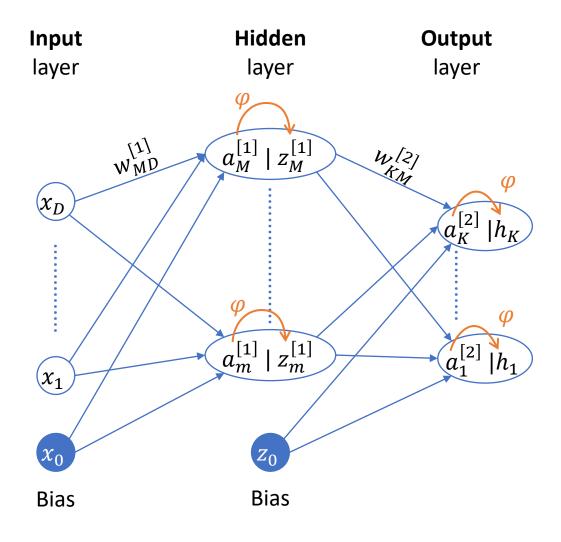
NEURAL NETWORKS REVIEW

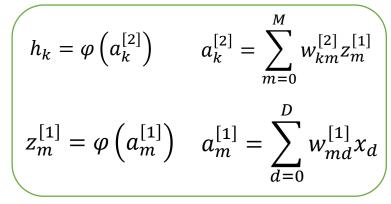


- $h_k=arphi\left(a_k^{[2]}\right)$, where $arphi(\cdot)$ is an **activation function** $\varphi(a)=\sigma(a)=\frac{1}{1+\exp(-a)} \text{ logistic function}$ $\varphi(a)=\tanh(a)=\frac{e^a-e^{-a}}{e^a+e^{-a}} \text{ hyperbolic tangent}$ $\varphi(a)=\max\{0,a\} \text{ Rectified Linear Unit (ReLU)}$
- $a_k^{[2]} = \sum_{m=0}^M w_{km}^{[2]} z_m$
 - $z_m = \varphi\left(a_m^{[1]}\right)$

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

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.

Input	Hidden	Output
layer	layer	layer

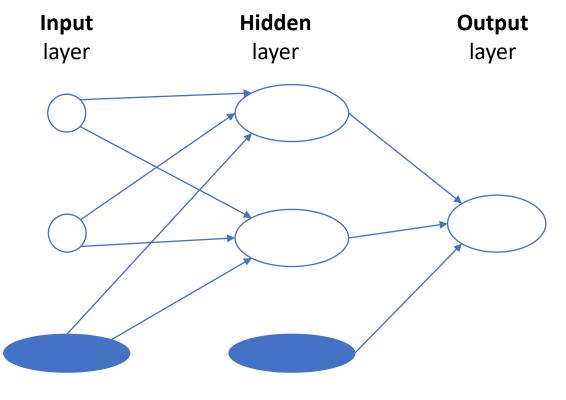
We have a fully connected network with the following characteristics:

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

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

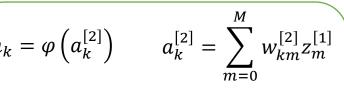
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \quad a_m^{[1]} = \sum_{d=0}^D w_{md}^{[1]} x_d$$

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



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.



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

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

Input layer	Hidden layer	Output layer

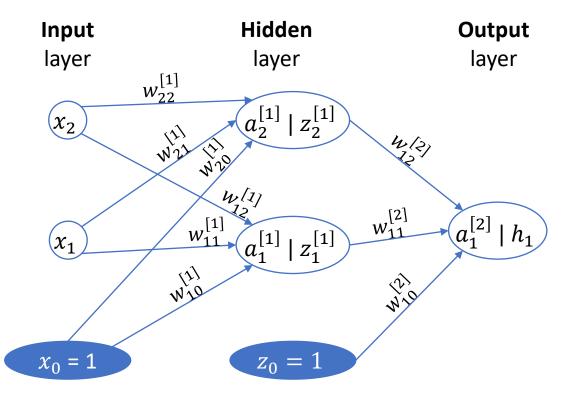
We have a fully connected network with the following characteristics:

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

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

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

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



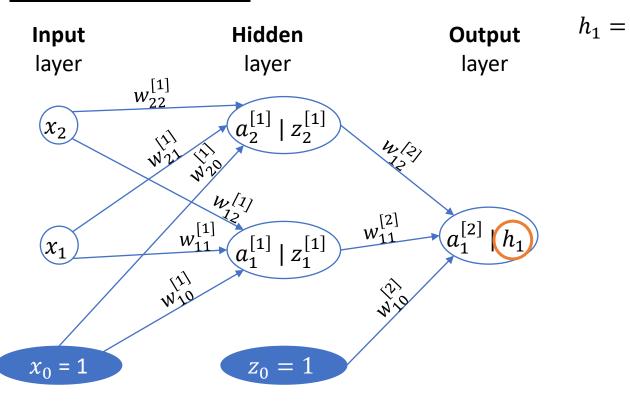
We have a fully connected network with the following characteristics:

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

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

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

(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.

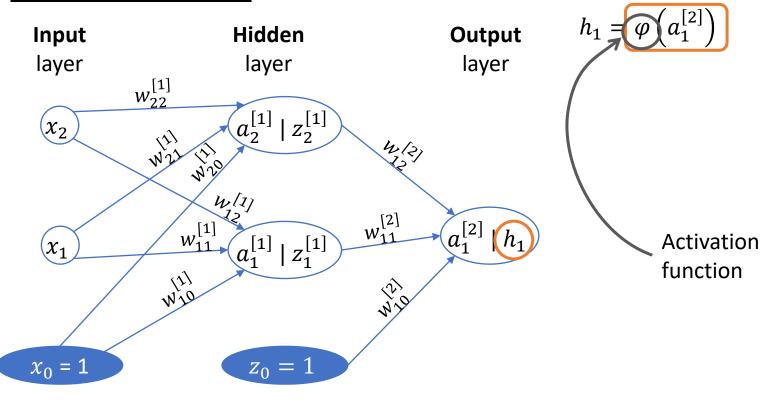


We have a fully connected network with the following characteristics:

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

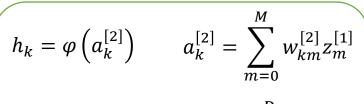
$$h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^M w_{km}^{[2]} z_m^{[1]}$$
 $z_m^{[1]} = \varphi\left(a_k^{[1]}\right) \qquad a_k^{[1]} = \sum_{m=0}^D w_{km}^{[1]} x_m^{[1]}$

(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.



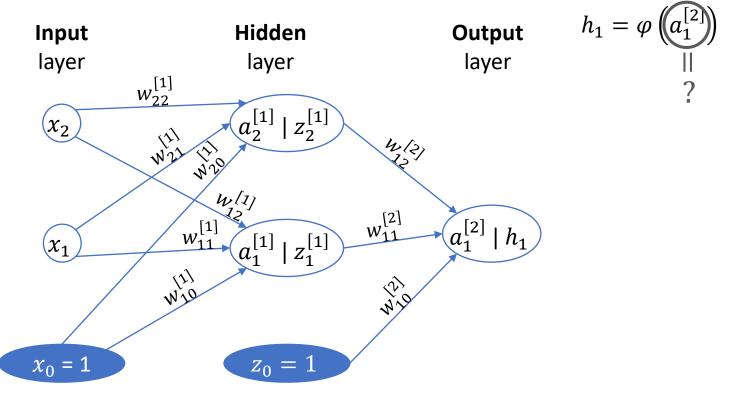
We have a fully connected network with the following characteristics:

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



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

(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.



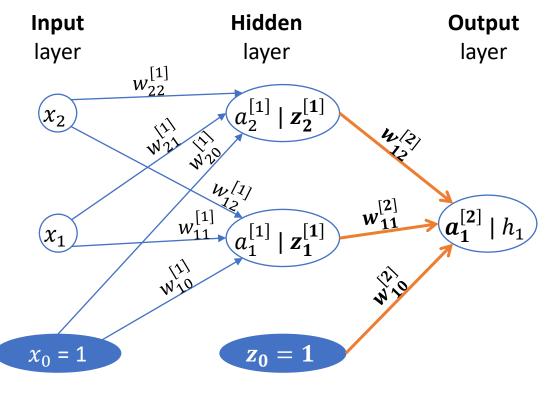
We have a fully connected network with the following characteristics:

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

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

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

(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.

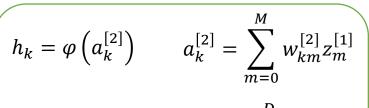


$$h_{1} = \varphi \left(a_{1}^{[2]}\right)$$

$$\sum_{m=0}^{2} w_{1m}^{[2]} z_{m}^{[1]} = w_{10}^{[2]} z_{0}^{[1]} + w_{11}^{[2]} z_{1}^{[1]} + w_{12}^{[2]} z_{2}^{[1]}$$

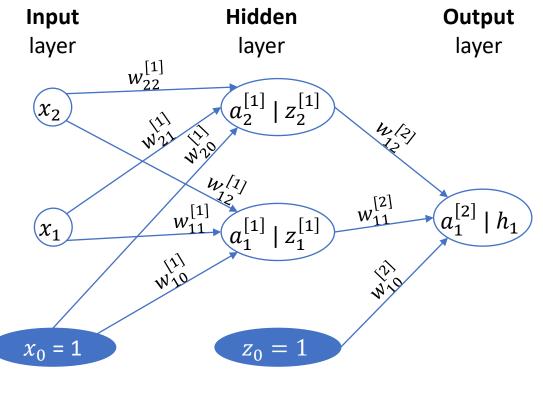
We have a fully connected network with the following characteristics:

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



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

(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.



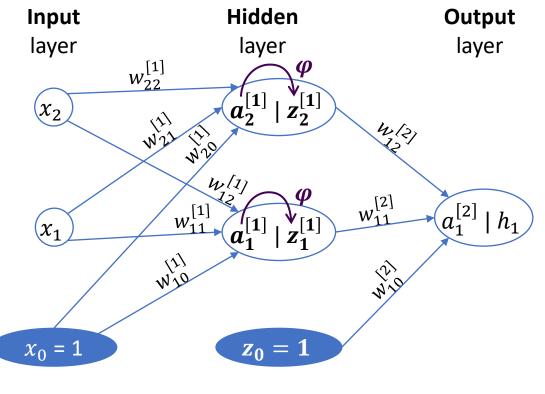
We have a fully connected network with the following characteristics:

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

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

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

(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.



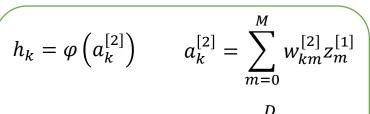
$$h_{1} = \varphi\left(a_{1}^{[2]}\right)$$

$$= \varphi\left(w_{10}^{[2]} \underbrace{z_{0}^{[1]}}_{||} + w_{11}^{[2]} \underbrace{z_{1}^{[1]}}_{||} + w_{12}^{[2]} \underbrace{z_{2}^{[1]}}_{||}\right)$$

$$\mathbf{1} \qquad \varphi\left(a_{1}^{[1]}\right) \qquad \varphi\left(a_{2}^{[1]}\right)$$

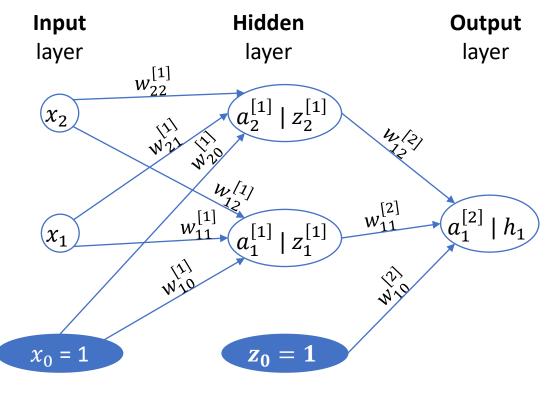
We have a fully connected network with the following characteristics:

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



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

(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.



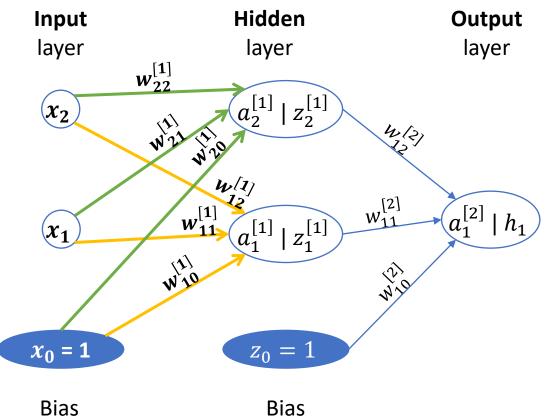
We have a fully connected network with the following characteristics:

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

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

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

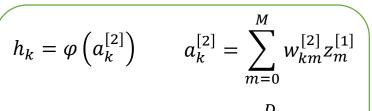
(b) Put the name at each element of the graph and write, as a closed formula, the output in function of all the elements of the network.



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

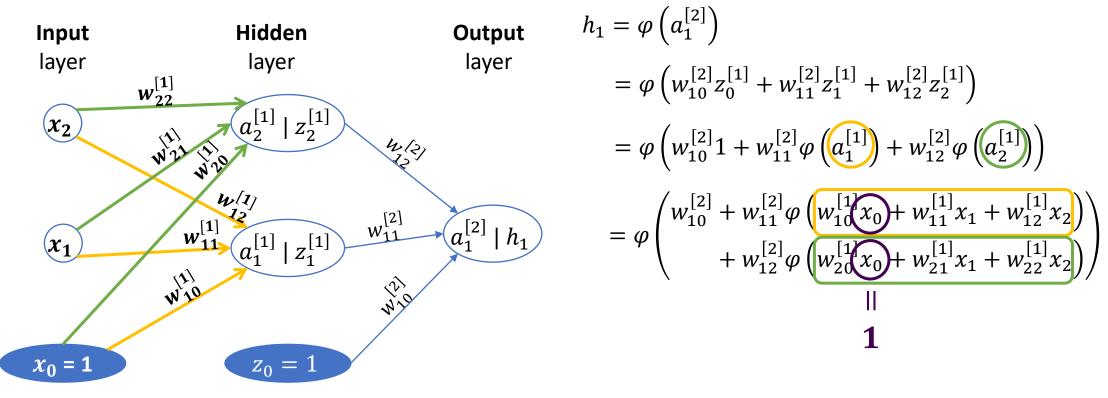
We have a fully connected network with the following characteristics:

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



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

(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.



Bias

We have a fully connected network with the following characteristics:

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

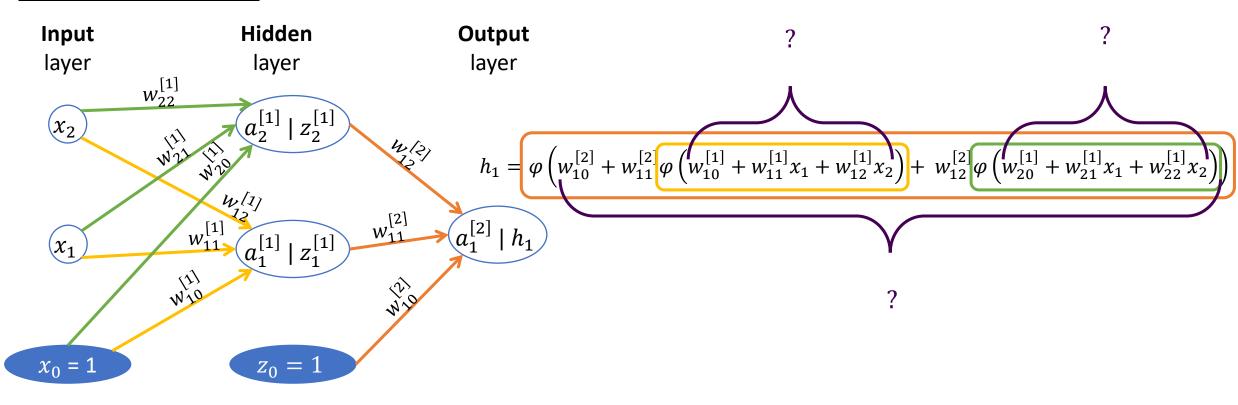
Bias

The output layer has one neuron.

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

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

(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.



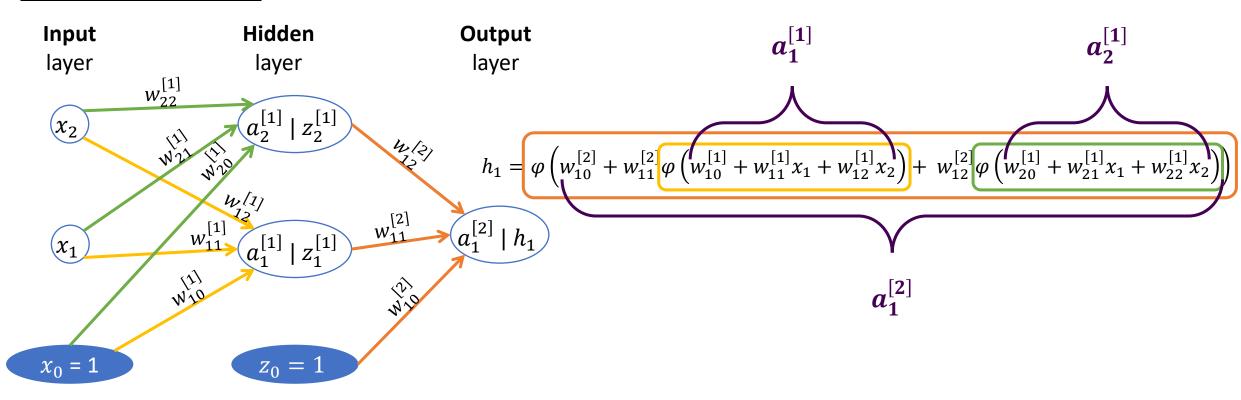
We have a fully connected network with the following characteristics:

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

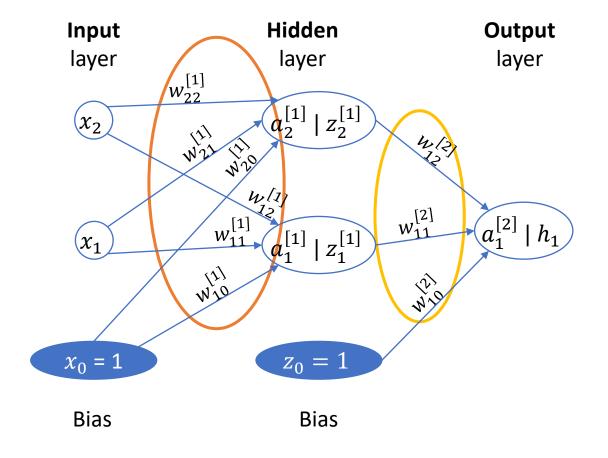
$$h_k = \varphi\left(a_k^{[2]}\right)$$
 $a_k^{[2]} = \sum_{m=0}^M w_{km}^{[2]} z_m^{[1]}$

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

(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.



(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 $h(\mathbf{x}^{(1)})$ when all the activation function are sigmoids except in the output layer, where h(a) = a.



NEW INFORMATION

$$\begin{aligned} w_{md}^{[1]} &= 0.1 \quad \forall m, d \\ w_{1m}^{[2]} &= 0.2 \quad \forall m \\ h_1 &= \varphi\left(a_1^{[2]}\right) = a_1^{[2]} \\ z_m &= \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right) = \frac{1}{1 + \exp\left(-a_m^{[1]}\right)} \\ \text{with } \sigma'(a) &= \sigma(a)(1 - \sigma(a)) \end{aligned}$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = ?$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

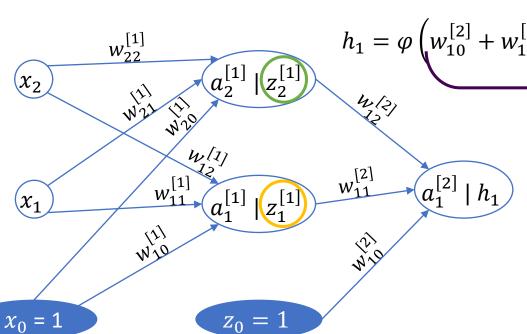
$$\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^M w_{km}^{[2]} z_m^{[1]} \right)$$

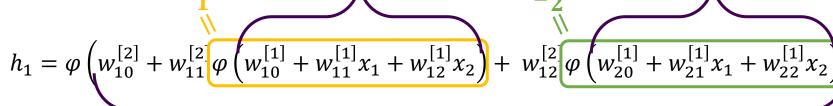
$$(a) = \sigma(a)(1 - \sigma(a))$$

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

 $a_2^{[1]}$







 $a_1^{[2]}$

 $a_1^{[1]}$

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

Bias

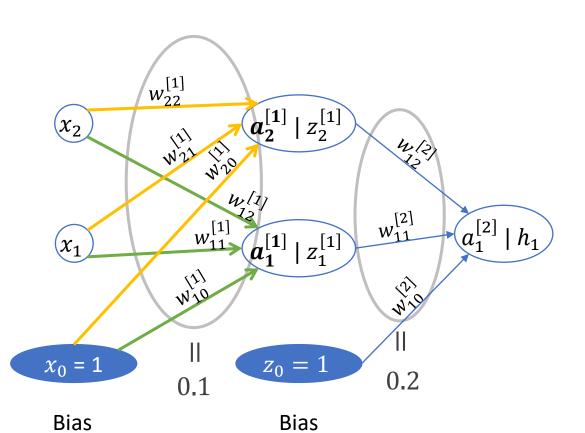
$$\sigma(a) = \frac{1}{1 + \exp(-a)} \begin{cases} h_k = \varphi\left(a_k^{[2]}\right) & a_k^{[2]} = \sum_{m=0}^M w_{km}^{[2]} z_m^{[1]} \\ \sigma'(a) = \sigma(a)(1 - \sigma(a)) \end{cases}$$

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = ?$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

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



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

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = ?$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

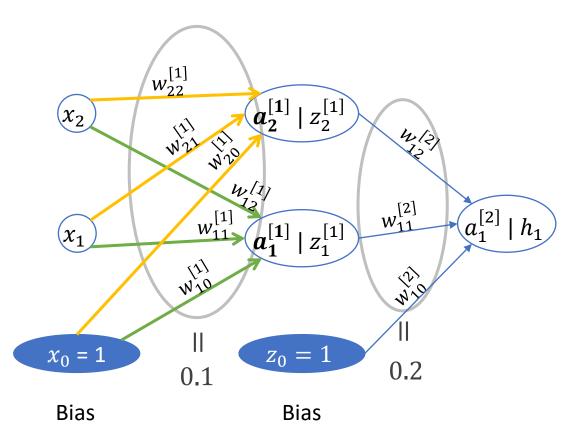
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

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

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

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

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



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

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

$$a_{2}^{[1]} = \sum_{d=0}^{2} w_{2d}^{[1]} x_{d} = w_{20}^{[1]} x_{0} + w_{21}^{[1]} x_{1} + w_{22}^{[1]} x_{2}$$
$$= 0.1 \cdot 1 + 0.1 \cdot 0.3 + 0.1 \cdot 0.5 = 0.18$$

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

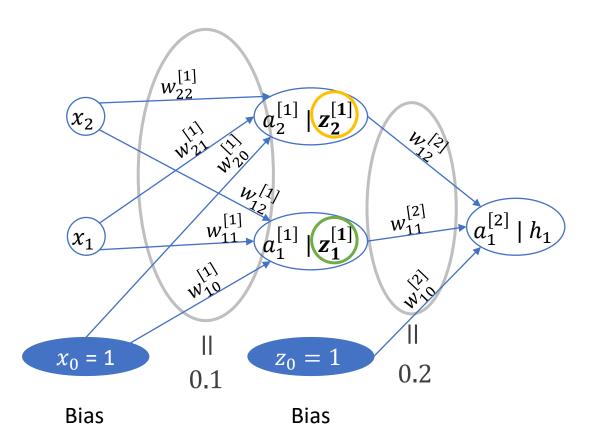
with
$$h_1\left(a_1^{[2]}\right)=a_1^{[2]}$$
 and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

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

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

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

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



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 $a_1^{[2]}$.
- Compute h_1

$$\mathbf{z}_{1}^{[1]} = \varphi\left(a_{1}^{[1]}\right) = \sigma\left(a_{1}^{[1]}\right) = ?$$

$$\mathbf{z}_{2}^{[1]} = \varphi\left(a_{2}^{[1]}\right) = \sigma\left(a_{2}^{[1]}\right) = ?$$

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

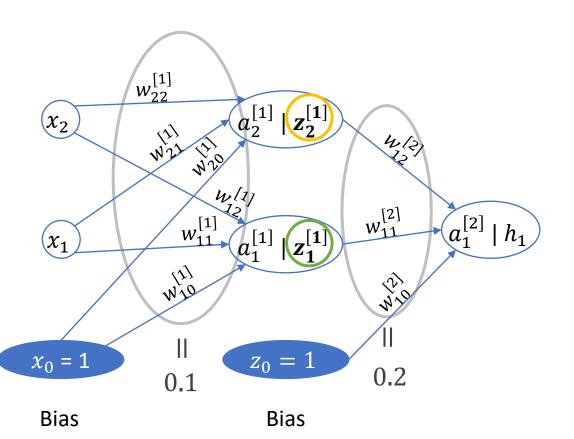
with
$$h_1\left(a_1^{[2]}\right)=a_1^{[2]}$$
 and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

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

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

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

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



- 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 $a_1^{[2]}$.
- Compute h_1

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

$$= \sigma(0.18) = \frac{1}{1 + \exp(-0.18)} = 0.54$$

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

$$= \sigma(0.18) = \frac{1}{1 + \exp(-0.18)} = 0.54$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = ?$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

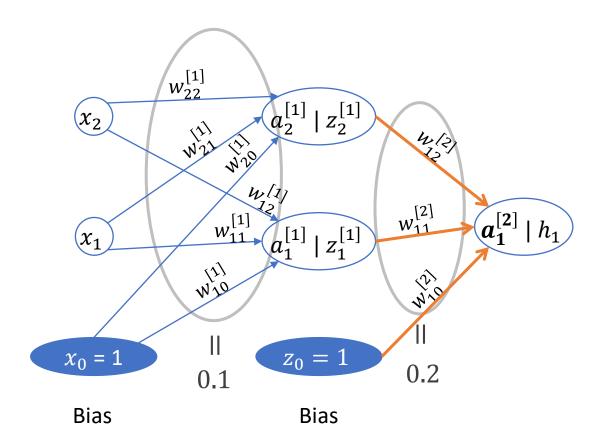
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

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

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

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

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



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 $a_1^{[2]}$.
- Compute h_1

$$a_{1}^{[2]} = \sum_{m=0}^{2} w_{1m}^{[2]} z_{m}^{[1]} = w_{10}^{[2]} z_{0}^{[1]} + w_{11}^{[2]} z_{1}^{[1]} + w_{12}^{[2]} z_{2}^{[1]} = ?$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = ?$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

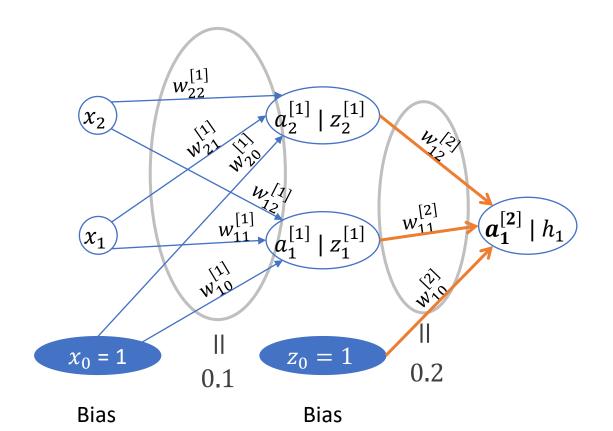
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

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

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

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

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



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 $a_1^{[2]}$.
- Compute h_1

$$a_{1}^{[2]} = \sum_{m=0}^{2} w_{1m}^{[2]} z_{m}^{[1]} = w_{10}^{[2]} z_{0}^{[1]} + w_{11}^{[2]} z_{1}^{[1]} + w_{12}^{[2]} z_{2}^{[1]}$$
$$= 0.2 \cdot 1 + 0.2 \cdot 0.54 + 0.2 \cdot 0.54 = 0.42$$

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

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

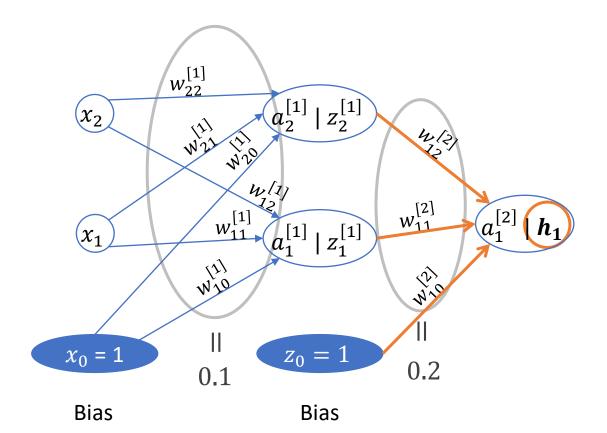
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

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

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

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

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



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 $a_1^{[2]}$.
- Compute h_1

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

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

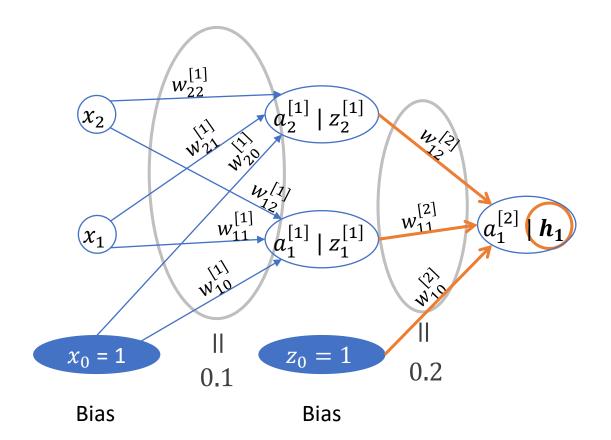
with
$$h_1\left(a_1^{[2]}\right)=a_1^{[2]}$$
 and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

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

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

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

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



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
$$h_1$$

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

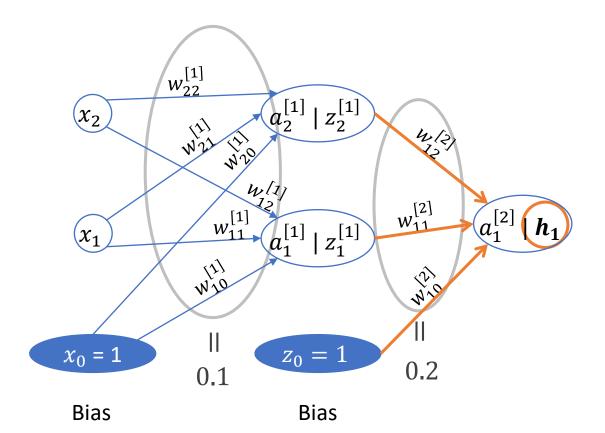
(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = ?$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

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

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

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

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



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
$$h_1$$

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

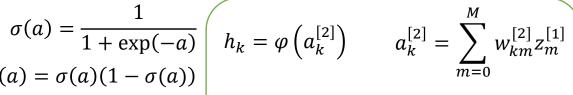
$$\Rightarrow h(\mathbf{x}^{(1)}) = 0.42$$

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ \longrightarrow $h(\mathbf{x}^{(1)}) = 0.42$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

(i) Calculate the error using LS Error.



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

LEAST SQUARES ERROR

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

$$x_{2}$$

$$x_{2}$$

$$x_{3}$$

$$x_{4}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{2}$$

$$x_{1}$$

$$x_{1}$$

$$x_{2}$$

$$x_{2}$$

$$x_{3}$$

$$x_{4}$$

$$x_{2}$$

$$x_{3}$$

$$x_{4}$$

$$x_{4}$$

$$x_{4}$$

$$x_{5}$$

$$x_{5}$$

$$x_{5}$$

$$x_{5}$$

$$x_{5}$$

$$x_{5}$$

$$x_{5}$$

$$x_{5$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ \longrightarrow $h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

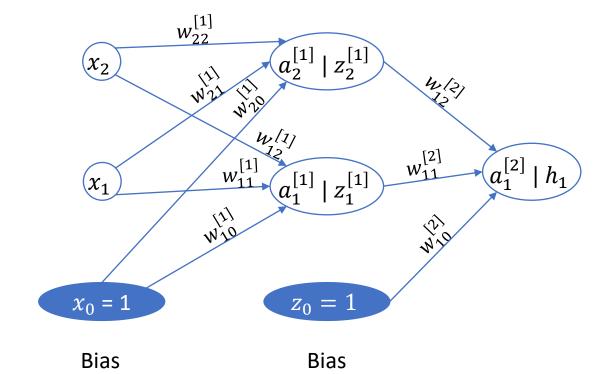
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

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

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

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

(i) Calculate the error using LS Error.



LEAST SQUARES ERROR

$$\mathbb{E}(\mathbf{x}^{(1)}) = \frac{1}{2} \sum_{n=1}^{N} (h(\mathbf{x}^{(n)}) - y^{(n)})^{2}$$
$$= \frac{1}{2} (h(\mathbf{x}^{(1)}) - y^{(1)})^{2} = \frac{1}{2} (0.42 - 1)^{2} = 0.17$$

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

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

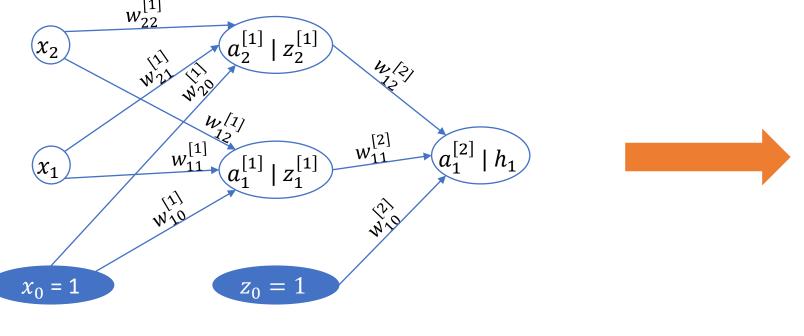
EXERCICE 1
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$
 (c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $\mathbf{y}^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$ $z_m^{[1]} = \varphi\left(a_m^{[1]}\right)$ $z_m^{[1]} = \varphi\left(a_m^{[1]}\right)$ (ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{md}^{[2]}}$. Draw only the part of the graph and the relations

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

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?



Bias

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

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

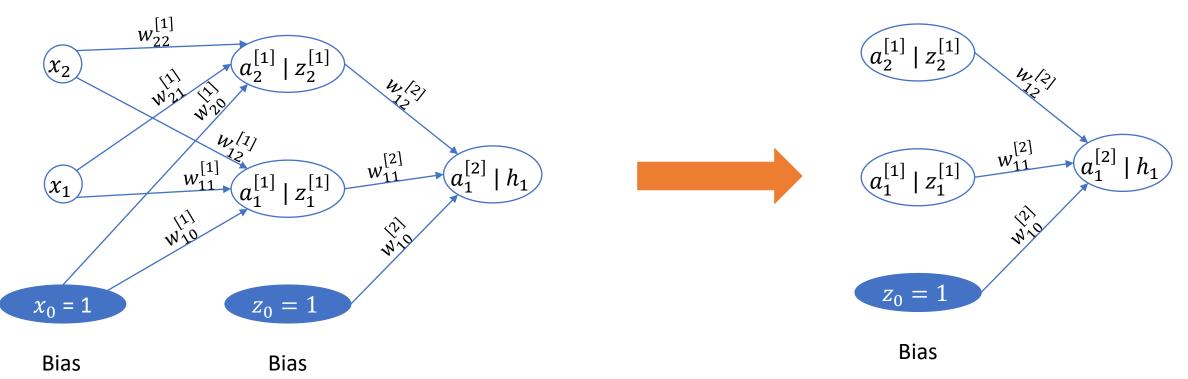
(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$

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

EXERCICE 1
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$
 (c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $\mathbf{y}^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ 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 $\frac{\partial \mathbb{E}_1}{\partial w^{[2]}}$. Draw only the part of the graph and the relations

(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?



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

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

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

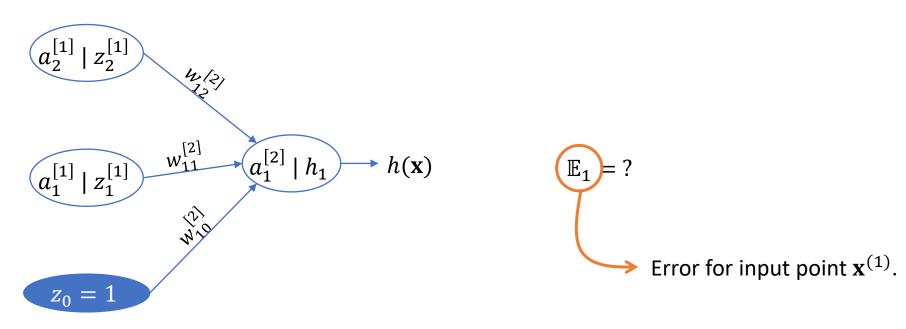
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \quad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

EXERCICE 1
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$
 (c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ 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 $\frac{\partial \mathbb{E}_1}{\partial w^{[2]}}$.

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

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

BACK-PROPAGATION



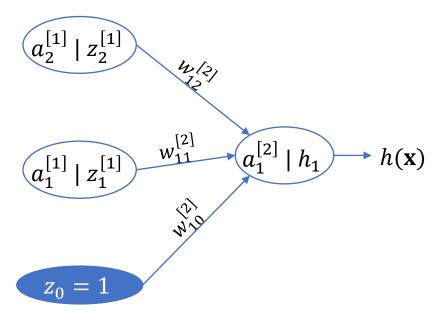
(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

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

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

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

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 $\mathbf{x}^{(1)}$.

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

 $h(\mathbf{x}^{(1)})$ 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.

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

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

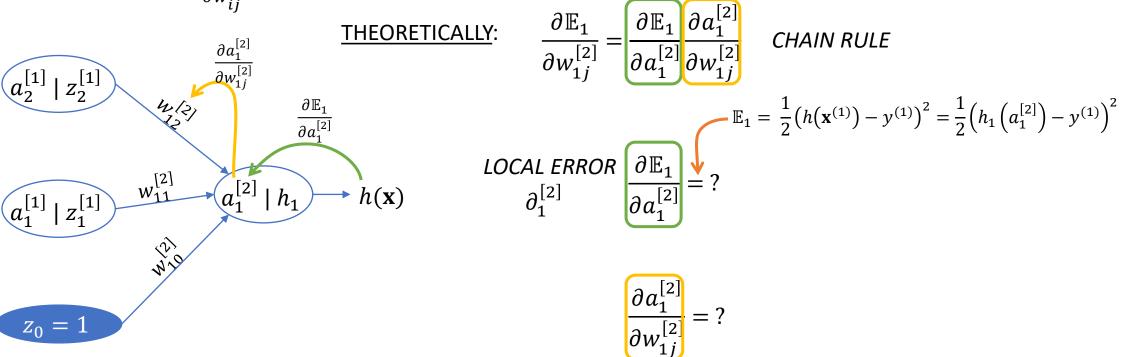
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \quad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$$

with
$$h_1(a_1^{[2]}) = a_1^{[2]}$$

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.



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

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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

with
$$h_1(a_1^{[2]}) = a_1^{[2]}$$

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 $\frac{\partial \mathbb{E}_1}{\partial w_{ii}^{[2]}}$.

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

$$a_{2}^{[1]} \mid z_{2}^{[1]}$$

$$a_{1}^{[1]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid h_{1}$$

THEORETICALLY:

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

CHAIN RULE

$$E_{1} = \frac{1}{2} (h(\mathbf{x}^{(1)}) - y^{(1)})^{2} = \frac{1}{2} (h_{1} (a_{1}^{[2]}) - y^{(1)})^{2}$$

$$UCCAL \ ERROR \ \partial_{1}^{[2]} = \left(h_{1} (a_{1}^{[2]}) - y^{(1)} \right) \cdot h'_{1} (a_{1}^{[2]}) = h_{1} (a_{1}^{[2]}) - y^{(1)}$$

$$h_{1}(a) = a$$

$$\frac{\partial a_{1}^{[2]}}{\partial w_{1j}^{[2]}} = \frac{\partial}{\partial w_{1j}^{[2]}} \left\{ w_{10}^{[2]} + w_{11}^{[2]} z_{1}^{[1]} + w_{12}^{[2]} z_{2}^{[1]} \right\} = z_{j}^{[1]}$$

$$a_{1}^{[2]}$$

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

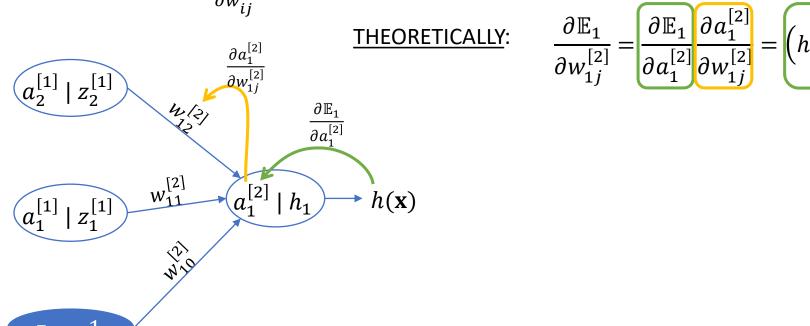
$$\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^M w_{km}^{[2]} z_m^{[1]} \right)$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

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

(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.



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

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

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{k=0}^{M} w_{km}^{[2]} z_m^{[1]} \right)$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

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

with
$$h_1(a_1^{[2]}) = a_1^{[2]}$$

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

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

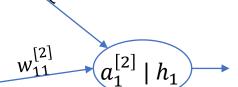
(ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ii}^{[2]}}$.

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

THEORETICALLY:

$$\frac{\partial \mathbb{E}_{1}}{\partial w_{1i}^{[2]}} = \frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} \frac{\partial a_{1}^{[2]}}{\partial w_{1i}^{[2]}} = \left(h_{1}\left(a_{1}^{[2]}\right) - y^{(1)}\right) \cdot z_{j}^{[1]}$$

$$a_{2}^{[1]} \mid z_{2}^{[1]}$$



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

 $z_0 = 1$

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

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

3.
$$a_1^{[2]} = 0.42$$

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

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

$$\overline{-a}$$
 h

$$a_k^{[2]} = \sum_{k=1}^{M} w_{km}^{[2]} z_k^{[2]}$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

$$= 0.42$$

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

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

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

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

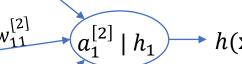
Compute $\partial_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ij}^{[2]}}$ for all the parameters involved.

THEORETICALLY:

$$\frac{\partial \mathbb{E}_{1}}{\partial w_{1j}^{[2]}} = \frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} \frac{\partial a_{1}^{[2]}}{\partial w_{1j}^{[2]}} = \left(h_{1} \left(a_{1}^{[2]} \right) - y^{(1)} \right) \cdot z_{j}^{[1]}$$

$$a_{2}^{[1]} \mid z_{2}^{[1]}$$

$$w_{11}^{[2]}$$



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

= $a_1^{[2]} - y^{(1)} = 0.42 - 1 = -0.58$

$$z_0 = 1$$

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

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

3.
$$a_1^{[2]} = 0.42$$

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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 $\frac{\partial \mathbb{E}_1}{\partial w_{ii}^{[2]}}$.

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

$$\partial_1^{[2]} = -0.58$$

THEORETICALLY:

$$\frac{\partial \mathbb{E}_{1}}{\partial w_{1j}^{[2]}} = \frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} \frac{\partial a_{1}^{[2]}}{\partial w_{1j}^{[2]}} = \left(h_{1}\left(a_{1}^{[2]}\right) - y^{(1)}\right) \cdot z_{j}^{[1]}$$

$$a_{2}^{[1]} \mid z_{2}^{[1]}$$

$$a_{1}^{[1]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid h_{1}$$

$$h(\mathbf{x})$$

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

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

$$z_0 = 1$$

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

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

3.
$$a_1^{[2]} = 0.42$$

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

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

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{k=0}^{M} w_{km}^{[2]} z_m^{[1]} \right)$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

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

with
$$h_1\left(a_1^{12}\right)=a_1^{12}$$
 and $z_m=\varphi\left(a_m^{12}\right)=\sigma\left(a_m^{12}\right)$

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

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

-0.58

THEORETICALLY:

$$\frac{\partial \mathbb{E}_{1}}{\partial w_{1j}^{[2]}} = \frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} \frac{\partial a_{1}^{[2]}}{\partial w_{1j}^{[2]}} = \left(h_{1}\left(a_{1}^{[2]}\right) - y^{(1)}\right) \cdot z_{j}^{[1]}$$

$$a_{2}^{[1]} \mid z_{2}^{[1]}$$

$$a_{1}^{[1]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid h_{1}$$

$$h(\mathbf{x})$$

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

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

$$z_0 = 1$$
 1. $a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$

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

3.
$$a_1^{[2]} = 0.42$$

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

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

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

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

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

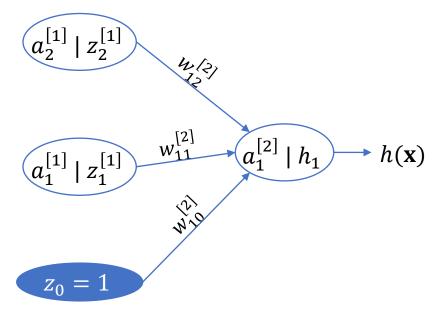
EXERCICE 1
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$
(c) $\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$ with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1\left(a_1^{[2]}\right) = a_1^{[2]}$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$ $z_m^{[1]} = \varphi\left(a_m^{[1]}\right)$ $z_m^{[1]} = \varphi\left(a_m^{[1]}\right)$ (ii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial y_m^{[2]}}$.

with
$$h_1(a_1^{[2]}) = a_1^{[2]}$$

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

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

Compute $\partial_1^{[2]}$ and $\frac{\partial \mathbb{E}_1}{\partial w_{ii}^{[2]}}$ 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 $\alpha = 0.5$?

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

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

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

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

$$\overline{a}$$
 h_k

$$a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

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

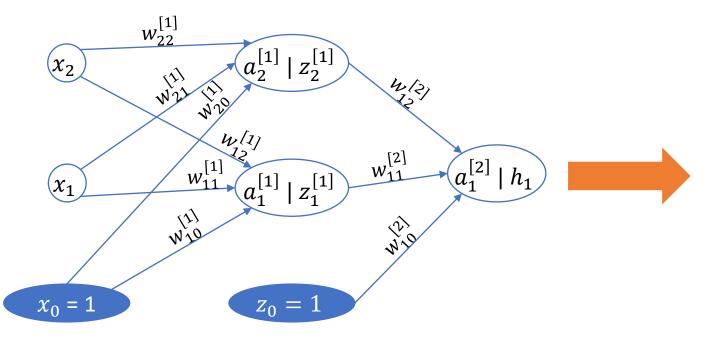
$$\sigma'(a) = \sigma(a)(1 - \sigma(a))$$

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

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

 $\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \quad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]} \right)$ $\text{(c) } \mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix} \text{ with } \mathbf{y}^{(1)} = 1 \longrightarrow h(\mathbf{x}^{(1)}) = 0.42 \quad \sigma'(a) = \sigma(a)(1 - \sigma(a))$ $\text{with } h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad \text{and} \quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$ $\text{(iii) We want to calculate the solution of the properties of the pro$ (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.



Bias

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

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

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

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

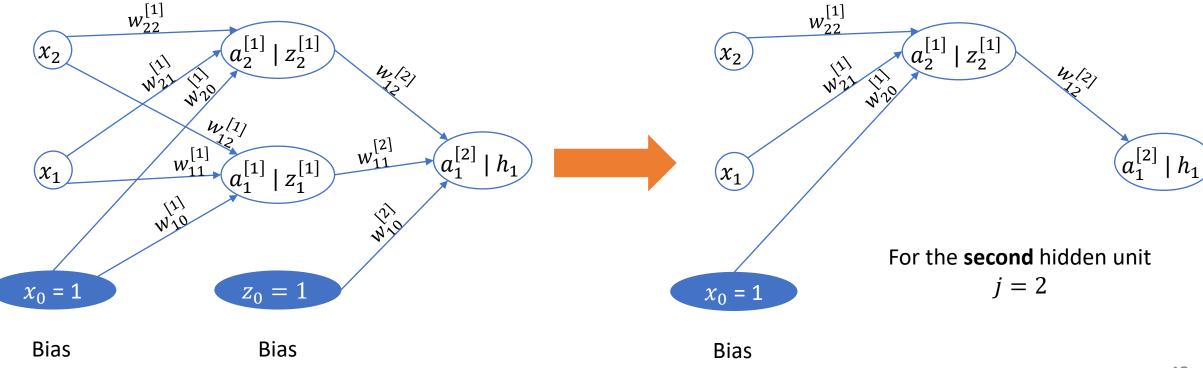
$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \quad a_m^{[1]} = \sum_{l=0}^{D} v_{l}^{[l]}$$

with
$$h_1(a_1^{[2]}) = a_1^{[2]}$$

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

 $\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \quad a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]} \right)$ $\text{(c) } \mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix} \text{ with } \mathbf{y}^{(1)} = 1 \longrightarrow h(\mathbf{x}^{(1)}) = 0.42 \quad \sigma'(a) = \sigma(a)(1 - \sigma(a))$ $\text{with } h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad \text{and} \quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$ $\text{(iii) We want to calculate the solution of the properties of the pro$ (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.



43

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

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

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

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

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

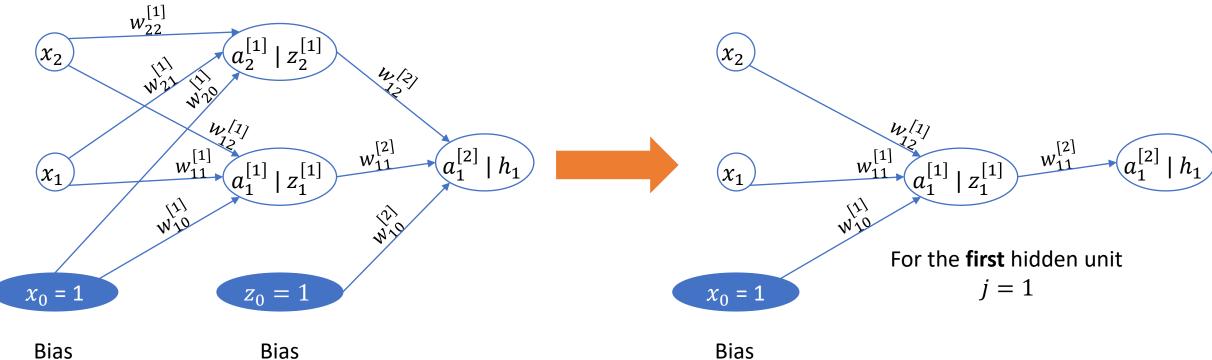
$$h_k = \varphi\left(a_k^{[2]}\right) \quad a_k^{[2]} = \sum_{m=0}^M w_{km}^{[2]} z_m^{[1]}$$

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

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \quad a_m^{[1]} = \sum_{d=0}^D w_{md}^{[1]} x_d$$
 (iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_1^{[1]}}$. Draw only the part of the graph and the relations involved.

(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.



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

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

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

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

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

$$\frac{\partial \mathbb{E}_1}{\partial a_m} = \frac{1}{1 + \exp(-a)}$$

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

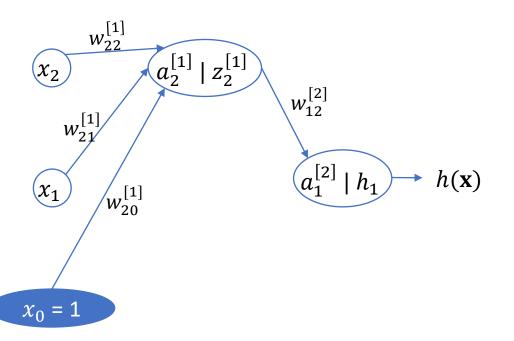
$$\sigma'(a) = \sigma(a)(1 - \sigma(a))$$

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

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

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



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

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

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

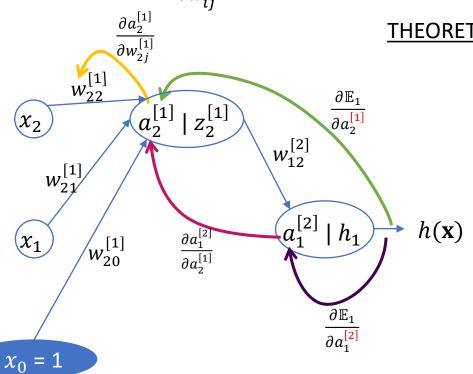
(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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 $\frac{\partial \mathbb{E}_1}{\partial w_{ii}^{[1]}}$.

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



THEORETICALLY:

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \underbrace{\frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}}}_{0} \underbrace{\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}}}_{0}$$

LOCAL ERROR
$$\partial_{2}^{[1]}$$

$$a_{1}^{[2]} = \sum_{m=0}^{2} w_{1m}^{[2]} z_{m}^{[1]} = \sum_{m=0}^{2} w_{1m}^{[2]} \varphi \left(a_{m}^{[1]} \right)$$

$$\mathbb{E}_{1} = \frac{1}{2} \left(h(\mathbf{x}^{(1)}) - y^{(1)} \right)^{2} = \frac{1}{2} \left(h_{1} \left(a_{1}^{[2]} \right) - y^{(1)} \right)^{2}$$

$$\frac{\partial \mathbb{E}_{1}}{\partial a_{2}^{[1]}} = \frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} \frac{\partial a_{1}^{[2]}}{\partial a_{2}^{[1]}}$$

$$\vdots$$

$$\vdots$$

$$\partial_{1}^{[2]}$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

 $\sigma(a) = \frac{1}{1 + \exp(-a)} \begin{pmatrix} h_k = \varphi\left(a_k^{[2]}\right) & a_k^{[2]} = \sum_{m=0}^{m} w_{km}^{[2]} z_m^{[1]} \\ z_m^{[1]} = \varphi\left(a_m^{[1]}\right) & a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d \end{pmatrix}$

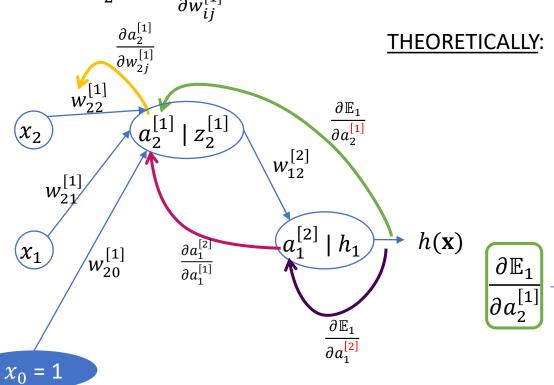
CHAIN RULE

(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.

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

 $\partial \mathbb{E}_1$



$$\frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} = ?$$

$$\mathbb{E}_{1} = \frac{1}{2} (h(\mathbf{x}^{(1)}) - y^{(1)})^{2} = \frac{1}{2} (h_{1} (a_{1}^{[2]}) - y^{(1)})^{2}$$

$$\frac{\partial a_{1}^{[2]}}{\partial a_{2}^{[1]}} = ?$$

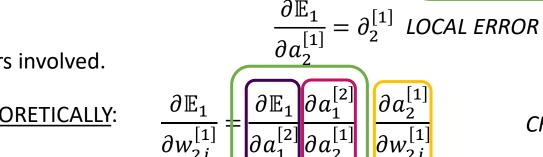
(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

 $\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^m w_{km}^{[2]} z_m^{[1]} \right)$ $z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \quad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$

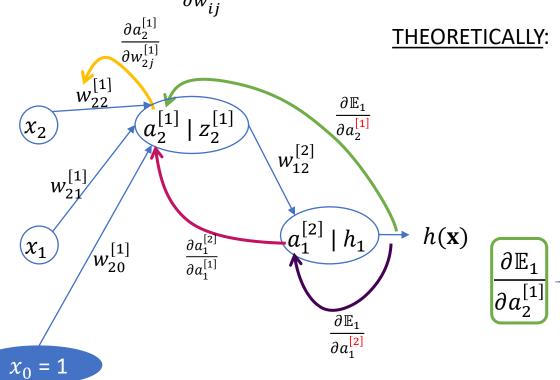
CHAIN RULE

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

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



 $\frac{\partial \mathbb{E}_1}{\partial w_{2i}^{[1]}}$



$$\frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} = \left(h_{1}\left(a_{1}^{[2]}\right) - y^{(1)}\right) \cdot h'_{1}\left(a_{1}^{[2]}\right) = h_{1}\left(a_{1}^{[2]}\right) - y^{(1)}$$

$$\mathbb{E}_{1} = \frac{1}{2}\left(h(\mathbf{x}^{(1)}) - y^{(1)}\right)^{2} = \frac{1}{2}\left(h_{1}\left(a_{1}^{[2]}\right) - y^{(1)}\right)^{2}$$

$$\frac{\partial a_{1}^{[2]}}{\partial a_{1}^{[1]}} = ?$$

 $x_0 = 1$

$$\begin{array}{l} \text{EXERCICE 1} \\ \text{(c) } \mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix} \text{ with } y^{(1)} = 1 \longrightarrow h(\mathbf{x}^{(1)}) = 0.42 \\ \text{with } h_1\left(a_1^{[2]}\right) = a_1^{[2]} \quad \text{and} \quad z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right) \end{array}$$

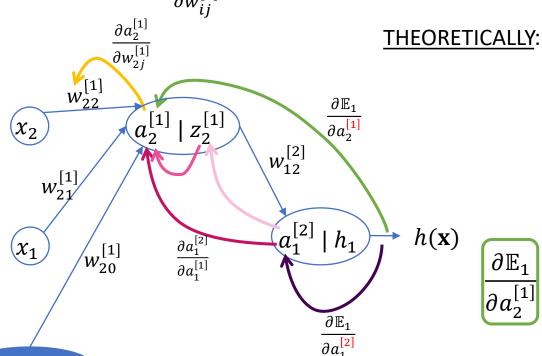
 $\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^m w_{km}^{[2]} z_m^{[1]} \right)$ $z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \quad a_m^{[1]} = \sum_{d=0}^{D} w_{md}^{[1]} x_d$

CHAIN RULE

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

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

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



$$\frac{\partial \mathbb{E}_{1}}{\partial w_{2j}^{[1]}} = \begin{bmatrix} \partial \mathbb{E}_{1} & \partial a_{1}^{[2]} \\ \partial a_{1}^{[2]} & \partial a_{2}^{[1]} \end{bmatrix} \begin{bmatrix} \partial a_{2}^{[1]} \\ \partial w_{2j}^{[1]} \end{bmatrix}$$

$$CHAIN R$$

$$\frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} = h_{1} \left(a_{1}^{[2]} \right) - y^{(1)} = \partial_{1}^{[2]}$$

$$a_{1}^{[2]} = \sum_{m=0}^{2} w_{1m}^{[2]} z_{m}^{[1]} = \sum_{m=0}^{2} w_{1m}^{[2]} \varphi \left(a_{m}^{[1]} \right)$$

$$\frac{\partial a_{1}^{[2]}}{\partial a_{2}^{[1]}} = \frac{\partial a_{1}^{[2]}}{\partial z_{2}^{[1]}} \frac{\partial z_{2}^{[1]}}{\partial a_{2}^{[1]}} = ?$$

 $x_0 = 1$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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

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

THEORETICALLY: $a_2^{[1]} \mid z_2^{[1]}$ $\partial a_2^{[1]}$ (x_2) $w_{12}^{[2]}$ $w_{21}^{[1]}$ $a_1^{[2]} \mid h_1$ $h(\mathbf{x})$ (x_1) $\frac{\partial a_1^{[2]}}{\partial a_1^{[1]}}$ $w_{20}^{[1]}$ $\partial \mathbb{E}_1$

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \begin{pmatrix} h_k = \varphi\left(a_k^{[2]}\right) & a_k^{[2]} = \sum_{m=0}^M w_{km}^{[2]} z_m^{[1]} \\ \sigma'(a) = \sigma(a)(1 - \sigma(a)) & a_k^{[1]} = \sum_{m=0}^D w_{km}^{[2]} z_m^{[1]} \end{pmatrix}$$

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

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

$$\frac{\partial \mathbb{E}_{1}}{\partial w_{2j}^{[1]}} = \begin{bmatrix} \partial \mathbb{E}_{1} & \partial a_{1}^{[2]} \\ \partial a_{1}^{[2]} & \partial a_{2}^{[1]} \end{bmatrix} \frac{\partial a_{2}^{[1]}}{\partial w_{2j}^{[1]}}$$

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

$$\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]}} = \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) = \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) = \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) = \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) = \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 z_{2}^{[1]}} \varphi \left(a_{2}^{[1]} \right) = \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 z_{2}^{[1]}} \varphi \left(a_{2}^{[1]} \right) = \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 z_{2}^{[1]}} \varphi \left(a_{2}^{[1]} \right) = \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 z_{2}^{[1]}} \varphi \left(a_{2}^{[1]} \right) = \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 z_{2}^{[1]}} \varphi \left(a_{2}^{[1]} \right) = \frac{\partial}{\partial z_{2}^{[1]}} \left\{ w_{10}^{[2]} z_{0}^{[2]} + w_{11}^{[2]} z_{1}^{[2]} + w_{12}^{[2]} z_{2}^{[2]} \right\} \frac{\partial}{\partial z_{2}^{[1]}} \varphi \left(a_{2}^{[1]} \right) = \frac{\partial}{\partial z_{2}^{[1]}} \left\{ w_{10}^{[2]} z_{0}^{[2]} + w_{11}^{[2]} z_{1}^{[2]} + w_{12}^{[2]} z_{2}^{[2]} \right\} \frac{\partial}{\partial z_{2}^{[2]}} \varphi \left(a_{2}^{[2]} z_{1}^{[2]} + w_{11}^{[2]} z_{2}^{[2]} \right\} \frac{\partial}{\partial z_{2}^{[2]}} \varphi \left(a_{2}^{[2]} z_{1}^{[2]} + w_{12}^{[2]} z_{2}^{[2]} \right\} \frac{\partial}{\partial z_{2}^{[2]}} \varphi \left(a_{2}^{[2]} z_{1}^{[2]} + w_{11}^{[2]} z_{2}^{[2]} \right)$$

 $x_0 = 1$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

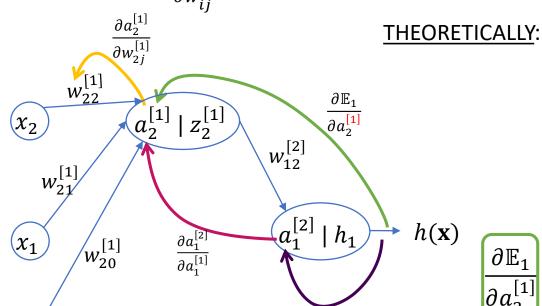
 $\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{m=0}^M w_{km}^{[2]} z_m^{[1]} \right)$ $(a) = \sigma(a)(1 - \sigma(a))$

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

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

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

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



$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \begin{bmatrix} \partial \mathbb{E}_1 & \partial a_1^{[2]} \\ \partial a_1^{[2]} & \partial a_2^{[1]} \end{bmatrix} \begin{bmatrix} \partial a_2^{[1]} \\ \partial w_2^{[2]} \end{bmatrix}$$

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

$$\frac{\partial a_1^{[2]}}{\partial a_2^{[1]}} = \frac{\partial a_1^{[2]}}{\partial z_2^{[1]}} \frac{\partial z_2^{[1]}}{\partial a_2^{[1]}} = w_{12}^{[2]} \sigma' \left(a_2^{[1]} \right)$$

 (x_1)

 $x_0 = 1$

 $w_{20}^{[1]}$

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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

with
$$h_1(a_1^{[2]}) = a_1^{[2]}$$

$$)=a_1^{\lfloor 2\rfloor} \qquad \mathsf{a}$$

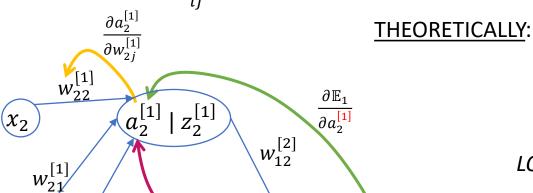
with
$$h_1\left(a_1^{[2]}\right)=a_1^{[2]}$$
 and $z_m=\varphi\left(a_m^{[1]}\right)=\sigma\left(a_m^{[1]}\right)$

 $h(\mathbf{x})$

(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.

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



 $a_1^{[2]} \mid h_1$

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \begin{bmatrix} \partial \mathbb{E}_1 & \partial a_1^{[2]} \\ \partial a_1^{[2]} & \partial a_2^{[1]} \\ \partial a_2^{[1]} & \partial w_{2j}^{[1]} \end{bmatrix} \begin{bmatrix} \partial a_2^{[1]} \\ \partial w_{2j}^{[1]} & \partial a_2^{[1]} \\ \partial a_2^{[1]} & \partial a_2^{[1]} \end{bmatrix}$$

LOCAL ERROR
$$\partial \mathbb{E}_{1}$$
 $\partial \mathbb{E}_{1}$ $\partial \mathbb{E}_{1}$ $\partial a_{1}^{[1]} = \partial a_{1}^{[2]} \partial a_{1}^{[2]} = \partial a_{1}^{[2]} w_{12}^{[2]} \sigma' \left(a_{2}^{[1]} \right)$

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

 (x_1)

 $x_0 = 1$

 $w_{20}^{[1]}$

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

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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 $\frac{\partial \mathbb{E}_1}{\partial w_{ii}^{[1]}}$.

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

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

$$\begin{array}{c}
\frac{\partial a_{2}^{[1]}}{\partial w_{2j}^{[1]}} & \underline{\text{THEORETICALLY}}: \\
x_{2} & a_{2}^{[1]} \mid z_{2}^{[1]} \\
w_{21}^{[1]} & w_{12}^{[2]} \\
\end{array}$$

$$\begin{array}{c}
\frac{\partial \mathbb{E}_{1}}{\partial a_{2}^{[1]}} \\
w_{12}^{[2]} & u_{12}^{[2]}
\end{array}$$

$$\begin{array}{c}
\mathcal{E}_{1} \\
w_{12}^{[2]} & u_{12}^{[2]}
\end{array}$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \begin{bmatrix} \partial \mathbb{E}_1 & \partial a_1^{[2]} \\ \partial a_1^{[2]} & \partial a_2^{[1]} \end{bmatrix} \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}$$

$$\frac{\text{LOCAL ERROR}}{\partial_{2}^{[1]}} \frac{\partial \mathbb{E}_{1}}{\partial a_{2}^{[1]}} = \underbrace{\frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}}}_{\frac{\partial a_{1}^{[2]}}{\partial a_{2}^{[1]}}} = \partial_{1}^{[2]} w_{12}^{[2]} \sigma' \left(a_{2}^{[1]} \right)$$

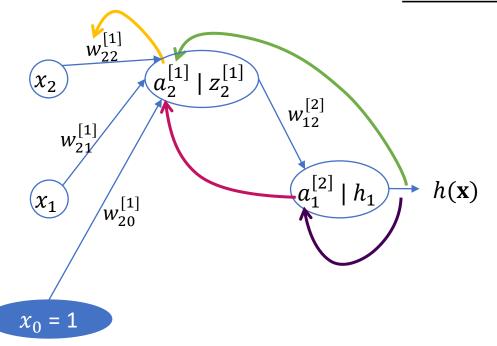
$$\frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}} = \frac{\partial}{\partial w_{2j}^{[1]}} \left\{ w_{10}^{[1]} x_0 + w_{11}^{[1]} x_1 + w_{12}^{[1]} x_2 \right\} = x_j$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

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

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

THEORETICALLY:



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

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

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

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

$$a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

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

$$\frac{\partial \mathbb{E}_{1}}{\partial a_{2}^{[1]}} = \partial_{2}^{[1]} \quad LOCAL \; ERROR$$

$$\frac{\partial \mathbb{E}_{1}}{\partial w_{2j}^{[1]}} = \begin{bmatrix} \partial \mathbb{E}_{1} & \partial a_{1}^{[2]} \\ \partial a_{1}^{[2]} & \partial a_{2}^{[1]} \end{bmatrix} \begin{bmatrix} \partial a_{2}^{[1]} \\ \partial w_{2j}^{[1]} \end{bmatrix} = \begin{bmatrix} \partial_{1}^{[2]} & w_{12}^{[2]} \sigma' & (a_{2}^{[1]}) \\ w_{12}^{[2]} & (a_{2}^{[1]}) \end{bmatrix} \begin{bmatrix} x_{j} \end{bmatrix}$$

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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

 $a_1^{[2]} \mid h_1$

LOCAL ERROR

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

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

THEORETICALLY:

$$\frac{\partial \mathbb{E}_{1}}{\partial w_{2j}^{[1]}} = \underbrace{\frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} \frac{\partial a_{1}^{[2]}}{\partial a_{2}^{[1]}}}_{\frac{\partial a_{2}^{[1]}}{\partial w_{2j}^{[1]}} = \underbrace{\frac{\partial \mathbb{E}_{1}}{\partial a_{2}^{[1]}}}_{\frac{\partial a_{2}^{[1]}}{\partial w_{2j}^{[1]}} = \underbrace{\frac{\partial \mathbb{E}_{1}}{\partial a_{2}^{[1]}}}_{\frac{\partial a_{2}^{[1]}}{\partial a_{2}^{[1]}} = \underbrace{\frac{\partial \mathbb{E}_{1}}{\partial a_{2}^{[1]}}}_{\frac{\partial a_{2}^{[1]}}{\partial a_{2}^{[1]}}}$$

 $w_{2^{-1}}^{[1]}$ (x_1) $w_{20}^{'[1]}$

$$h(\mathbf{x})$$
 $\partial_2^{[1]} =$

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

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

3.
$$a_1^{[2]} = 0.42$$

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

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

LOCAL ERROR

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

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 $\frac{\partial \mathbb{E}_1}{\partial w_{ii}^{[2]}}$.

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

THEORETICALLY:

$$\frac{\partial \mathbb{E}_1}{\partial w_{2j}^{[1]}} = \left(\frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}}\right) \frac{\partial a_2^{[1]}}{\partial w_{2j}^{[1]}} =$$

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

$$\begin{array}{c|c}
w_{22}^{[1]} \\
x_2 & a_2^{[1]} \mid z_2^{[1]} \\
w_{21}^{[1]} & w_{12}^{[2]} \\
\hline
x_1 & w_{20}^{[1]} & a_1^{[2]} \mid h_1 \longrightarrow h(\mathbf{x})
\end{array}$$

$$\frac{\partial^{[1]}}{\partial a_2^{[1]}} = \frac{\partial \mathbb{E}_1}{\partial a_2^{[1]}} = \frac{\partial \mathbb{E}_1}{\partial a_1^{[2]}} \frac{\partial a_1^{[2]}}{\partial a_2^{[1]}} = \partial_1^{[2]} w_{12}^{[2]} \sigma' \left(a_2^{[1]} \right) \\
= \left(h_1 \left(a_1^{[2]} \right) - y^{(1)} \right) w_{12}^{[2]} \sigma \left(a_2^{[1]} \right) \left(1 - \sigma \left(a_2^{[2]} \right)^2 \right) \\
= (0.42 - 1)0.2 \sigma (0.18) (1 - \sigma (0.18)) = -0.03$$

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

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

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

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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

with
$$h_1(a_1^{[2]}) = a_1^{[2]}$$

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 $\frac{\partial \mathbb{E}_1}{\partial w_{ii}^{[2]}}$.

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

 $\partial_2^{[1]} = -0.03$

THEORETICALLY:

$$\frac{\partial \mathbb{E}_{1}}{\partial w_{2j}^{[1]}} = \frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} \frac{\partial a_{1}^{[2]}}{\partial a_{2}^{[1]}} \frac{\partial a_{2}^{[1]}}{\partial w_{2j}^{[1]}} = \frac{\partial^{[2]}}{\partial a_{1}^{[2]}} w_{12}^{[2]} \sigma' \left(a_{2}^{[1]}\right) x_{j}$$

$$w_{2j}^{[1]} = ?$$

$$w_{21}^{[1]} = ?$$

$$w_{21}^{[1]} = ?$$

$$w_{20}^{[1]} = ?$$

$$a_{1}^{[2]} \mid h_{1} \rightarrow h(\mathbf{x})$$

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

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

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

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

3.
$$a_1^{[2]} = 0.42$$

$$\frac{\partial \mathbb{E}_1}{\partial w_{22}^{[1]}} = 3$$

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

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{k=0}^{M} w_{km}^{[2]} z_m^{[1]} \right)$$

(c)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.42$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = a_1^{[2]}$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

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

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

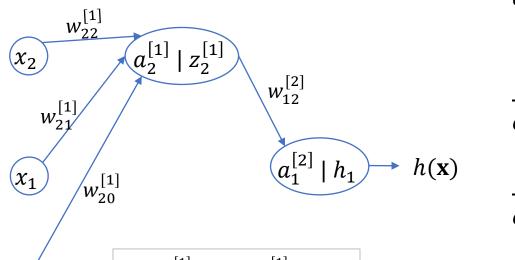
(iii) We want to calculate the $\frac{\partial \mathbb{E}_1}{\partial w_{ii}^{[2]}}$.

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

$$\partial_2^{[1]} = -0.03$$

THEORETICALLY:

$$\frac{\partial \mathbb{E}_{1}}{\partial w_{2j}^{[1]}} = \frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} \frac{\partial a_{1}^{[2]}}{\partial a_{2}^{[1]}} \frac{\partial a_{2}^{[1]}}{\partial w_{2j}^{[1]}} = \partial_{1}^{[2]} w_{12}^{[2]} \sigma' \left(a_{2}^{[1]} \right) x_{j}$$



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

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

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

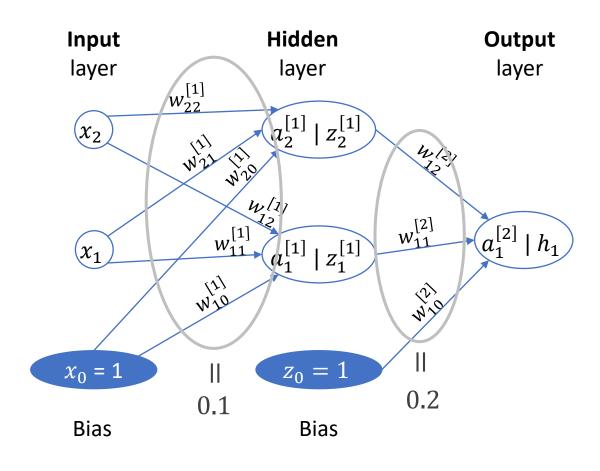
1.
$$a_1^{[1]} = 0.18, a_2^{[1]} = 0.18$$

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

3.
$$a_1^{[2]} = 0.42$$

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

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



NEW INFORMATION

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

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

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

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

(d)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = ?$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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

 $w_{22}^{[1]}$

W10

0.1

 (x_2)

 (x_1)

 $x_0 = 1$

Bias

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

 $w_{11}^{[2]}$

0.2

 $a_1^{[2]} \mid h_1$

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

 $a_2^{[1]} \mid z_2^{[1]}$

 $a_1^{[1]} \mid z_1^{[1]}$

 $z_0 = 1$

Bias

W/11

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

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

$$z_1^{[1]} = ?$$

$$z_2^{[1]} = ?$$

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

$$h_1 = ?$$

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

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

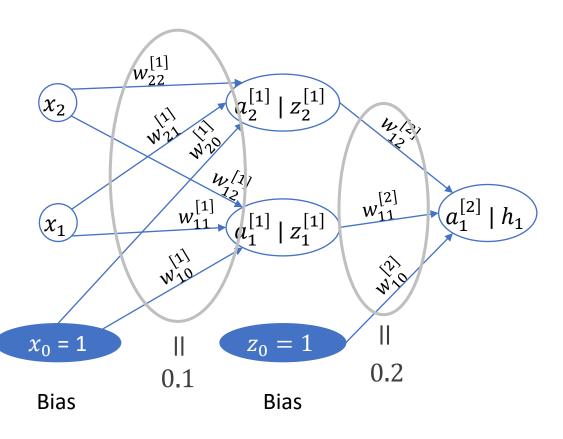
- 1. Compute $a_1^{[1]}$ and $a_2^{[1]}$.
- 2. Compute $z_1^{[1]}$ and $z_2^{[1]}$.
- 3. Compute $a_1^{[2]}$.
- Compute h_1

(d)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = ?$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$ $\mathbb{E}_n = -y^{(n)} \ln h(\mathbf{x}^{(n)}) - (1 - y^{(n)}) \ln \left(1 - h(\mathbf{x}^{(n)})\right)$

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

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

$$z_m^{[1]} = \varphi\left(a_m^{[1]}\right) \quad a_m^{[1]} = \sum^D w_{md}^{[1]} x_d$$



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

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

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

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

$$h_1 = \sigma(0.42) = 0.60$$
 $\Longrightarrow h(\mathbf{x}^{(1)}) = 0.60$

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
$$h_1$$

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

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

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

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

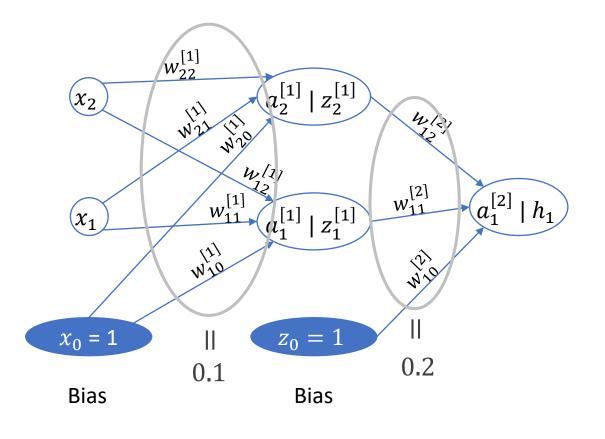
with
$$h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$$
 and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

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

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

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

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



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

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \left(h_k = \varphi\left(a_k^{[2]}\right) \qquad a_k^{[2]} = \sum_{k=0}^{M} w_{km}^{[2]} z_m^{[1]} \right)$$

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

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

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

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

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

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

$$z_{m}^{[1]} = \sigma(a)(1 - \sigma(a))$$

$$z_{m}^{[1]} = \varphi\left(a_{m}^{[1]}\right) \quad a_{m}^{[1]} = \sum_{m=0}^{D} w_{md}^{[1]} x_{d}$$

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

$$x_{2}$$
 x_{2}
 x_{2}
 x_{3}
 x_{4}
 x_{2}
 x_{2}
 x_{1}
 x_{2}
 x_{3}
 x_{4}
 x_{1}
 x_{1}
 x_{1}
 x_{2}
 x_{1}
 x_{1}
 x_{2}
 x_{1}
 x_{1}
 x_{2}
 x_{3}
 x_{4}
 x_{1}
 x_{1}
 x_{1}
 x_{2}
 x_{3}
 x_{4}
 x_{1}
 x_{1}
 x_{1}
 x_{2}
 x_{3}
 x_{4}
 x_{1}
 x_{1}
 x_{2}
 x_{3}
 x_{1}
 x_{2}
 x_{3}
 x_{4}
 x_{1}
 x_{1}
 x_{2}
 x_{3}
 x_{1}
 x_{2}
 x_{3}
 x_{4}
 x_{1}
 x_{2}
 x_{1}
 x_{2}
 x_{1}
 x_{2}
 x_{3}
 x_{4}
 x_{1}
 x_{2}
 x_{3}
 x_{4}
 x_{1}
 x_{2}
 x_{3}
 x_{4}
 x_{4}
 x_{1}
 x_{4}
 x_{4}
 x_{1}
 x_{4}
 x_{4

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

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

(d)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\rightarrow h(\mathbf{x}^{(1)}) = 0.60$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$

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

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

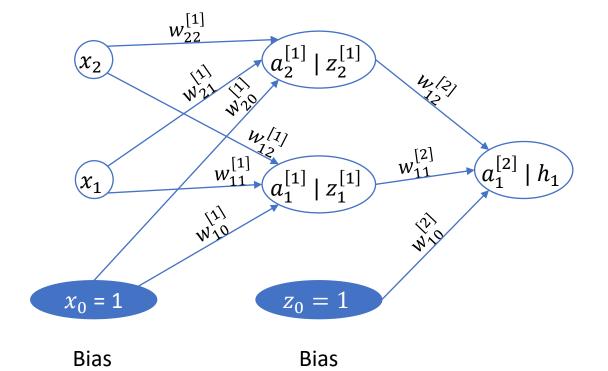
and
$$z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$$

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

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

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

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



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

(d)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.60$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$

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

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \begin{cases} h_k = \varphi\left(a_k^{[2]}\right) & a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]} \\ (a) = \sigma(a)(1 - \sigma(a)) \end{cases}$$

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

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

$$a_{2}^{[1]} \mid z_{2}^{[1]}$$

$$a_{1}^{[1]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid h_{1}$$

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

$$\begin{cases} \partial \mathbb{E}_1 \\ \partial a_1^{[2]} \end{cases} \stackrel{\checkmark}{=} ?$$

$$\text{LOCAL ERROR } \frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} = ?$$

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

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

$$= ?$$

$$\frac{\partial a_1^{[2]}}{\partial w_{1j}^{[2]}} = 3$$

(d)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.60$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$ $\mathbb{E}_n = -y^{(n)} \ln h(\mathbf{x}^{(n)}) - (1 - y^{(n)}) \ln \left(1 - h(\mathbf{x}^{(n)})\right)$

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

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

$$h_k = \varphi\left(a_k^{[2]}\right)$$

$$a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

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

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

THEORETICALLY:
$$\frac{\partial \mathbb{E}_{1}}{\partial w_{1i}^{[2]}} = \frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} \frac{\partial a_{1}^{[2]}}{\partial w_{1i}^{[2]}} \quad CHAIN \, RULE$$

$$a_{2}^{[1]} \mid z_{2}^{[1]}$$

$$a_{1}^{[1]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid h_{1}$$

LOCAL ERROR
$$\frac{\partial \mathbb{E}_{1}}{\partial a_{1}^{[2]}} = \frac{y^{(n)}}{h_{1}(a_{1}^{[2]})} h'_{1}(a_{1}^{[2]}) + \frac{1 - y^{(n)}}{1 - h_{1}(a_{1}^{[2]})} h'_{1}(a_{1}^{[2]})$$

$$= -y^{(n)}(1 - h_{1}(a_{1}^{[2]})) + (1 - y^{(n)})h_{1}(a_{1}^{[2]})$$

$$= -y^{(n)} + h_{1}(a_{1}^{[2]})$$

$$\frac{\partial a_{1}^{[2]}}{\partial w_{1j}^{[2]}} = z_{j}^{[1]}$$

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

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

$$\sigma(a)$$

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

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

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

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

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

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

$$a_{2}^{[1]} \mid z_{2}^{[1]}$$

$$a_{1}^{[1]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid z_{1}^{[1]}$$

$$a_{1}^{[2]} \mid h_{1}$$

$$a_{1}^{[2]} \mid h_{1}$$

$$h(\mathbf{x})$$

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

$$a_1^{[2]} \mid h_1 \longrightarrow h(\mathbf{x}) \qquad \frac{\partial \mathbb{E}_1}{\partial w_{12}^{[2]}} = -0.4 \cdot 1 = -0.4$$

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

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

 $z_0 = 1$

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

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

3.
$$a_1^{[2]} = 0.42$$

(d)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ $\longrightarrow h(\mathbf{x}^{(1)}) = 0.60$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1\left(a_1^{[2]}\right) = \sigma\left(a_1^{[2]}\right)$ and $z_m = \varphi\left(a_m^{[1]}\right) = \sigma\left(a_m^{[1]}\right)$ $\mathbb{E}_n = -y^{(n)} \ln h(\mathbf{x}^{(n)}) - (1 - y^{(n)}) \ln \left(1 - h(\mathbf{x}^{(n)})\right)$

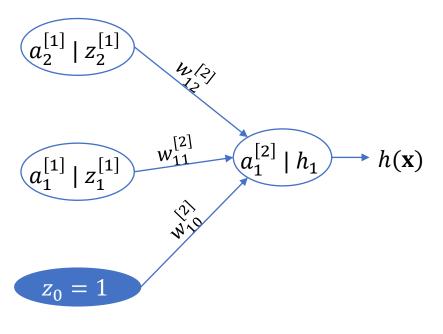
$$\sigma(a) = \frac{1}{1 + \exp(-a)}$$

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

$$a_k^{[2]} = \sum_{m=0}^{M} w_{km}^{[2]} z_m^{[1]}$$

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

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



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

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

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

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

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

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

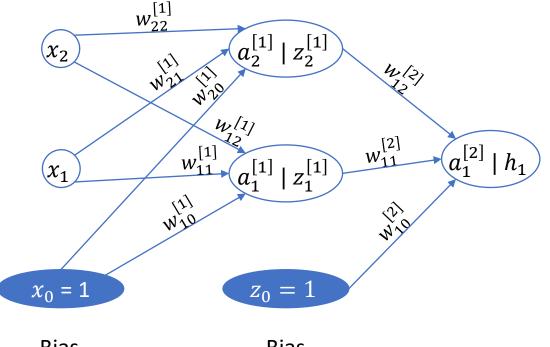
$$\frac{1}{+\exp(-a)}$$

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

(d)
$$\mathbf{x}^{(1)} = \begin{pmatrix} 0.3 \\ 0.5 \end{pmatrix}$$
 with $y^{(1)} = 1$ \longrightarrow $h(\mathbf{x}^{(1)}) = 0.60$ $\sigma'(a) = \sigma(a)(1 - \sigma(a))$ with $h_1(a_1^{[2]}) = \sigma(a_1^{[2]})$ and $z_m = \varphi(a_m^{[1]}) = \sigma(a_m^{[1]})$

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

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



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

SAME AS BEFORE

Bias