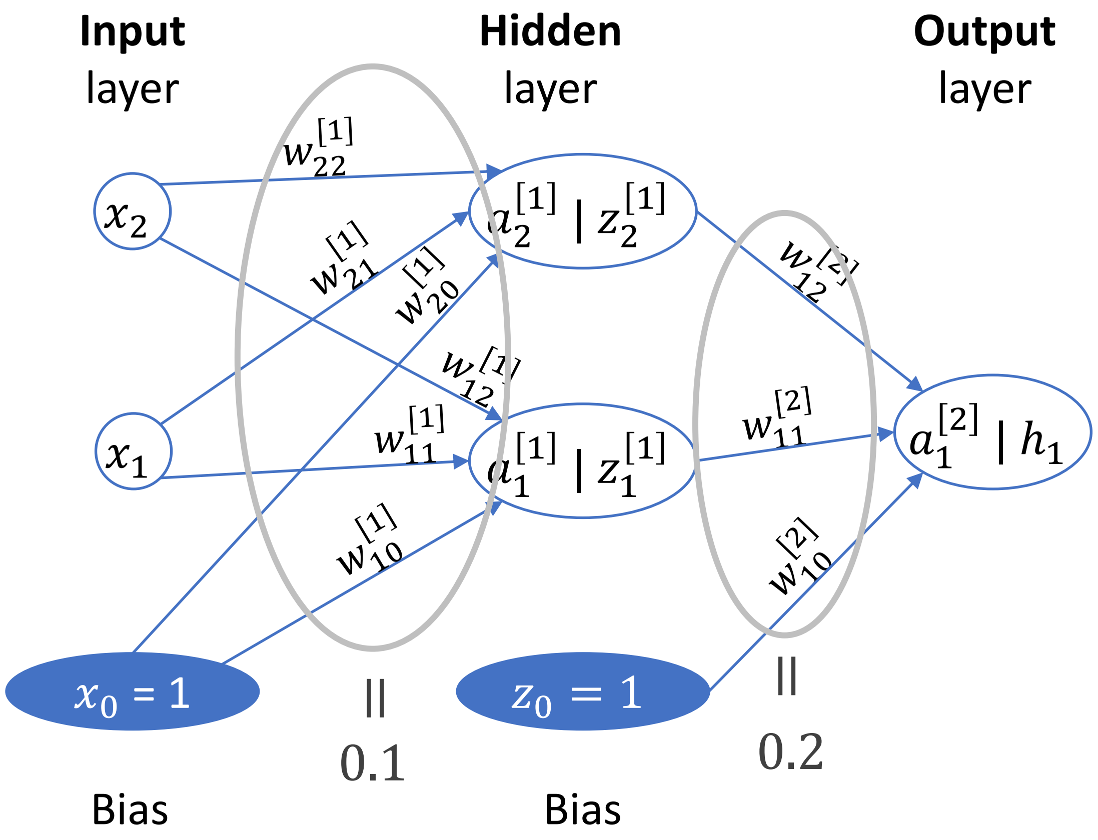
$$\frac{\partial \mathbb{E}_1}{\partial w_{20}^{[1]}} = -0.03 \cdot 1 = -0.03$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{21}^{[1]}} = -0.03 \cdot 0.3 = -0.009$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{22}^{[1]}} = -0.03 \cdot 0.5 = -0.015$$

EXERCICE 1

(d) Same exercise as (c) but with the **sigmoid** for the activation function in the output layer and the **cross-entropy error**.



NEW INFORMATION

$$h_1 = \varphi\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$$

$$\mathbb{E}_n = -y^{(n)} \ln h(\mathbf{x}^{(n)}) - \left(1 - y^{(n)}\right) \ln\left(1 - h(\mathbf{x}^{(n)})\right)$$

TODO

$h\left(\mathbf{x}^{(1)}\right) = ?$

(i) $\mathbb{E}_1 = ?$

(ii) $\partial_1^{[2]} = ?$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}} = ?$

(iii) $\partial_1^{[1]} = ?$, $\partial_2^{[1]} = ?$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}} = ?$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

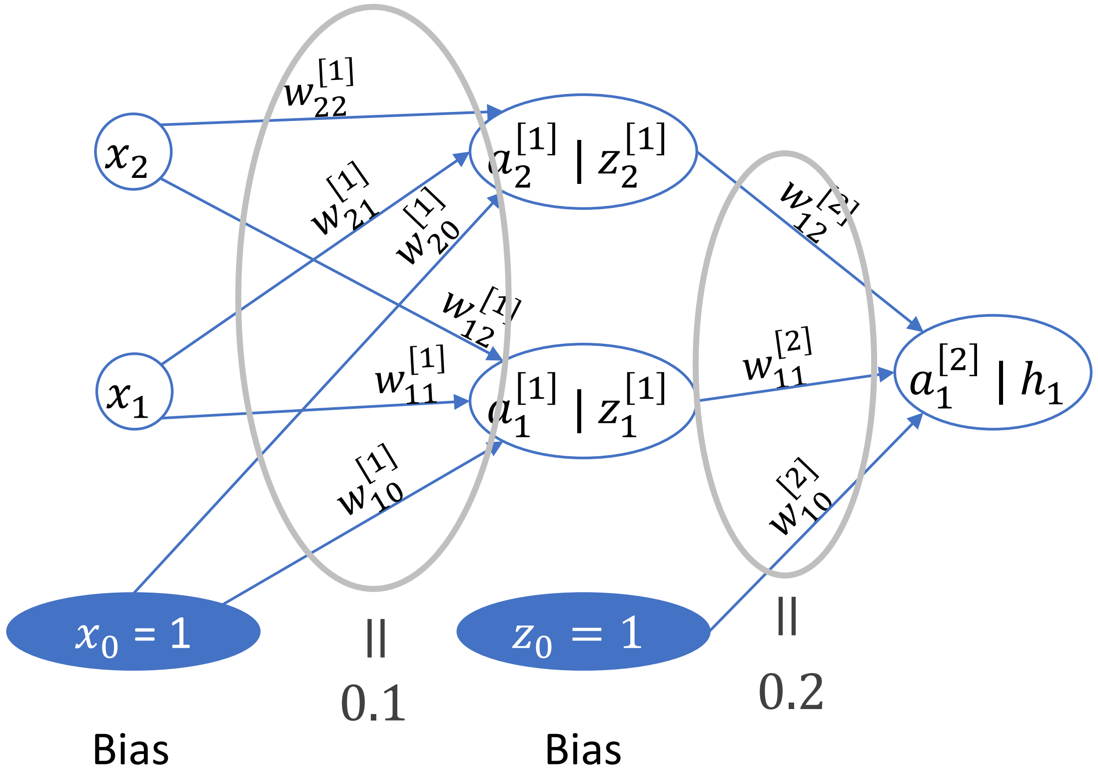
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(d) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = ?$

with $h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n = -y^{(n)} \ln h\left(\mathbf{x}^{(n)}\right) - \left(1-y^{(n)}\right) \ln\left(1 - h\left(\mathbf{x}^{(n)}\right)\right)$$



$a_1^{[1]} = ?$

$a_2^{[1]} = ?$

$z_1^{[1]} = ?$

$z_2^{[1]} = ?$

$a_1^{[2]} = ?$

$h_1 = ?$

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. Compute $a_1^{[2]}$.
4. Compute h_1

EXERCICE 1

(d) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1 \longrightarrow h\left(\mathbf{x}^{(1)}\right)=?$

$$\sigma(a)=\frac{1}{1+\exp(-a)}$$

$$\sigma'(a)=\sigma(a)(1-\sigma(a))$$

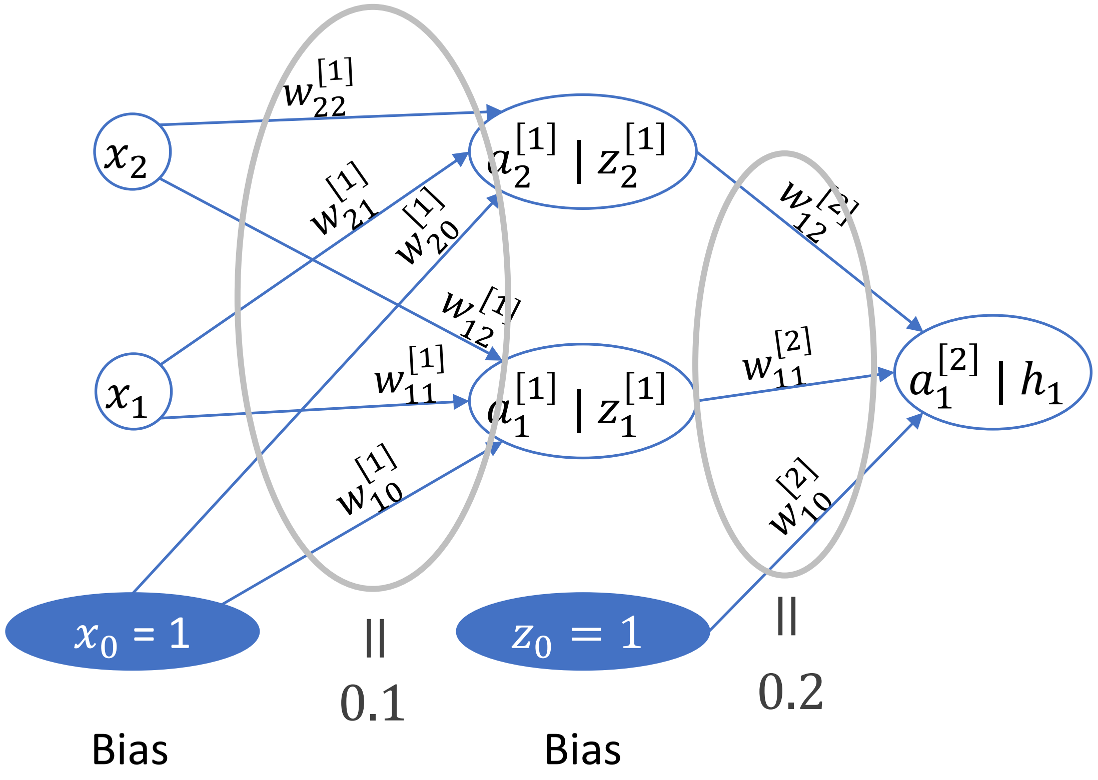
$$h_k=\varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]}=\sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]}=\varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]}=\sum_{d=0}^{D} w_{md}^{[1]} x_d$$

with $h_1\left(a_1^{[2]}\right)=\sigma\left(a_1^{[2]}\right)$ and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n=-y^{(n)}\ln h\left(\mathbf{x}^{(n)}\right)-\left(1-y^{(n)}\right)\ln\left(1-h\left(\mathbf{x}^{(n)}\right)\right)$$

1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
3. Compute $a_1^{[2]}$.
4. Compute h_1



$$a_1^{[1]}=0.18$$

$$a_2^{[1]}=0.18$$

$$z_1^{[1]}=\sigma(0.18)=0.54$$

$$z_2^{[1]}=\sigma(0.18)=0.54$$

$$a_1^{[2]}=0.42$$

$$h_1=\sigma(0.42)=0.60$$

$$\boxed{\Longrightarrow h\left(\mathbf{x}^{(1)}\right)=0.60}$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

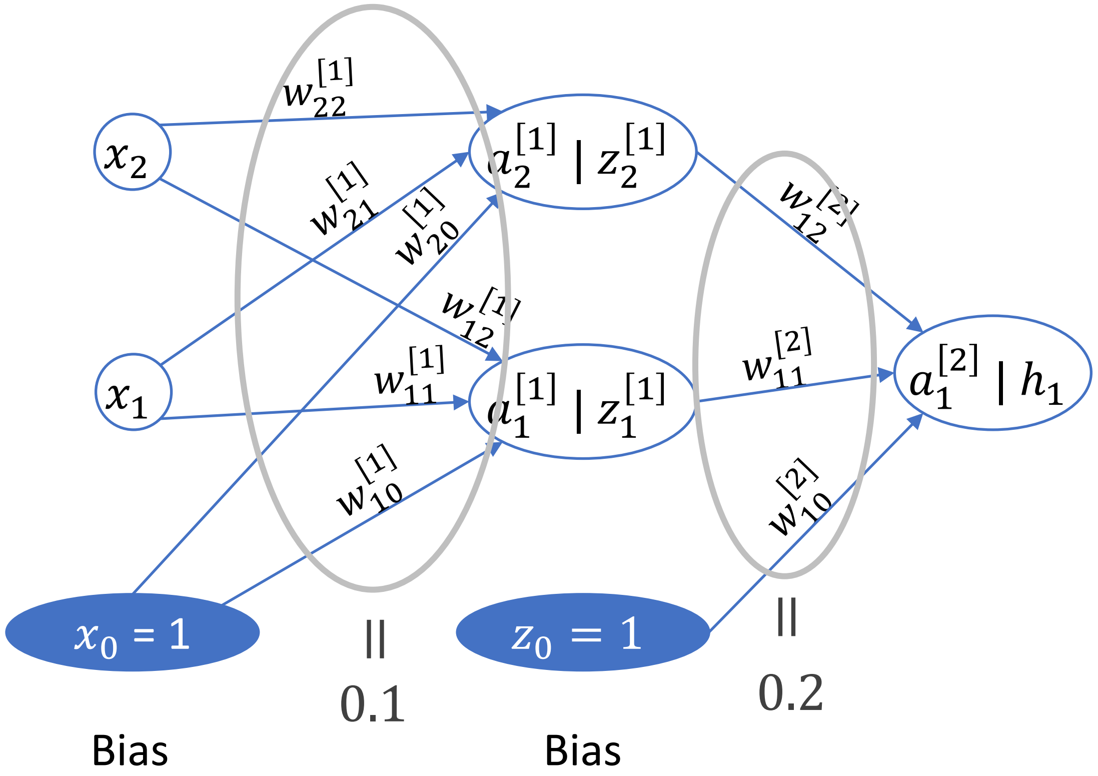
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(d) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow \boxed{h\left(\mathbf{x}^{(1)}\right) = 0.60}$

with $h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n = -y^{(n)} \ln h\left(\mathbf{x}^{(n)}\right) - \left(1 - y^{(n)}\right) \ln\left(1 - h\left(\mathbf{x}^{(n)}\right)\right)$$



(i) $\mathbb{E}_1 = ?$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

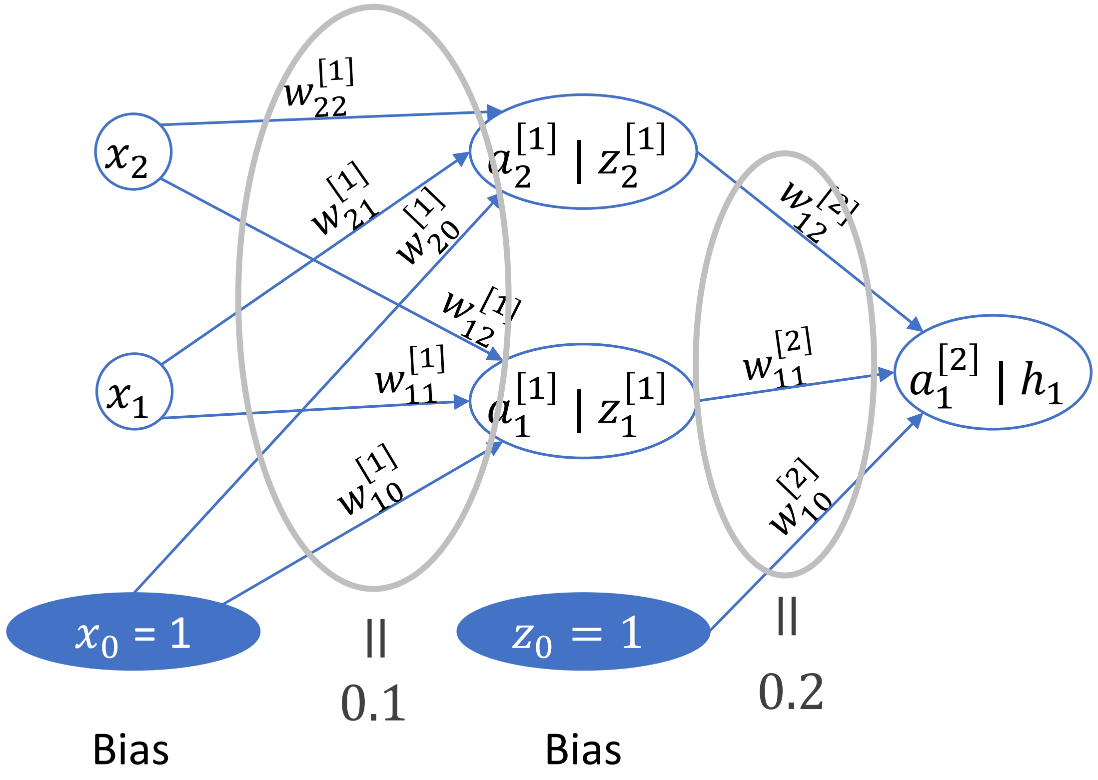
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(d) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow \boxed{h\left(\mathbf{x}^{(1)}\right) = 0.60}$

with $\quad h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right) \quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n = -y^{(n)} \ln h\left(\mathbf{x}^{(n)}\right) - \left(1 - y^{(n)}\right) \ln\left(1 - h\left(\mathbf{x}^{(n)}\right)\right)$$



(i) $\mathbb{E}_1 = -1 \ln 0.60 - (1-1)\ln(1-0.60) = 0.51$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

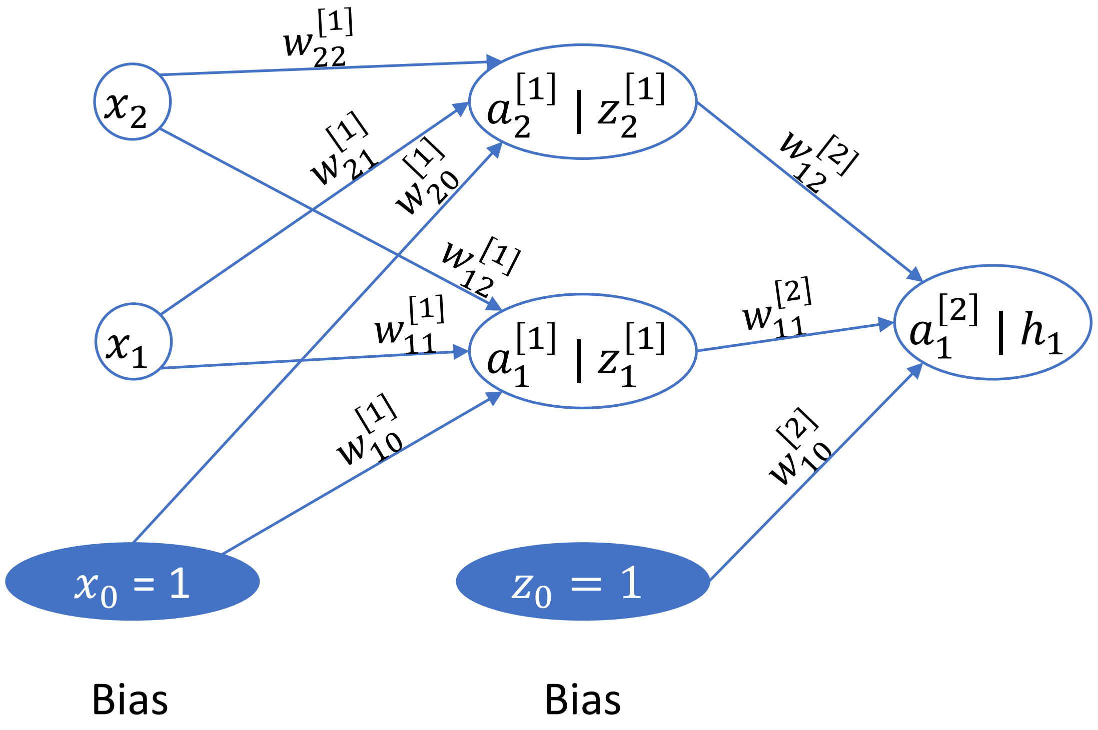
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(d) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1$ ⟶ $h\left(\mathbf{x}^{(1)}\right) = 0.60$

with $\quad h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n = -y^{(n)} \ln h\left(\mathbf{x}^{(n)}\right) - \left(1-y^{(n)}\right) \ln\left(1 - h\left(\mathbf{x}^{(n)}\right)\right)$$



(ii) $\partial_1^{[2]} = ?$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}} = ?$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

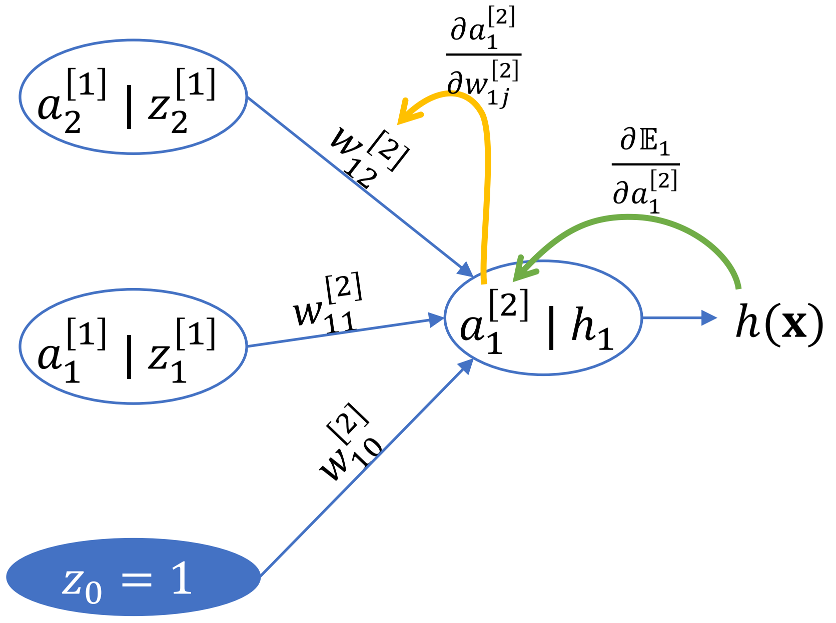
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(d) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.60$

with $\quad h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ $\quad$ and $\quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n = -y^{(n)} \ln h\left(\mathbf{x}^{(n)}\right) - \left(1-y^{(n)}\right) \ln\left(1-h\left(\mathbf{x}^{(n)}\right)\right)$$



Bias

<u>THEORETICALLY:</u> $\quad \dfrac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \boxed{\dfrac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\dfrac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}}$ $\quad$ *CHAIN RULE*

$$\mathbb{E}_1 = -y^{(n)} \ln h\left(\mathbf{x}^{(n)}\right) - \left(1-y^{(n)}\right) \ln\left(1-h\left(\mathbf{x}^{(n)}\right)\right)$$

$$= -y^{(n)} \ln h_1\left(a_1^{[2]}\right) - \left(1-y^{(n)}\right) \ln\left(1-h_1\left(a_1^{[2]}\right)\right)$$

LOCAL ERROR $\quad \partial_1^{[2]}$ $\quad \boxed{\dfrac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}} = ?$

$$\boxed{\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}} = ?$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(d) $\mathbf{x}^{(1)} = \binom{0.3}{0.5}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.60$

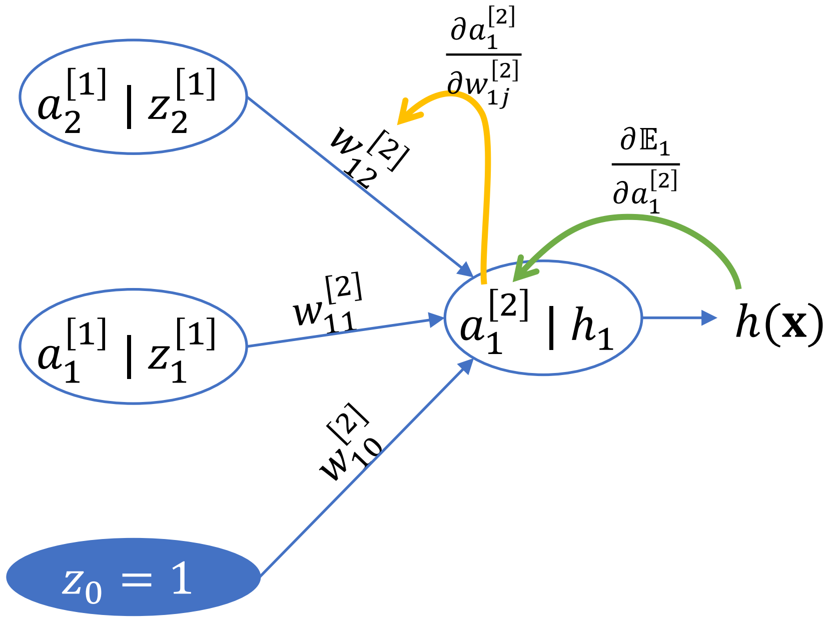
with $h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n = -y^{(n)} \ln h\left(\mathbf{x}^{(n)}\right) - \left(1-y^{(n)}\right) \ln\left(1 - h\left(\mathbf{x}^{(n)}\right)\right)$$

<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}}$$

CHAIN RULE



Bias

$$h(a) = \sigma(a)$$

LOCAL ERROR $\partial_1^{[2]}$

$$\boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}} = -\frac{y^{(n)}}{h_1\left(a_1^{[2]}\right)} h_1'\left(a_1^{[2]}\right) + \frac{1-y^{(n)}}{1-h_1\left(a_1^{[2]}\right)} h_1'\left(a_1^{[2]}\right)$$

$$= -y^{(n)}\left(1 - h_1\left(a_1^{[2]}\right)\right) + \left(1-y^{(n)}\right) h_1\left(a_1^{[2]}\right)$$

$$= -y^{(n)} + h_1\left(a_1^{[2]}\right)$$

$$\boxed{\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}} = z_j^{[1]}$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

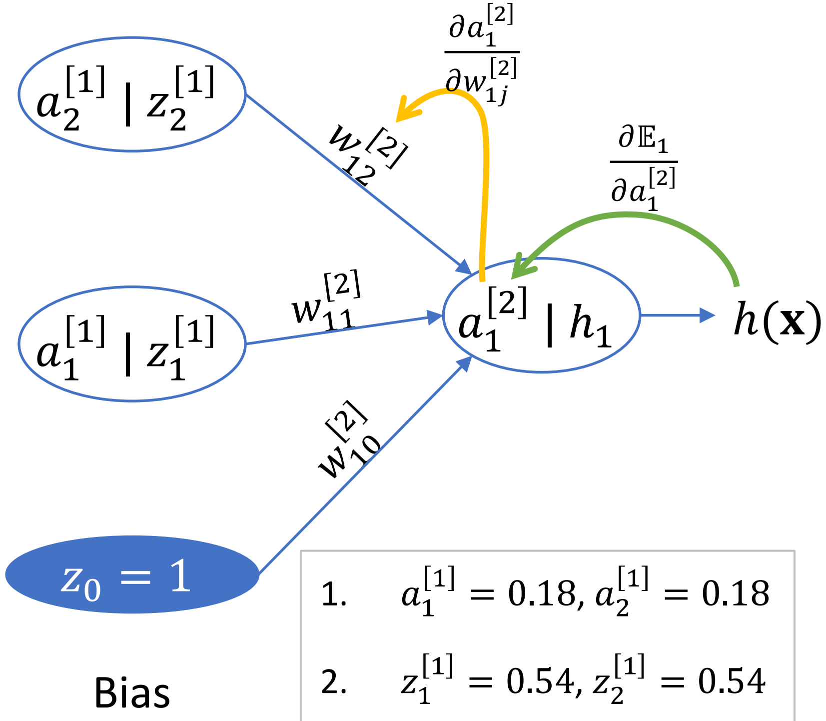
(d) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1 \longrightarrow h\left(\mathbf{x}^{(1)}\right)=0.60$

with $\quad h_1\left(a_1^{[2]}\right)=\sigma\left(a_1^{[2]}\right)$ and $\quad z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n = -y^{(n)}\ln h\left(\mathbf{x}^{(n)}\right) - \left(1-y^{(n)}\right)\ln\left(1-h\left(\mathbf{x}^{(n)}\right)\right)$$

<u>THEORETICALLY:</u>

$$\frac{\partial \mathbb{E}_1}{\partial w_{1j}^{[2]}} = \boxed{\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}}}\boxed{\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}}} = \boxed{\left(h_1\left(a_1^{[2]}\right)-y^{(n)}\right)}\boxed{z_j^{[1]}}$$



1. $a_1^{[1]} = 0.18,\ a_2^{[1]} = 0.18$
2. $z_1^{[1]} = 0.54,\ z_2^{[1]} = 0.54$
3. $a_1^{[2]} = 0.42$

LOCAL ERROR $\ \partial_1^{[2]} = h_1\left(a_1^{[2]}\right) - y^{(n)} = \sigma(0.42) - 1 = -0.4$

$$\frac{\partial \mathbb{E}_1}{\partial w_{10}^{[2]}} = -0.4 \cdot 1 = -0.4$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{11}^{[2]}} = -0.21$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{12}^{[2]}} = -0.21$$

EXERCICE 1

(d) $\mathbf{x}^{(1)}=\begin{pmatrix}0.3\\0.5\end{pmatrix}$ with $y^{(1)}=1 \longrightarrow h(\mathbf{x}^{(1)}) = 0.60$

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

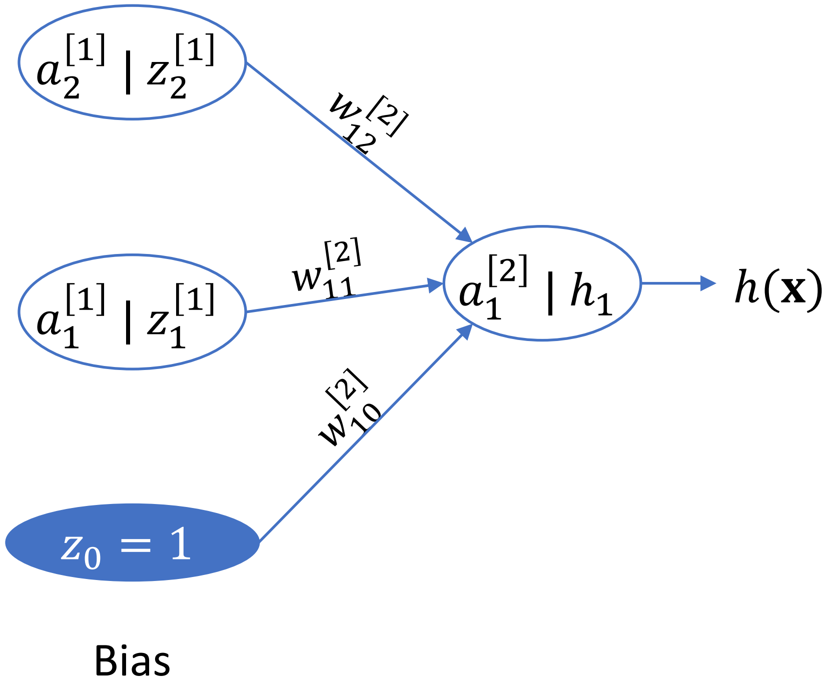
$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

with $\quad h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n = -y^{(n)} \ln h\left(\mathbf{x}^{(n)}\right) - \left(1-y^{(n)}\right) \ln\left(1-h\left(\mathbf{x}^{(n)}\right)\right)$$



$$\frac{\partial \mathbb{E}_1}{\partial w_{10}^{[2]}} = -0.4 \qquad \frac{\partial \mathbb{E}_1}{\partial w_{11}^{[2]}} = -0.2 \qquad \frac{\partial \mathbb{E}_1}{\partial w_{12}^{[2]}} = -0.2$$

Gradient descent for the weights of the output layer, with $\alpha = 0.5$?

$$w_{10}^{[2]t+1} = w_{10}^{[2]t} - 0.5 \cdot (-0.4) = 0.2 + 0.2 = 0.6$$

$$w_{11}^{[2]t+1} = w_{11}^{[2]t} - 0.5 \cdot (-0.2) = 0.2 + 0.1 = 0.3$$

$$w_{12}^{[2]t+1} = w_{12}^{[2]t} - 0.5 \cdot (-0.2) = 0.2 + 0.1 = 0.3$$

EXERCICE 1

$$\sigma(a) = \frac{1}{1+\exp(-a)}$$

$$\sigma'(a) = \sigma(a)(1-\sigma(a))$$

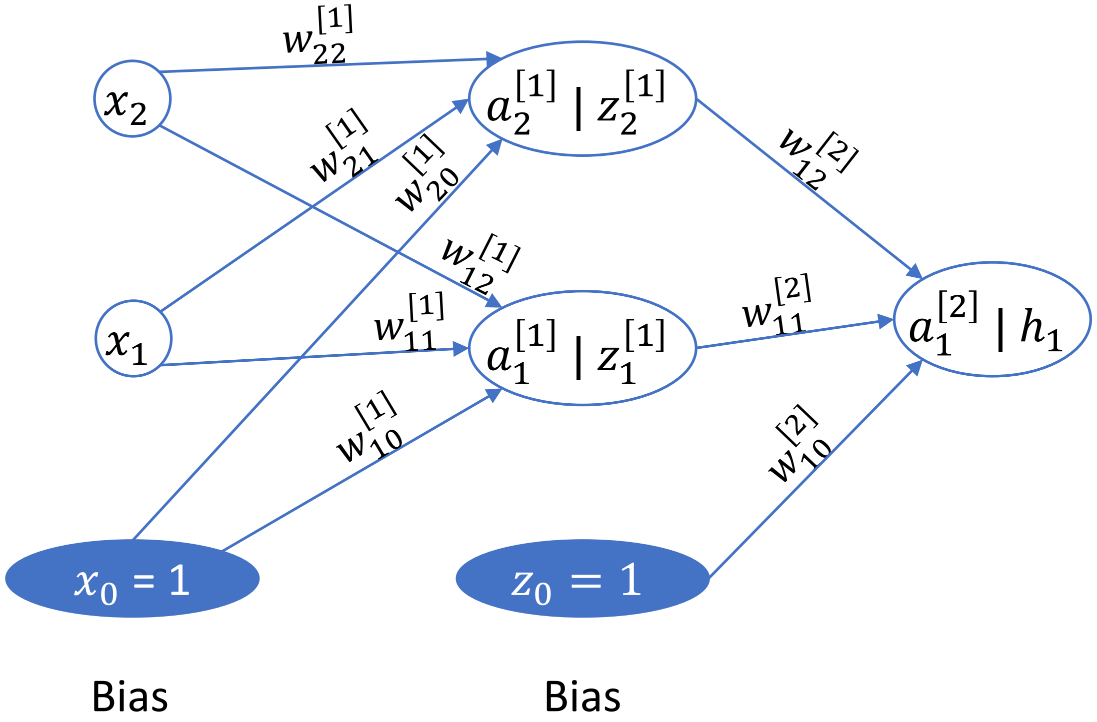
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \qquad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

(d) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1 \longrightarrow h\left(\mathbf{x}^{(1)}\right) = 0.60$

with $h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

$$\mathbb{E}_n = -y^{(n)} \ln h\left(\mathbf{x}^{(n)}\right) - \left(1-y^{(n)}\right) \ln\left(1 - h\left(\mathbf{x}^{(n)}\right)\right)$$



(iii) $\partial_1^{[1]} = ?, \partial_2^{[1]} = ?$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[1]}} = ?$

SAME AS BEFORE