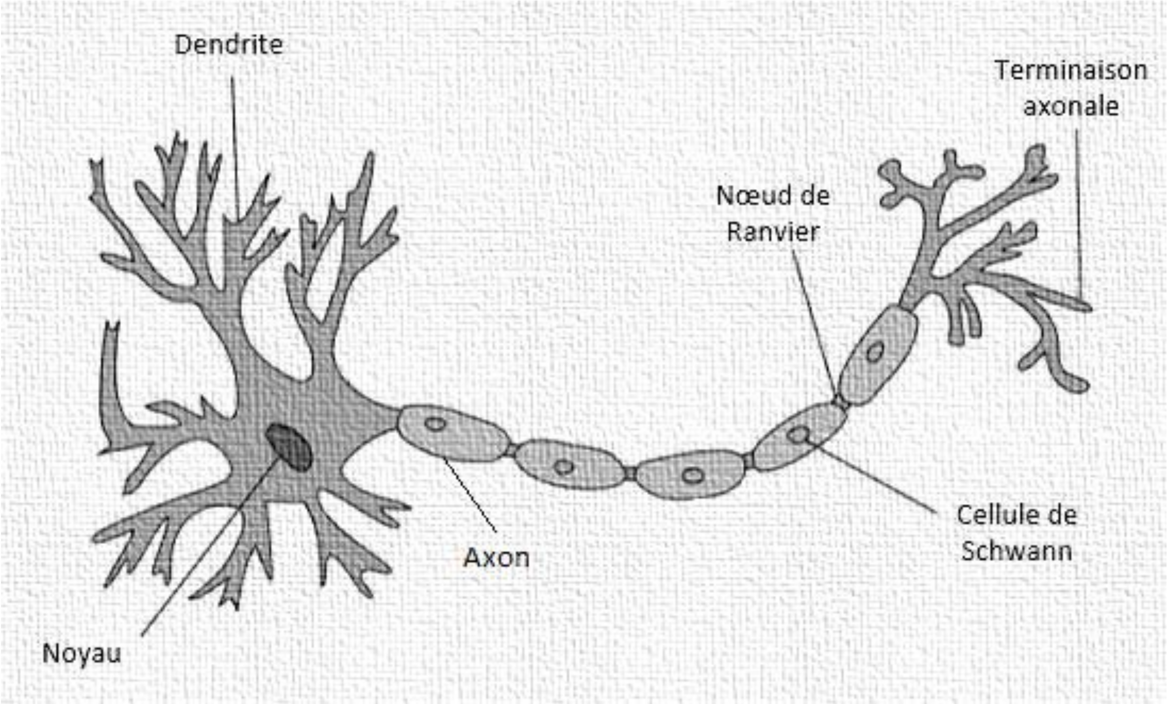
Réseaux de neurones artificiels

Aziz KHAMJANE

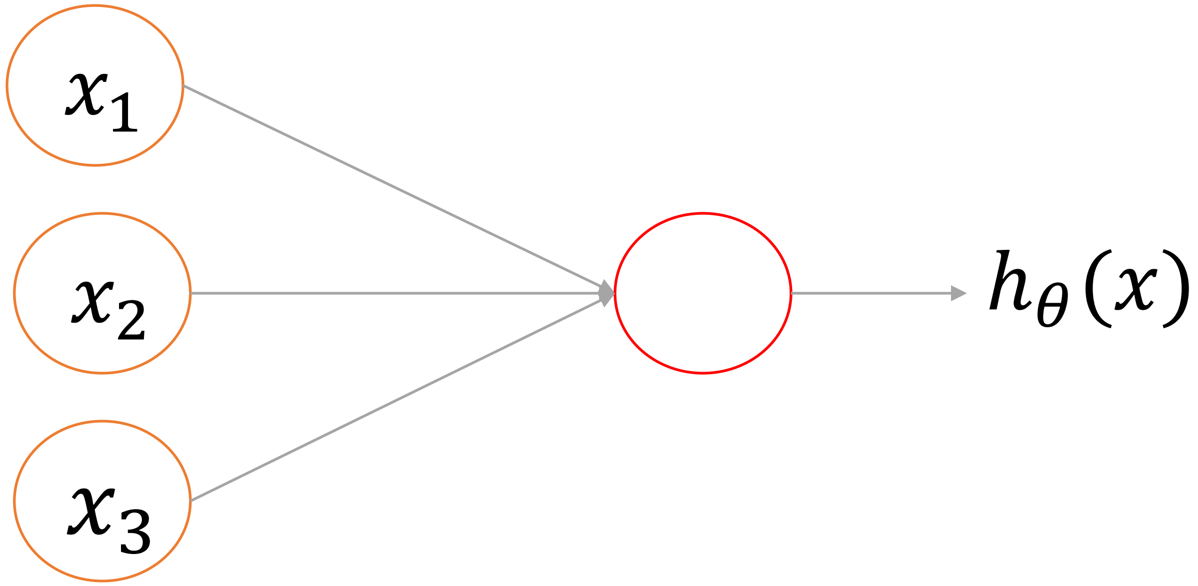
Réseau de neurones artificiels

- Des algorithmes essayant d'imiter le cerveau humain

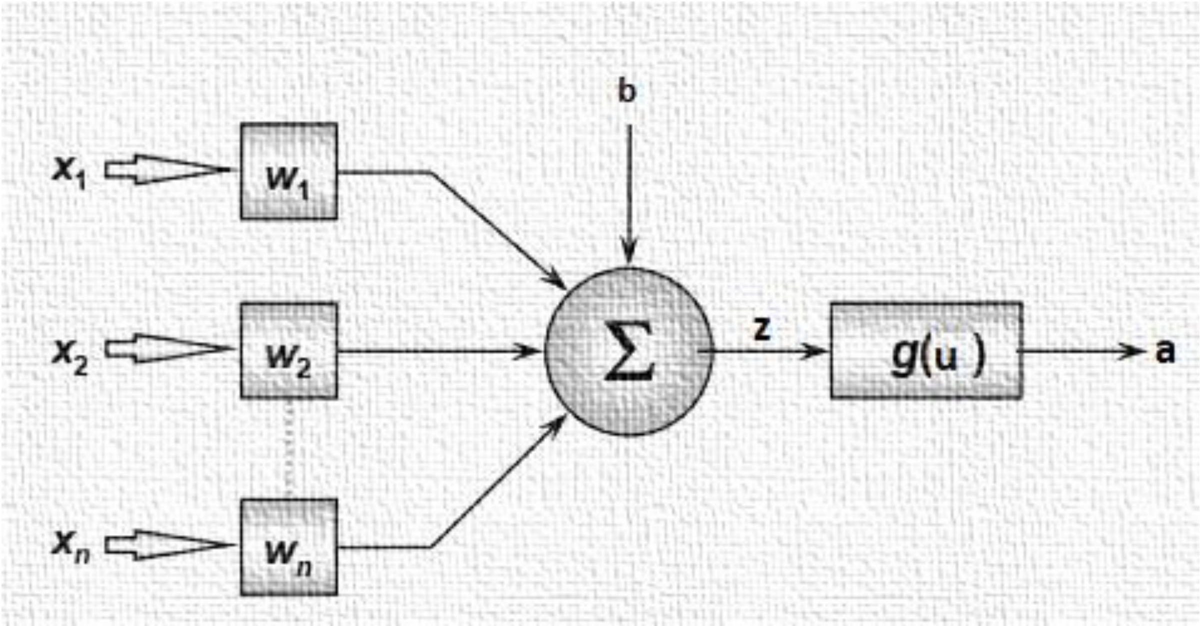
Neurone biologique



Neurone formel



Neurone formel

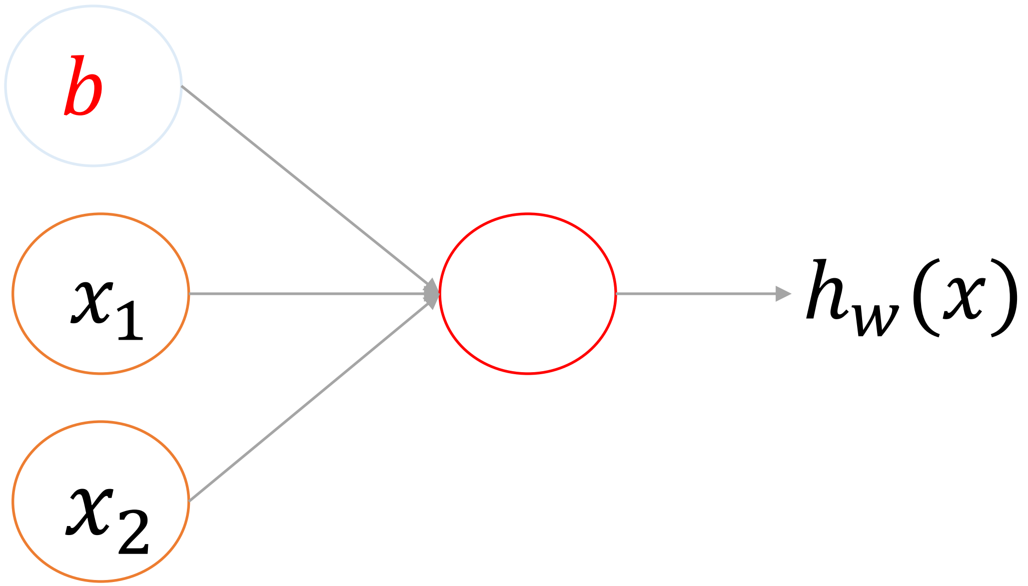


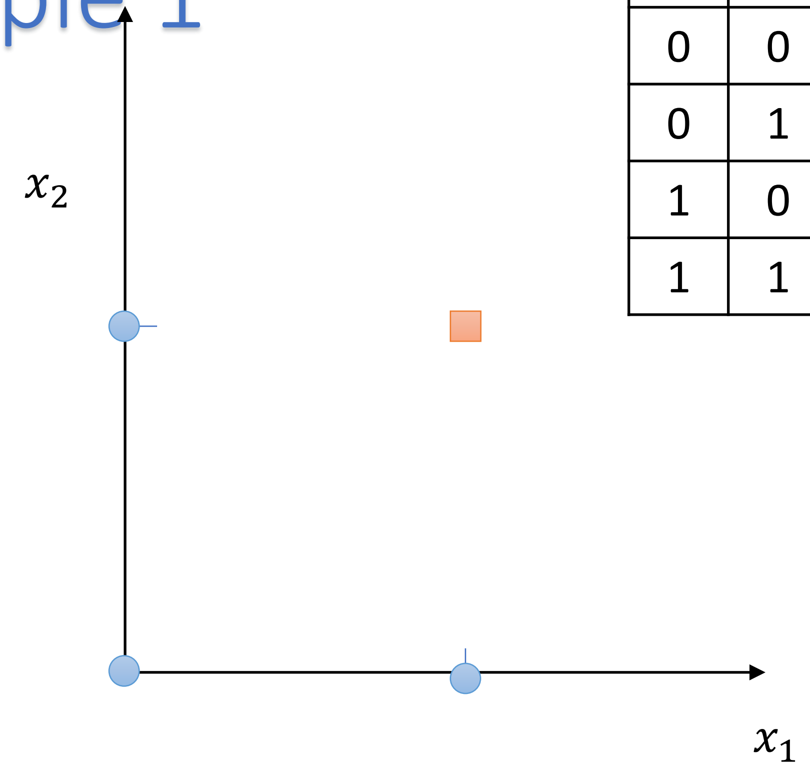
$$y = g\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

Perceptron: exemple 1

$$y = x1\ et\ x2$$

x1	x2	Y
0	0	0
0	1	0
1	0	0
1	1	1





$w_1 = 10 \quad w_2 = 10 \qquad b = -15$

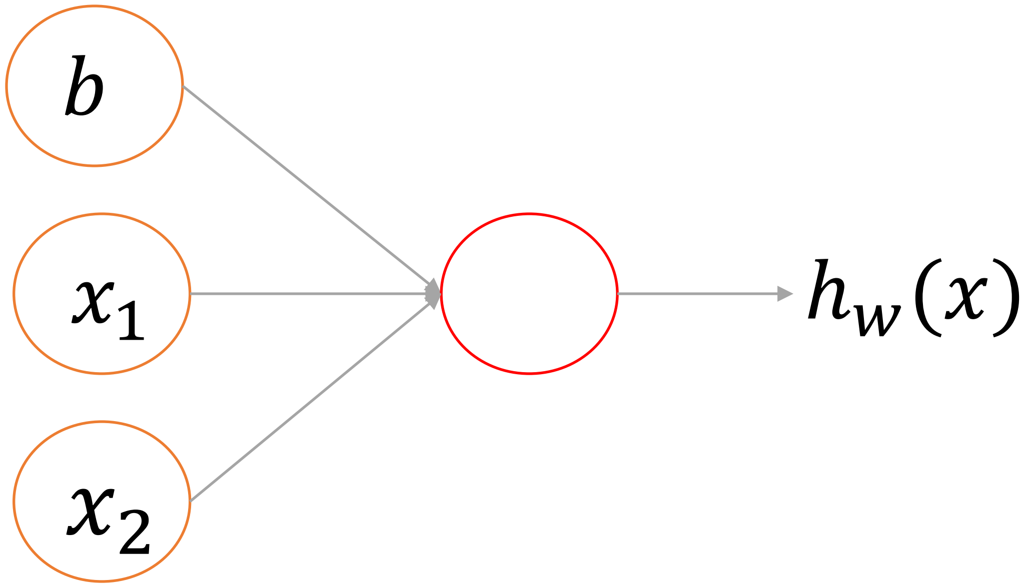
$$y = g\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

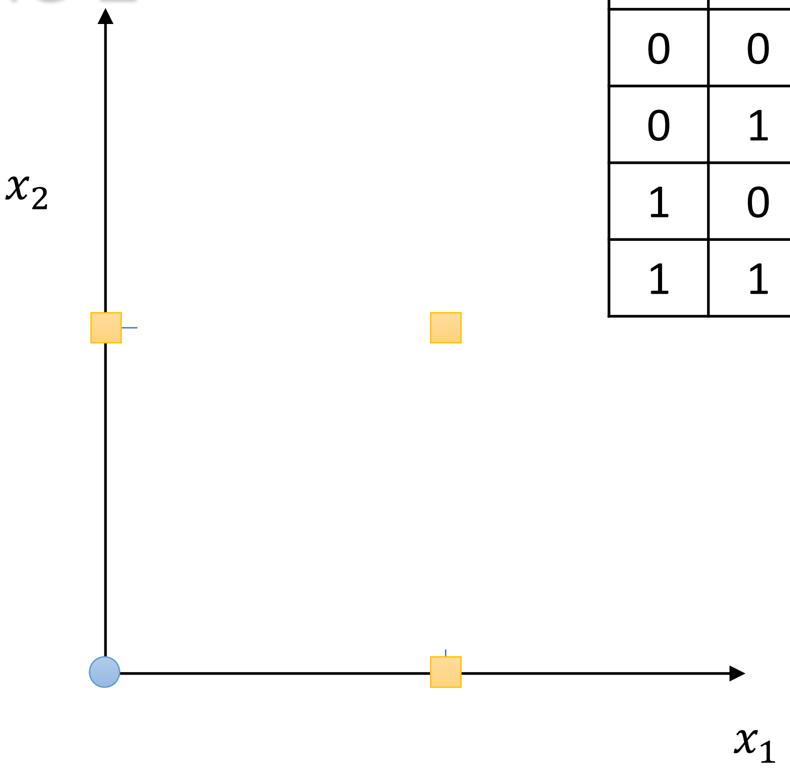
$$g(z) = \frac{1}{1 + e^{-z}}$$

x1	X2	z	$h_w(x)$	$\hat{y}$
0	0	$z = 0 \times 10 + 0 \times 10 - 15 = -15$	0,0000003	0
0	1	$z = 0 \times 10 + 1 \times 10 - 15 = -5$	0,00669	0
1	0	-$z = 1 \times 10 + 0 \times 10 - 15 = -5$	0,00669	0
1	1	$z = 1 \times 10 + 1 \times 10 - 15 = 5$	0,99330	1

Perceptron : exemple 2

$$y = x1 \ ou \ x2$$





x1	x2	Y
0	0	0
0	1	1
1	0	1
1	1	1

$w_1 = 15 \quad w_2 = 15 \qquad b = -10$

$$y = g\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

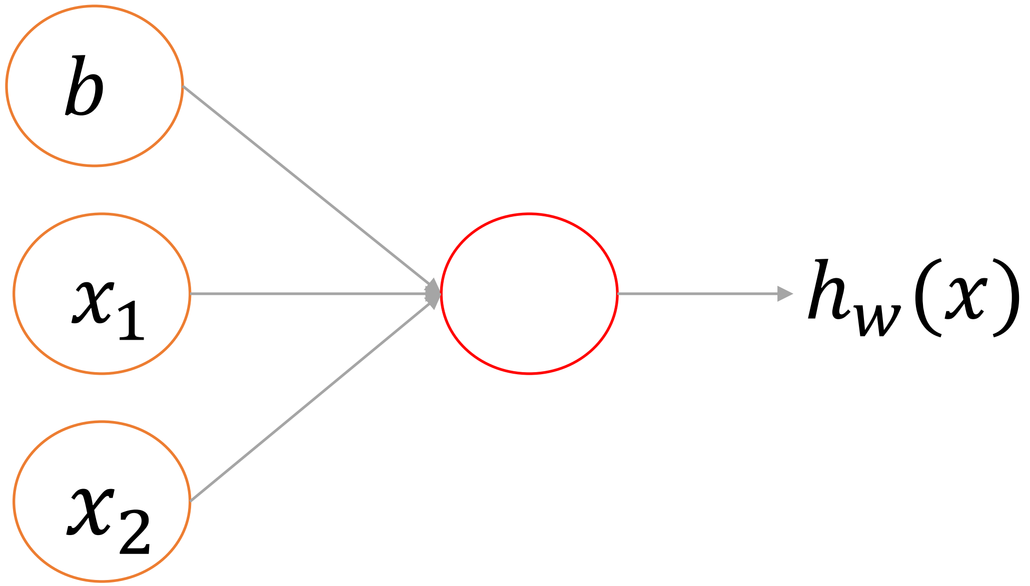
$$g(z) = \frac{1}{1 + e^{-z}}$$

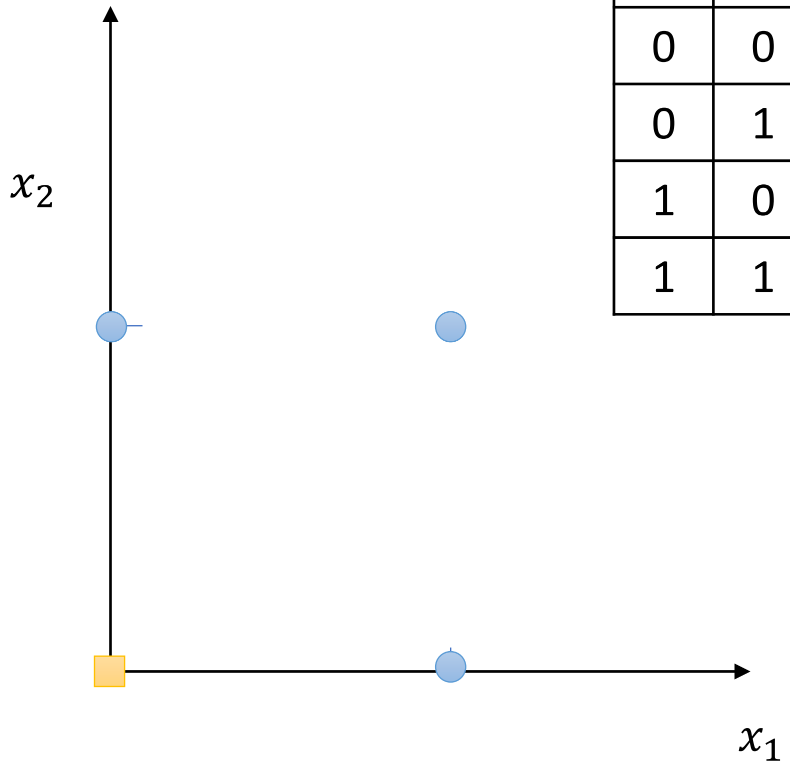
x1	X2	z	$h_w(x)$	$\hat{y}$
0	0	$z = 0 \times 15 + 0 \times 15 - 10 = -10$	0,000045	0
0	1	$z = 0 \times 15 + 1 \times 15 - 10 = 5$	0,9933	1
1	0	$z = 1 \times 15 + 0 \times 15 - 10 = 5$	0,9933	1
1	1	$z = 1 \times 15 + 1 \times 15 - 10 = 20$	0,99999	1

Perceptron : exemple 3

$$y = non\ (x1) et\ non(x2)$$

x1	x2	Y
0	0	1
0	1	0
1	0	0
1	1	0





$w_1 = -15 \quad w_2 = -15 \quad b = 10$

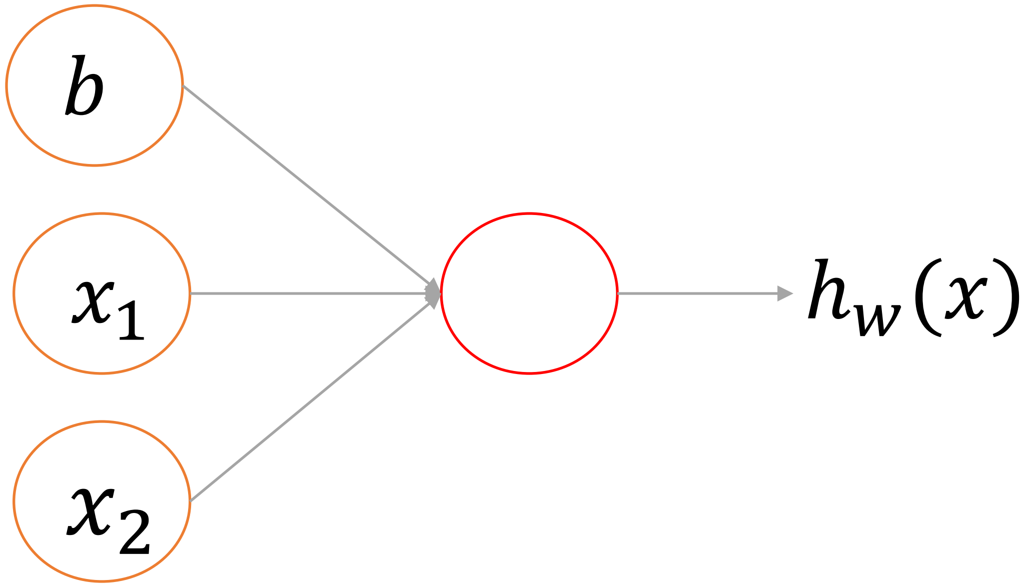
$$y = g\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

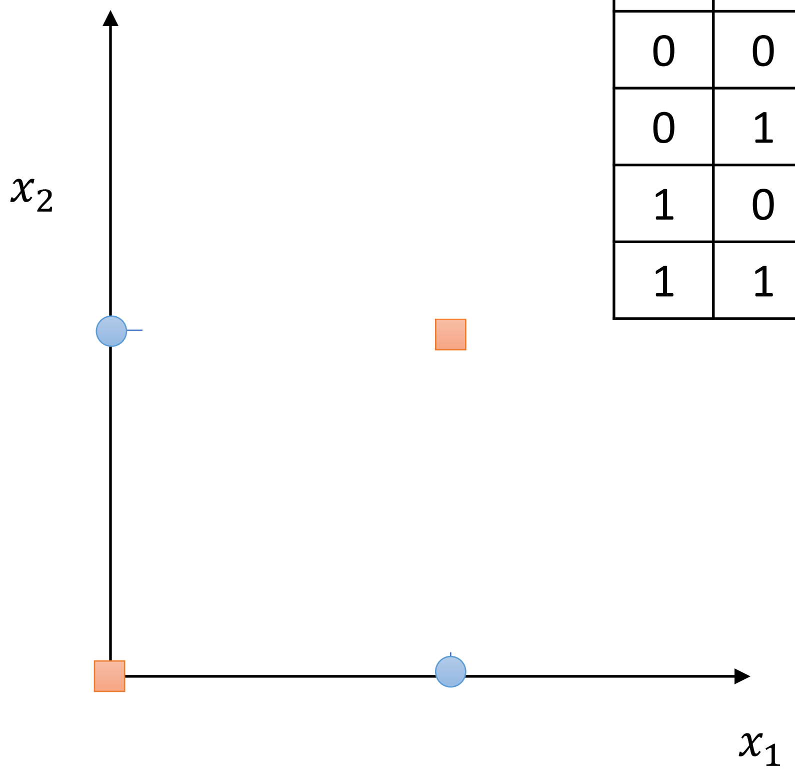
$$g(z) = \frac{1}{1 + e^{-z}}$$

x1	X2	z	$h_w(x)$	$\hat{y}$
0	0	$z = 0 \times (-15) + 0 \times (-15) + 10 = 10$	0,999954	1
0	1	$z = 0 \times (-15) + 1 \times (-15) + 10 = -5$	0,006692	0
1	0	$z = 1 \times (-15) + 0 \times (-15) + 10 = -5$	0,06692	0
1	1	$z = 1 \times (-15) + 1 \times (-15) + 10 = -20$	0,00000002	0

Problème non linéairement séparable

$$y = x1\ XNOR\ x2$$



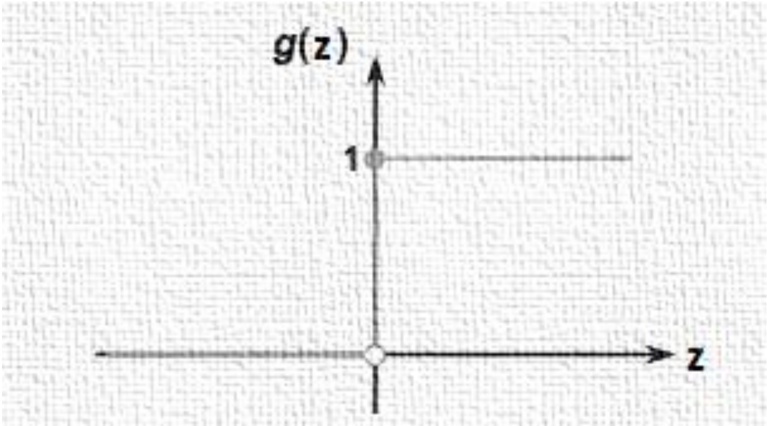


x1	x2	Y
0	0	1
0	1	0
1	0	0
1	1	1

Les fonctions d’activation

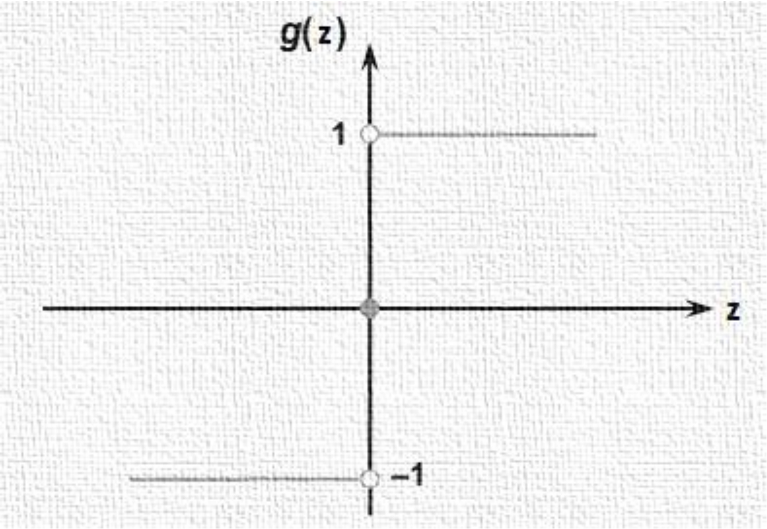
Fonction de Heaviside :

$$g(z) = \begin{cases} 1 \; si \; z \geq \; 0 \\ 0 \; si \; z < 0 \end{cases}$$



Fonction signe:

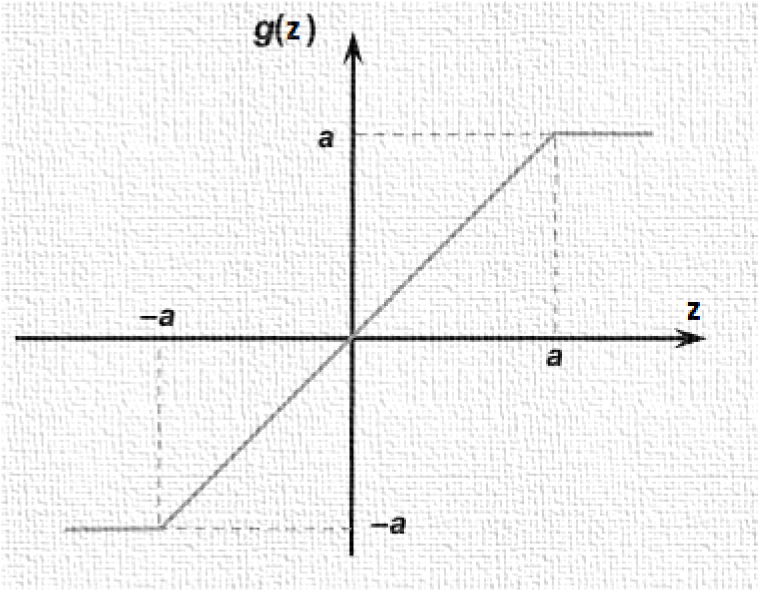
$$g(z) = \begin{cases} 1 \; si \; z \geq \; 0 \\ -1 \; si \; z < 0 \end{cases}$$



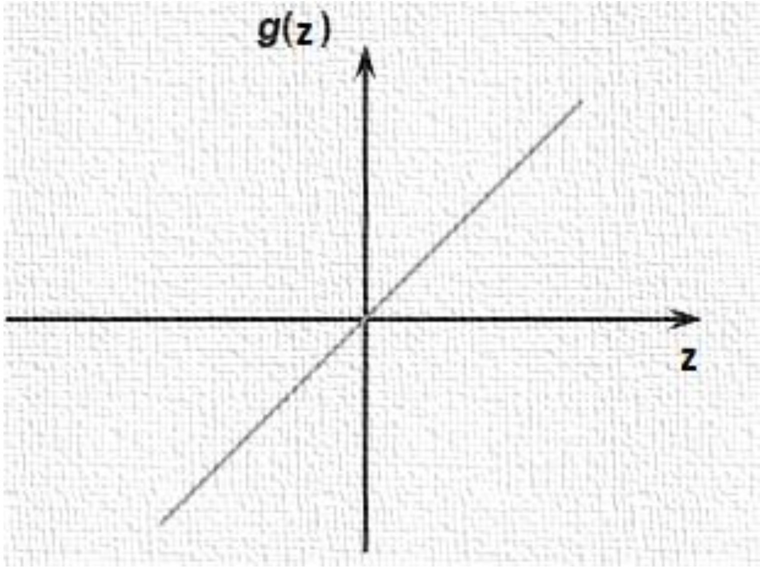
Les fonctions d'activation

Fonction linéaire par morceaux
:

$$g(z) = \begin{cases} a \ si \ z > a \\ z \ si \ -a \leq z \leq a \\ -a \ si \ z < -a \end{cases}$$



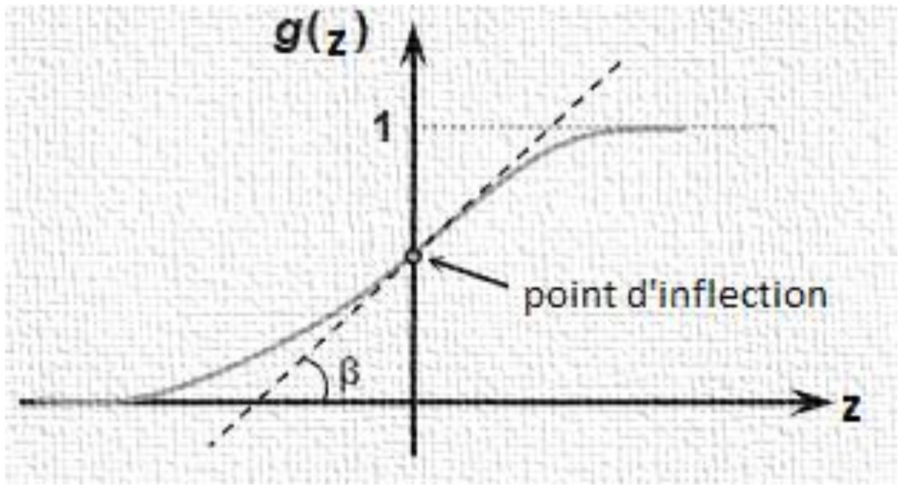
Fonction linéaire:
$g(z) = z$

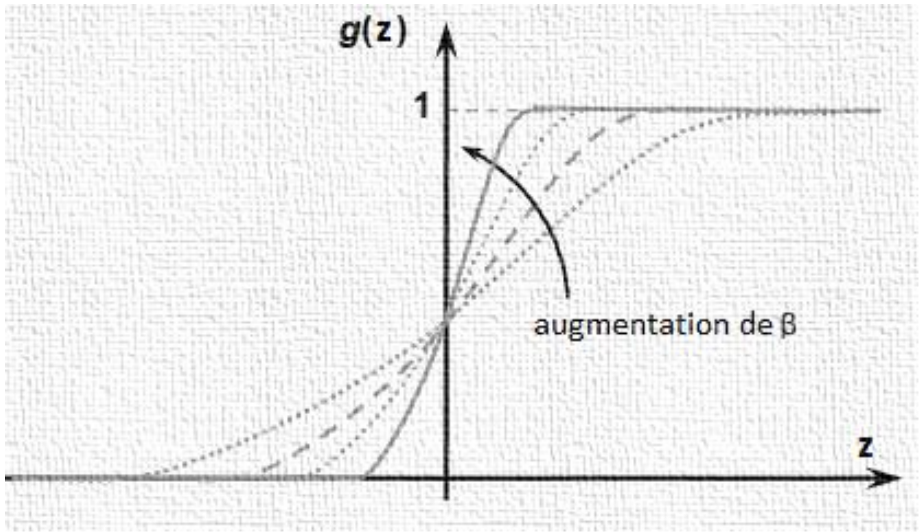


Les fonctions d'activation

Fonction sigmoïde

$$g(z) = \frac{1}{1 + e^{-\beta z}}$$

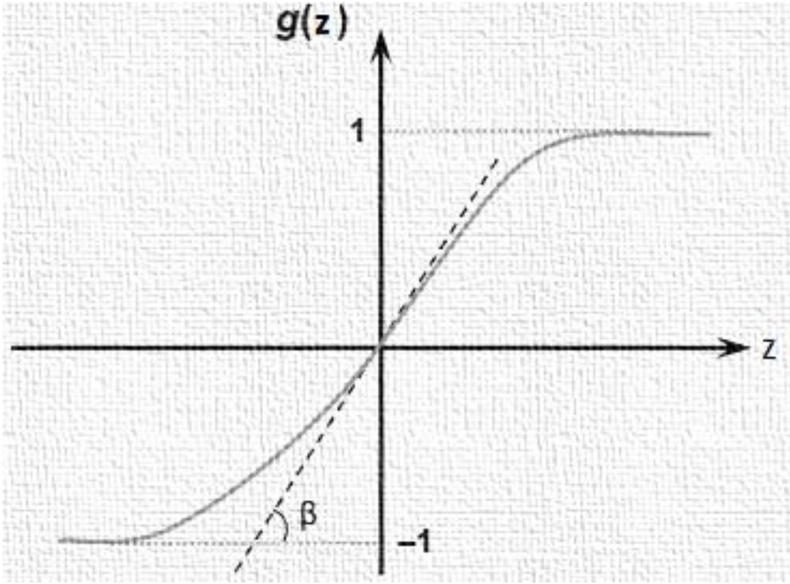


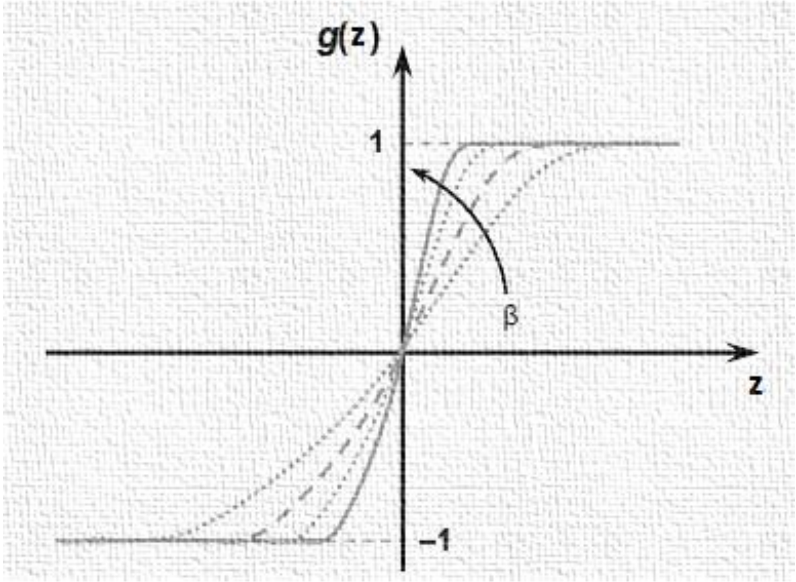


Les fonctions d'activation

Tangente hyperbolique

$$g(z) = \frac{1 - e^{-\beta z}}{1 + e^{-\beta z}}$$

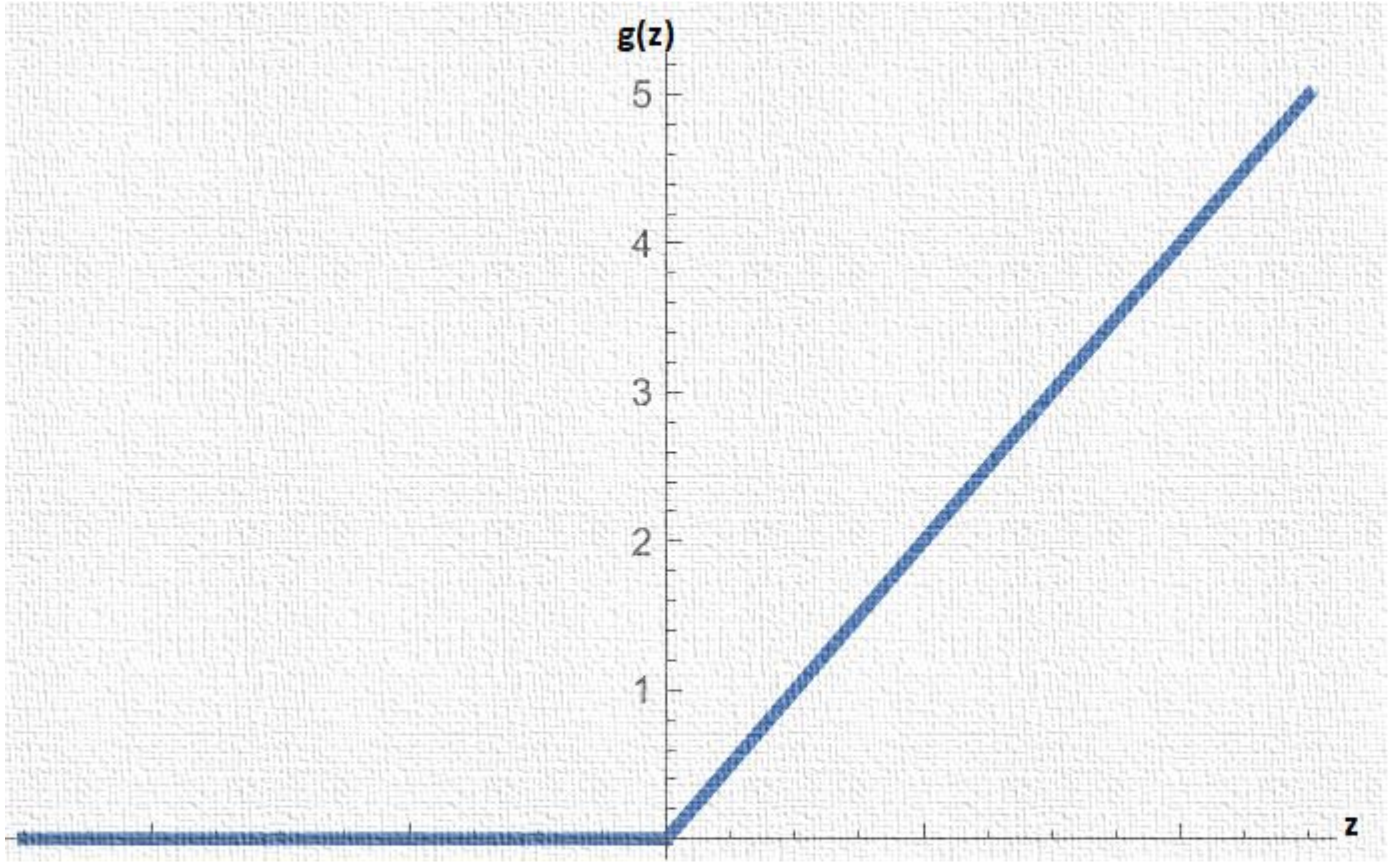




Les fonctions d'activation

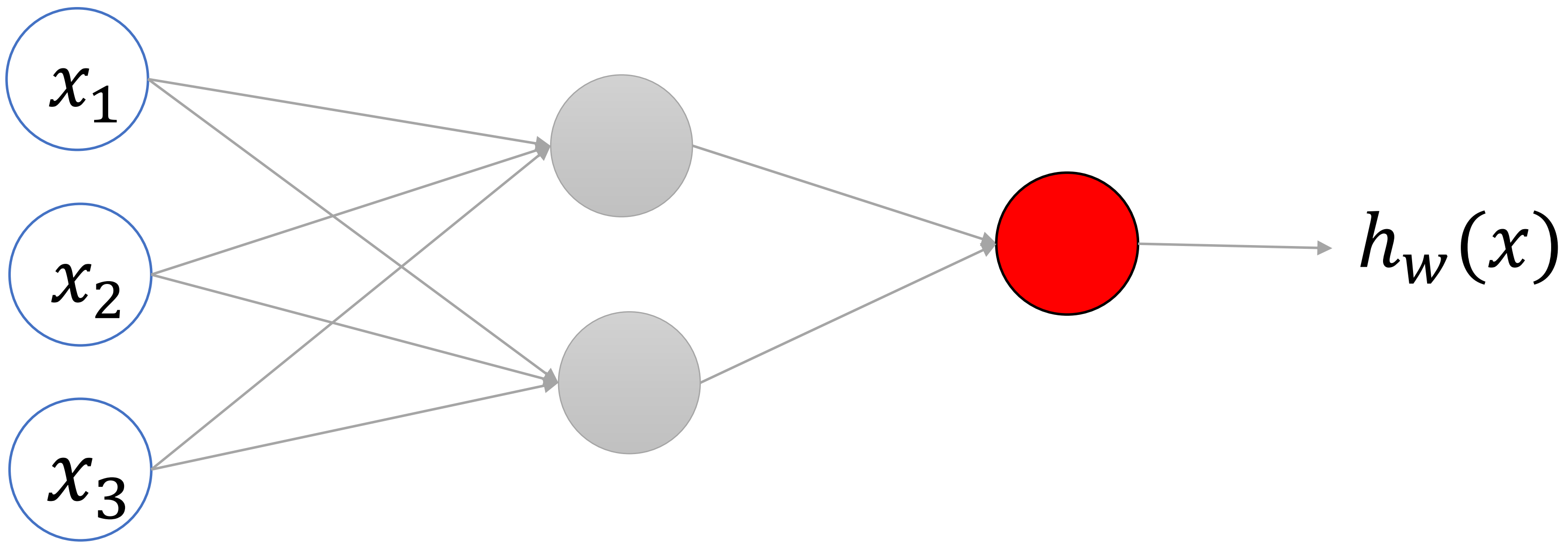
Relu

$$g(z) = \max(0, z)$$

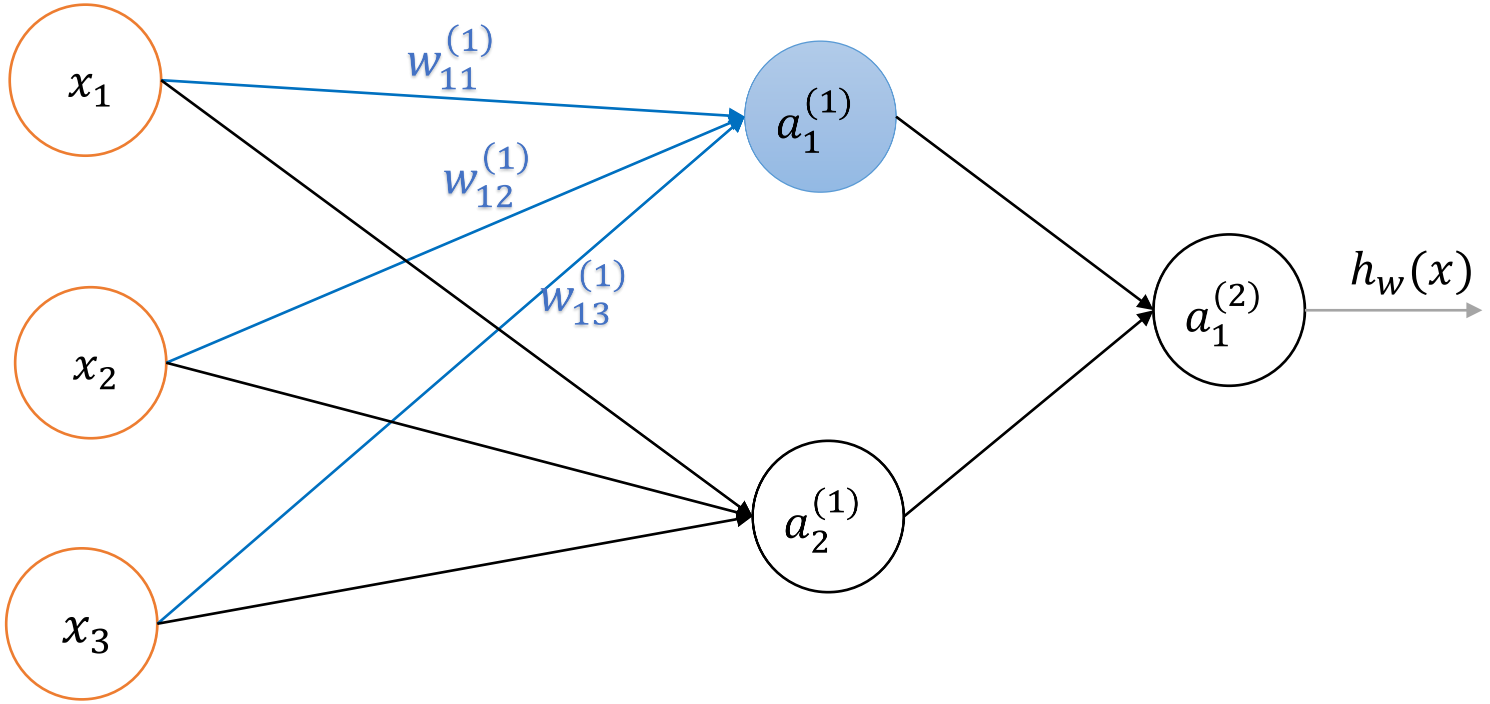


Architecture d’un réseau de neurones

- La couche d’entrée
- Les couches cachées
- La couche de sortie

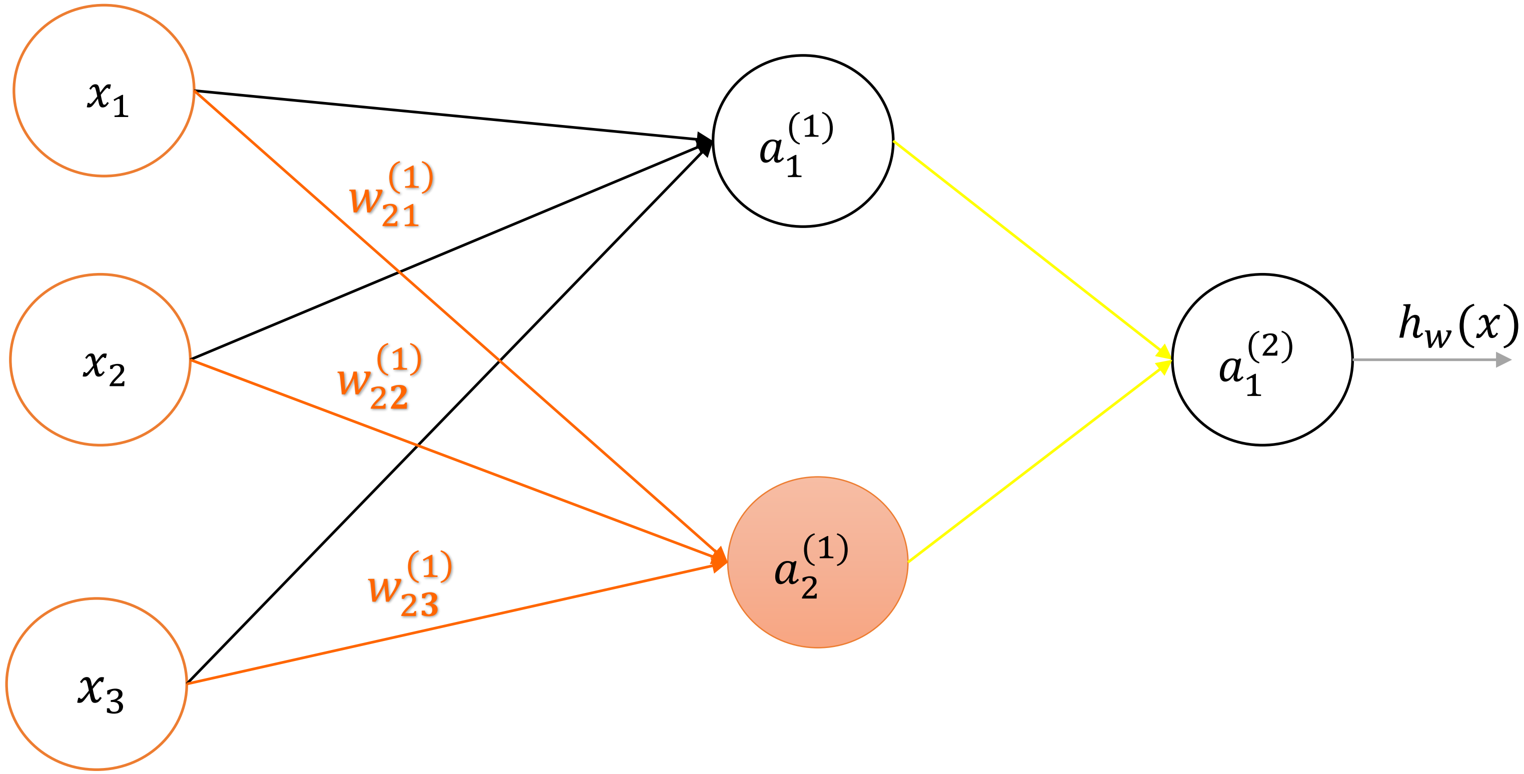


Réseau de neurones: propagation



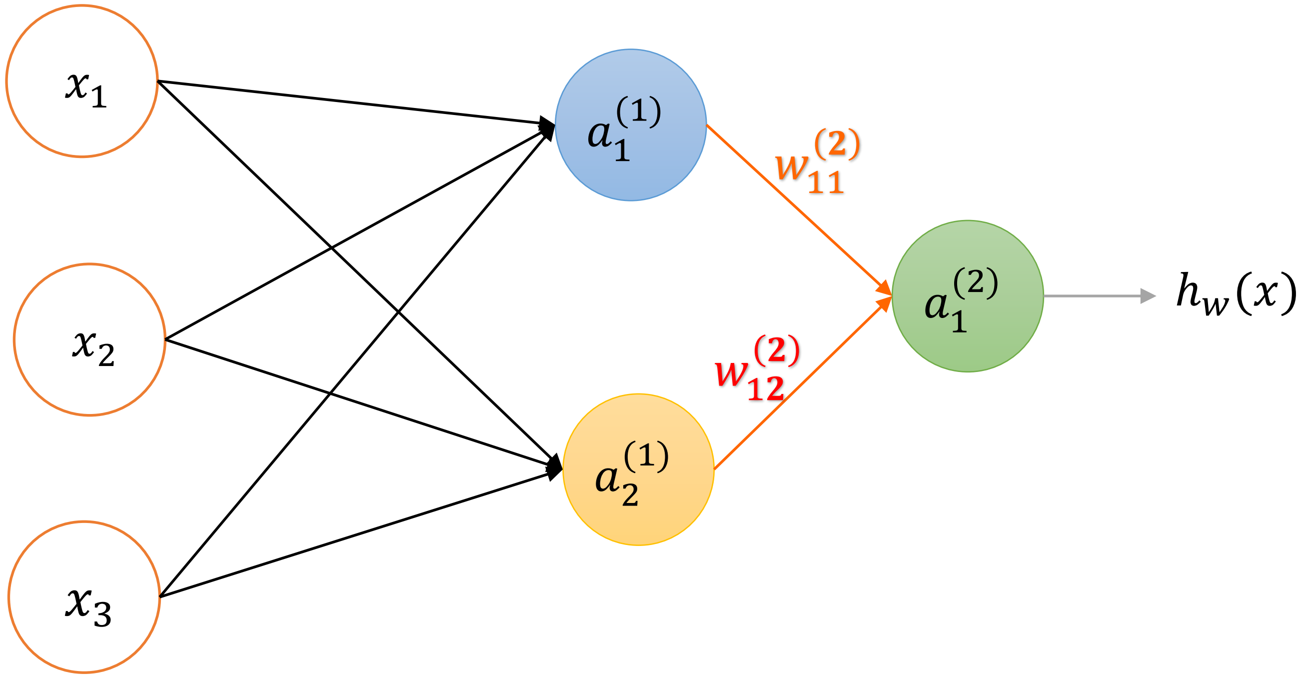
$$a_1^{(1)} = g\left(w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2 + w_{13}^{(1)}x_3 + b_1^{(1)}\right)$$

Réseau de neurones: propagation



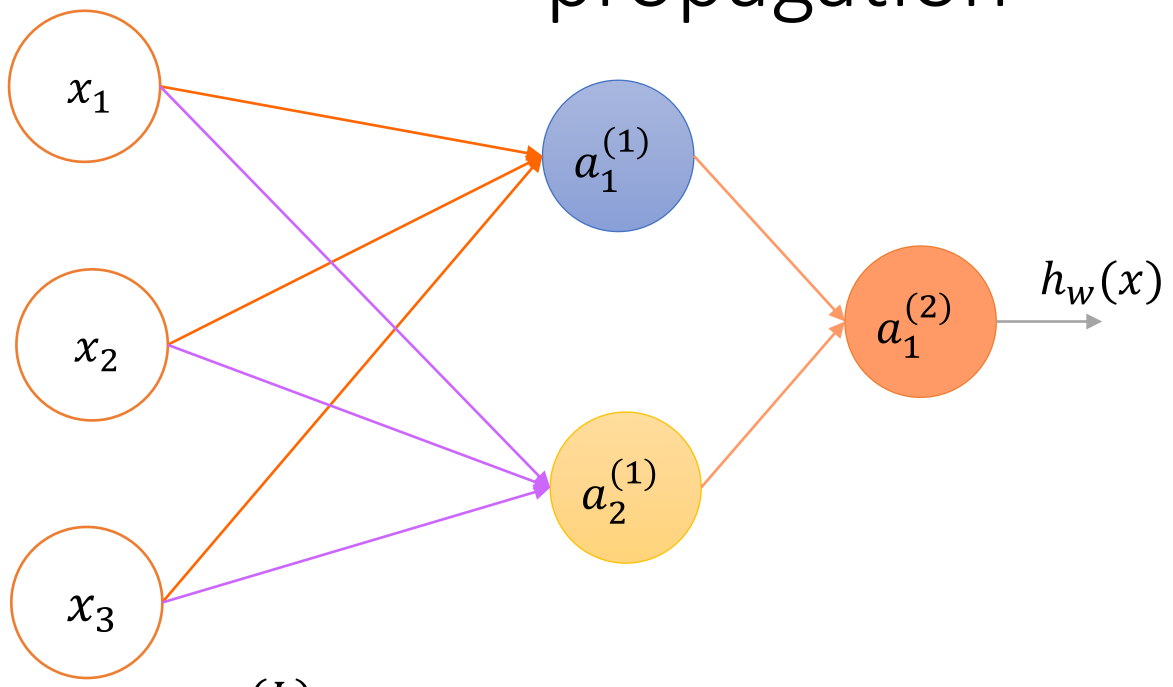
$$a_2^{(1)} = g\left(w_{21}^{(1)}x_1 + w_{22}^{(1)}x_2 + w_{23}^{(1)}x_3 + b_2^{(1)}\right)$$

Réseau de neurones: propagation



$$a_1^{(2)} = g\left(w_{10}^{(2)} a_1^{(1)} + +w_{12}^{(2)} a_2^{(1)} + b_1^{(2)}\right)$$

Réseau de neurones: propagation



$$a_1^{(1)} = g\left(w_{11}^{(1)}x_1 + w_{12}^{(1)}x_2 + w_{13}^{(1)}x_3 + b_1^{(1)}\right)$$

$$a_2^{(1)} = g\left(w_{21}^{(1)}x_1 + w_{22}^{(1)}x_2 + w_{23}^{(1)}x_3 + b_2^{(1)}\right)$$

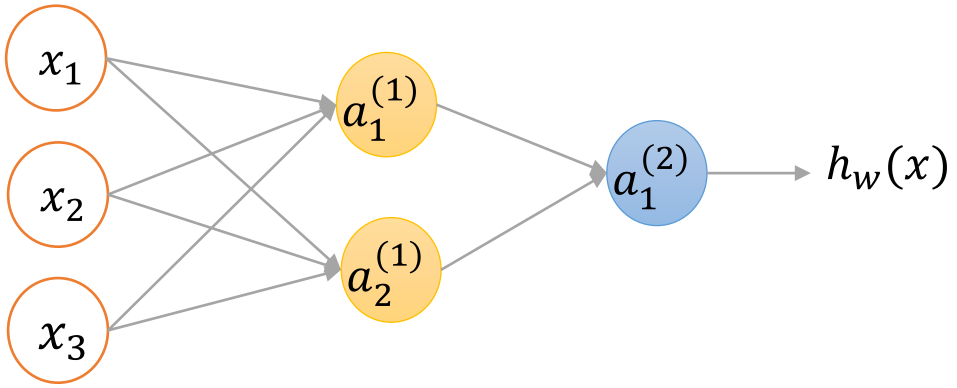
$$a_1^{(2)} = g\left(w_{11}^{(2)}a_1^{(1)} + w_{12}^{(2)}a_2^{(1)} + b_1^{(2)}\right)$$

$a_i^{(L)}$ est la sortie du neurone i de la couche L (la couche d'entré est la couche 0)

$w^{(k)}$ est la matrice des poids des connexions entre les neurones de la couche k-1 et la couche k.

$w_{ij}^{(1)}$ est le poids synaptique qui connecte le j$^{\text{ème}}$ neurone de la couche 0 et le i$^{\text{ème}}$ neurone de la couche 1

Réseau de neurones: propagation



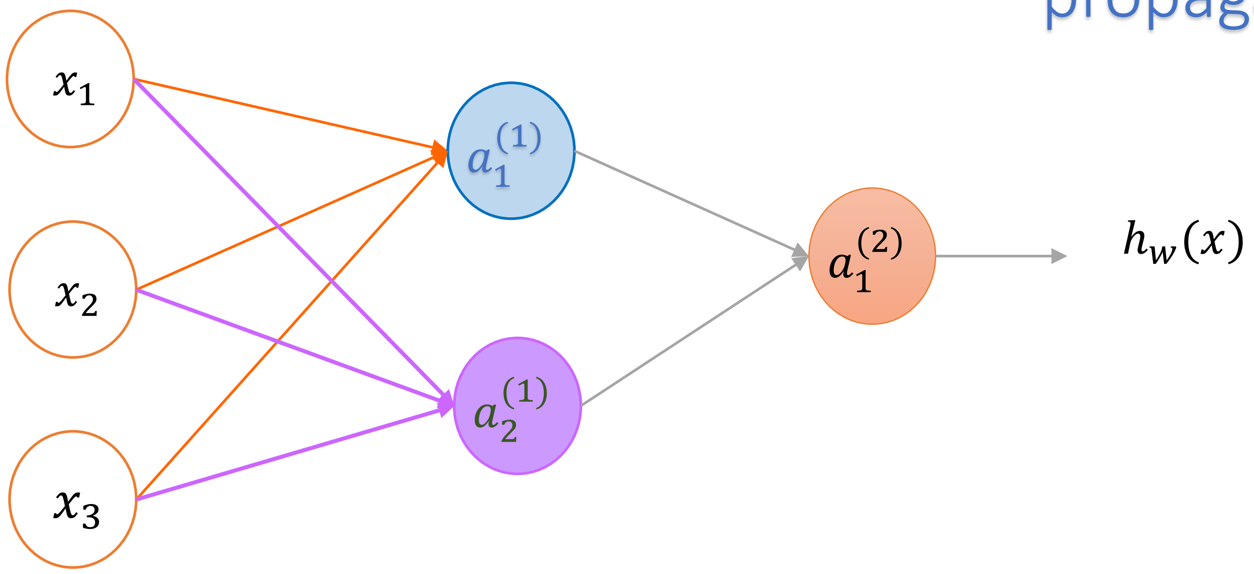
$$Z_1^{(1)} = w_{11}^{(1)} x_1 + w_{12}^{(1)} x_2 + w_{13}^{(1)} x_3 + b_1^{(1)}$$

$$Z_2^{(1)} = w_{21}^{(1)} x_1 + w_{22}^{(1)} x_2 + w_{23}^{(1)} x_3 + b_2^{(1)}$$

$$\Rightarrow \quad Z^{(1)} = \begin{bmatrix} Z_1^{(1)} \\ Z_2^{(1)} \end{bmatrix} \quad \Rightarrow \quad a^{(1)} = \begin{bmatrix} a_1^{(1)} \\ a_2^{(1)} \end{bmatrix} = \begin{bmatrix} g\left(Z_1^{(1)}\right) \\ g\left(Z_2^{(1)}\right) \end{bmatrix}$$

$$Z_1^{(2)} = +w_{11}^{(2)} a_1^{(1)} + w_{12}^{(2)} a_2^{(1)} + b_1^{(2)} \quad \Rightarrow \quad h_w(x) = a_1^{(2)} = g\left(Z_1^{(2)}\right)$$

Réseau de neurones propagation



$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \end{bmatrix} \qquad b^{(1)} = \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \end{bmatrix}$$

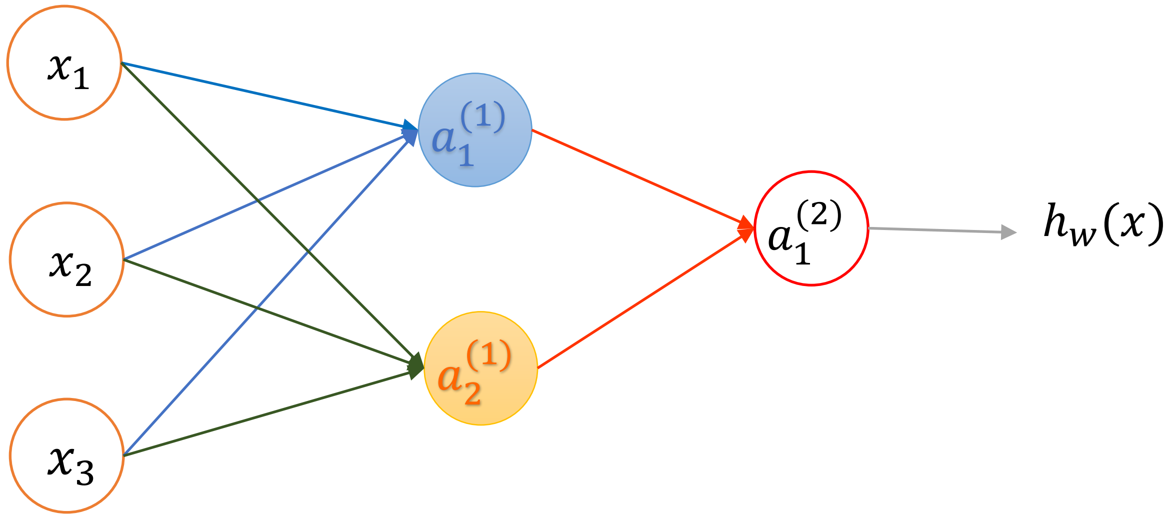
$$Z_1^{(1)} = w_{11}^{(1)} x_1 + w_{12}^{(1)} x_2 + w_{13}^{(1)} x_3 + b_1^{(1)}$$

$$Z_2^{(1)} = w_{21}^{(1)} x_1 + w_{22}^{(1)} x_2 + w_{23}^{(1)} x_3 + b_2^{(1)}$$

$$\Rightarrow \quad Z^{(1)} = \begin{bmatrix} Z_1^{(1)} \\ Z_2^{(2)} \end{bmatrix} = W^{(1)} x + b^{(1)}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_1^{(0)} \\ a_2^{(0)} \\ a_3^{(0)} \end{bmatrix} = a^{(0)} \qquad Z^{(1)} = W^{(1)} a^{(0)} + b^{(1)} \qquad a^{(1)} = \begin{bmatrix} a_1^{(1)} \\ a_2^{(1)} \end{bmatrix} = \begin{bmatrix} g(Z_1^{(2)}) \\ g(Z_2^{(2)}) \end{bmatrix}$$

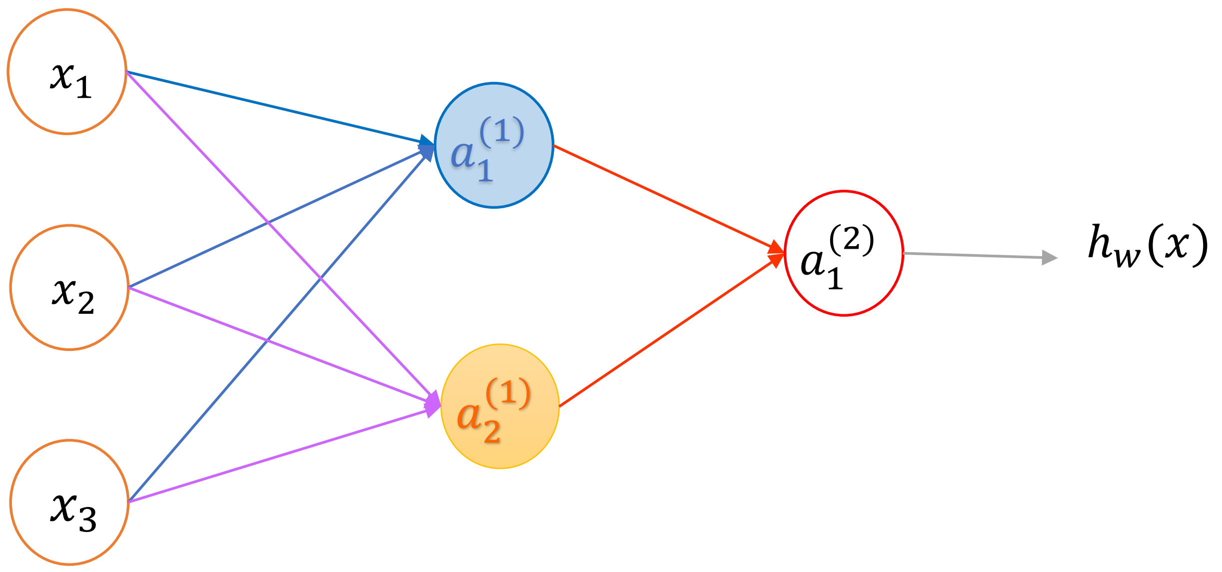
Réseau de neurones: propagation



$$a^{(1)} = \begin{bmatrix} a_1^{(1)} \\ a_2^{(1)} \end{bmatrix} \qquad W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} \end{bmatrix} \qquad b^{(2)} = \begin{bmatrix} b_1^{(2)} \end{bmatrix}$$

$$Z_1^{(2)} = W^{(2)} a^{(1)} + b^{(2)} \Rightarrow Z^{(2)} = \begin{bmatrix} Z_1^{(2)} \end{bmatrix} \Rightarrow h_w(x) = a_1^{(2)} = g\left(Z_1^{(2)}\right)$$

Réseau de neurones



$$W^{(1)} = \begin{bmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} \end{bmatrix} \quad b^{(1)} = \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \end{bmatrix} \quad Z^{(1)} = \begin{bmatrix} Z_1^{(1)} \\ Z_2^{(1)} \end{bmatrix} = W^{(1)}x + b^{(1)} \quad a^{(1)} = g(Z^{(1)})$$

$$W^{(2)} = \begin{bmatrix} w_{11}^{(2)} & w_{12}^{(2)} \end{bmatrix} \quad b^{(2)} = \begin{bmatrix} b_1^{(2)} \end{bmatrix} \quad Z_1^{(2)} = W^{(2)}a^{(1)} + b^{(2)} \quad a^{(2)} = g(Z^{(2)})$$

Réseau de neurones: propagation

$$a^{(0)} = X$$

$$Z^{(1)} = W^{(1)}a^{(0)} + b^{(0)}$$

$$a^{(1)} = g(Z^{(1)})$$

$$Z^{(2)} = W^{(2)}a^{(1)} + b^{(2)}$$

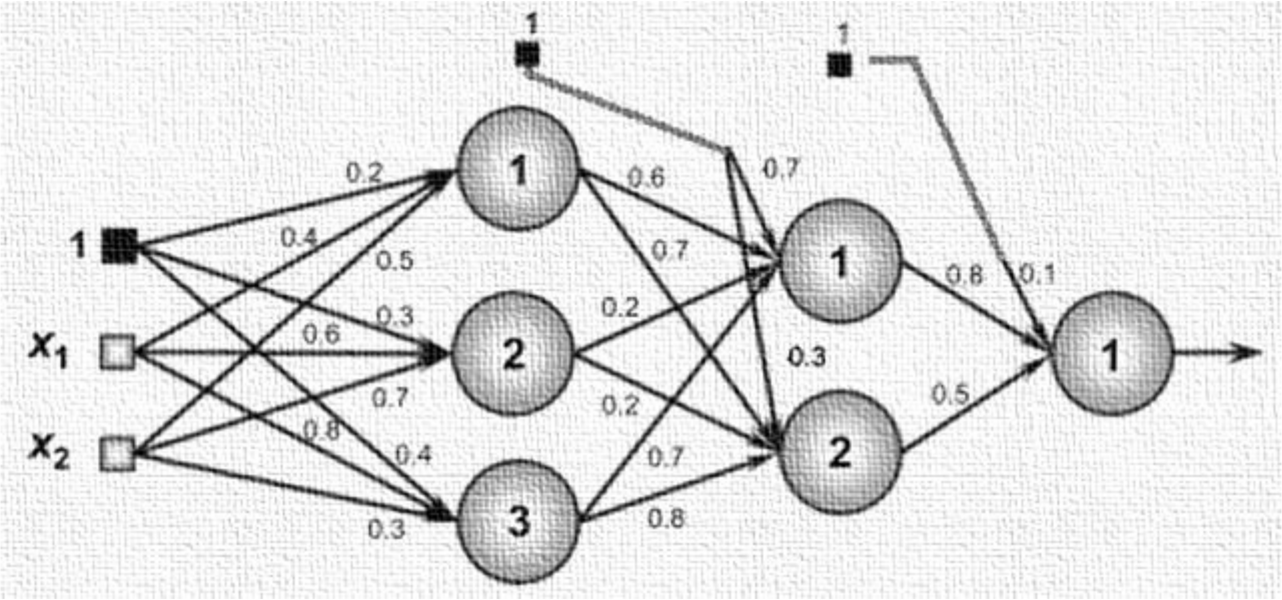
$$a^{(2)} = g(Z^{(2)})$$

$$Z^{(L)} = W^{(L)}a^{(L-1)} + b^{(L)}$$

$$a^{(L)} = g(Z^{(L)})$$

Réseau de neurones.
Exemple

$$X = a^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$



$$W^{(1)} = \begin{bmatrix} 0,4 & 0,5 \\ 0,6 & 0,7 \\ 0,8 & 0,3 \end{bmatrix} \qquad b^{(1)} = \begin{bmatrix} 0,2 \\ 0,3 \\ 0,4 \end{bmatrix}$$

$$W^{(2)} = \begin{bmatrix} 0,6 & 0,2 & 0,7 \\ 0,7 & 0,2 & 0,8 \end{bmatrix} \quad b^{(2)} = \begin{bmatrix} 0,7 \\ 0,3 \end{bmatrix}$$

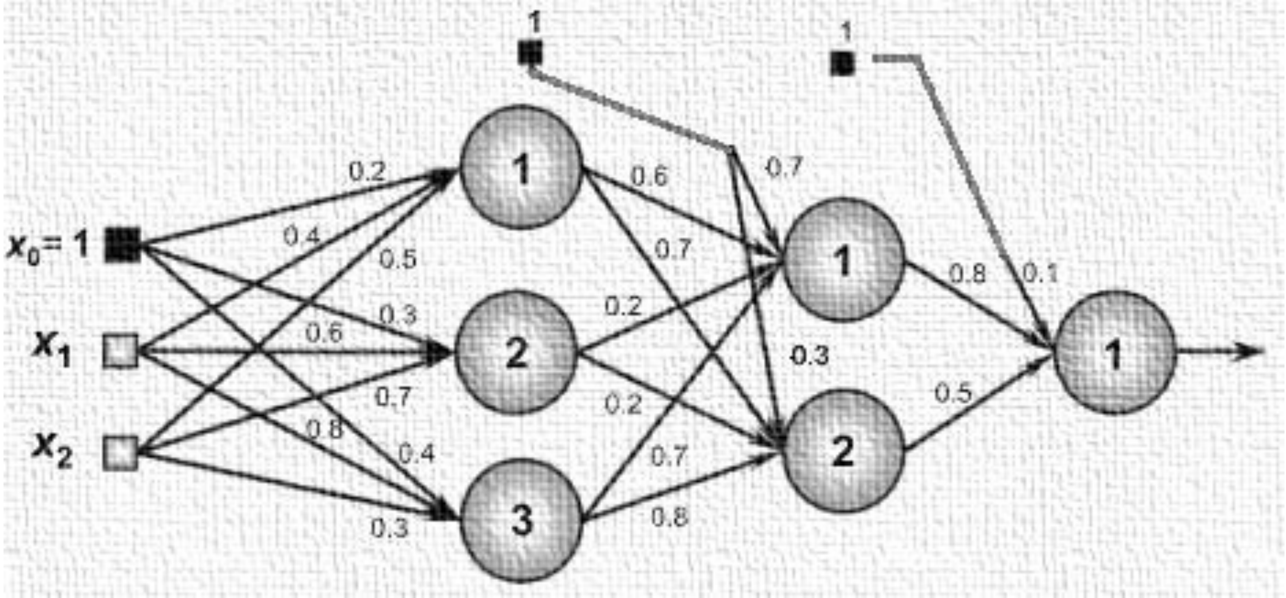
$$W^{(3)} = \begin{bmatrix} 0,8 & 0,5 \end{bmatrix} \qquad b^{(3)} = \begin{bmatrix} 0,1 \end{bmatrix}$$

La fonction d’activation utilisée dans les deux couches cachées est la fonction **tanh**

La fonction d’activation utilisée dans la couche de sortie est la fonction **sigmoide**

Réseau de neurones.
Exemple

$$X = a^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

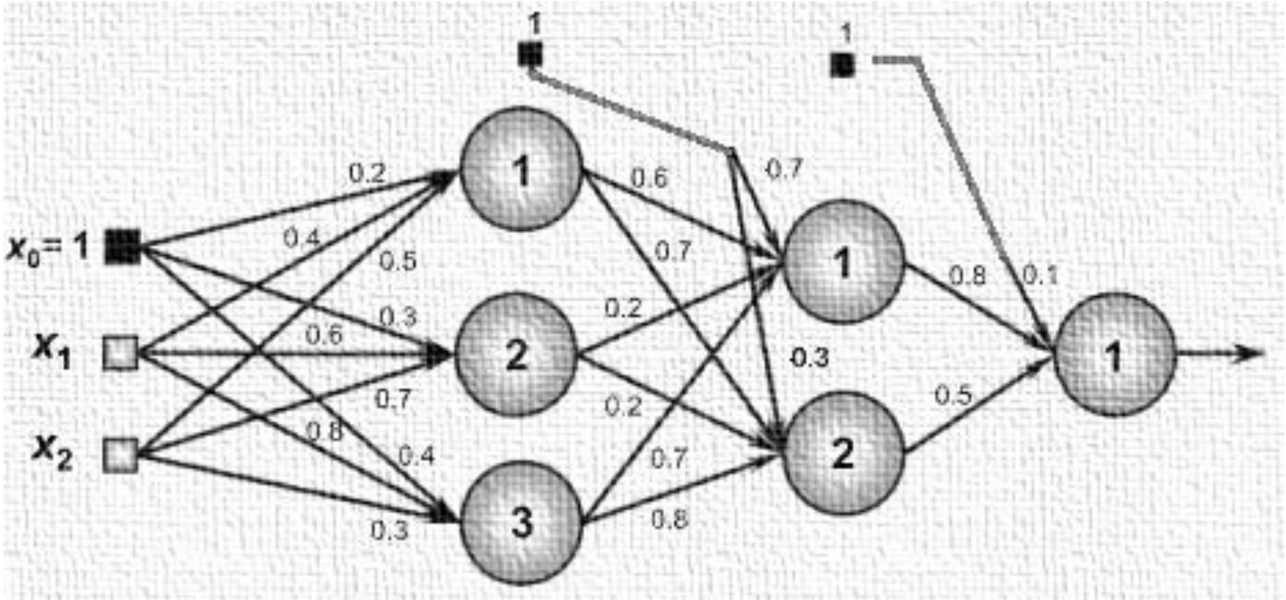


$$Z^{(1)} = W^{(1)} a^{(0)} + b^{(1)} = \begin{bmatrix} 0,4 & 0,5 \\ 0,6 & 0,7 \\ 0,8 & 0,3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0,2 \\ 0,3 \\ 0,4 \end{bmatrix}$$

$$Z^{(1)} = \begin{bmatrix} 0,6 \\ 0,9 \\ 1,2 \end{bmatrix} \Rightarrow a^{(1)} = \begin{bmatrix} g(0,6) \\ g(0,9) \\ g(1,2) \end{bmatrix} = \begin{bmatrix} \tanh(0,6) \\ \tanh(0,9) \\ \tanh(1,2) \end{bmatrix} = \begin{bmatrix} 0,537 \\ 0,716 \\ 0,833 \end{bmatrix}$$

$$a^{(1)} = \begin{bmatrix} 0,537 \\ 0,716 \\ 0,833 \end{bmatrix}$$

Réseau de neurones. Exemple



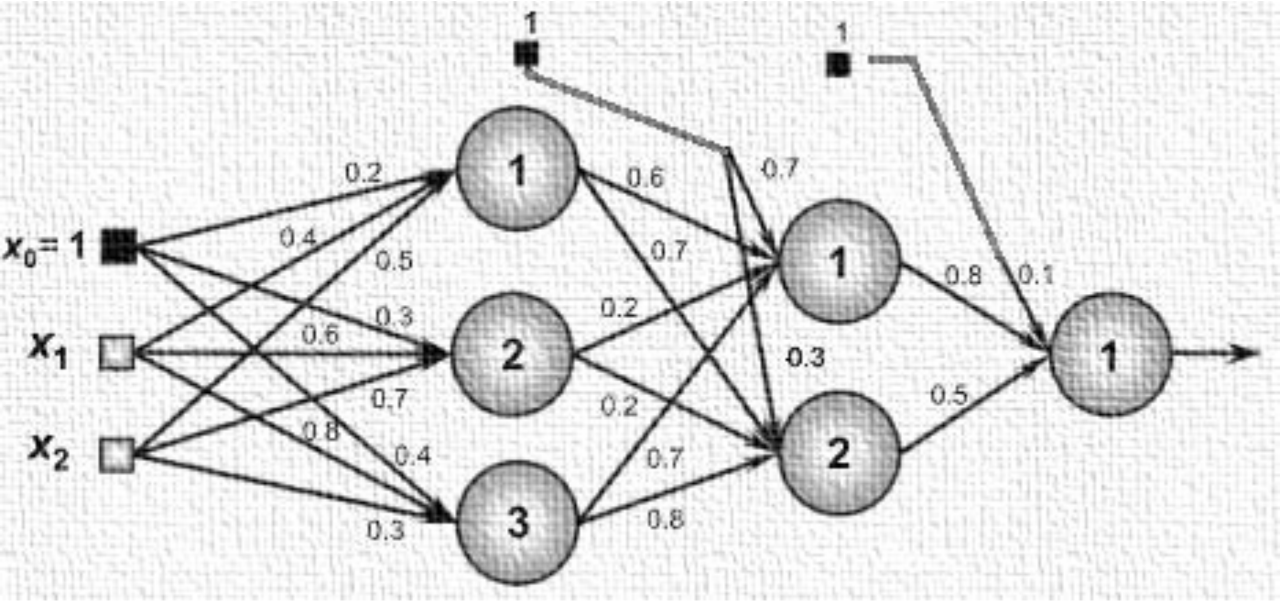
$$a^{(1)} = \begin{bmatrix} 0,537 \\ 0,716 \\ 0,833 \end{bmatrix}$$

$$Z^{(2)} = W^{(2)}a^{(1)} + b^{(2)} = \begin{bmatrix} 0,6 & 0,2 & 0,7 \\ 0,7 & 0,2 & 0,8 \end{bmatrix} \begin{bmatrix} 0,537 \\ 0,716 \\ 0,833 \end{bmatrix} + \begin{bmatrix} 0,7 \\ 0,3 \end{bmatrix}$$

$$Z^{(2)} = \begin{bmatrix} 1,749 \\ 1,486 \end{bmatrix} \Rightarrow \begin{bmatrix} g(1,749) \\ g(1,486) \end{bmatrix} = \begin{bmatrix} \tanh(1,749) \\ \tanh(1,486) \end{bmatrix} = \begin{bmatrix} 0,941 \\ 0,902 \end{bmatrix}$$

$$a^{(2)} = \begin{bmatrix} 0,941 \\ 0,902 \end{bmatrix}$$

Réseau de neurones : propagation
Exemple

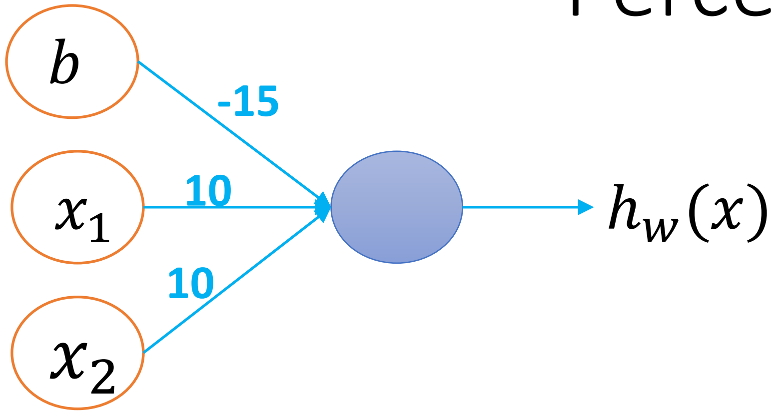


$$Z^{(3)} = W^{(3)}a^{(2)} + b^{(3)} = [0{,}8 \quad 0{,}5] \begin{bmatrix} 0{,}941 \\ 0{,}902 \end{bmatrix} + [0{,}1]$$

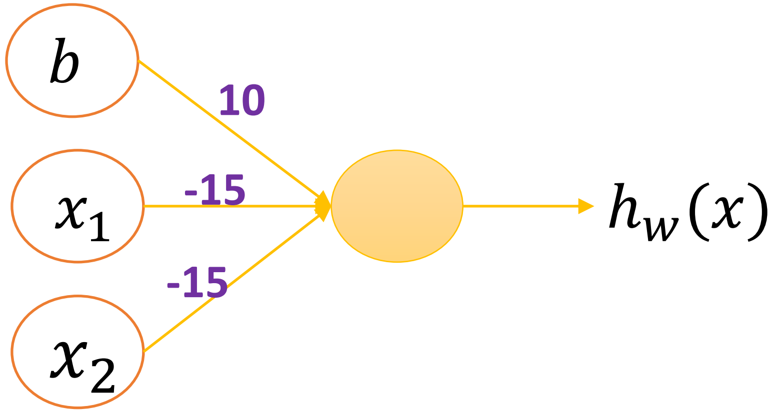
$$Z^{(3)} = [1{,}29] \qquad [g(1{,}29)] = [\sigma(1{,}29)] = [0{,}72]$$

$$h_w(x) = a^{(3)} = [\textcolor{red}{0{,}784}]$$

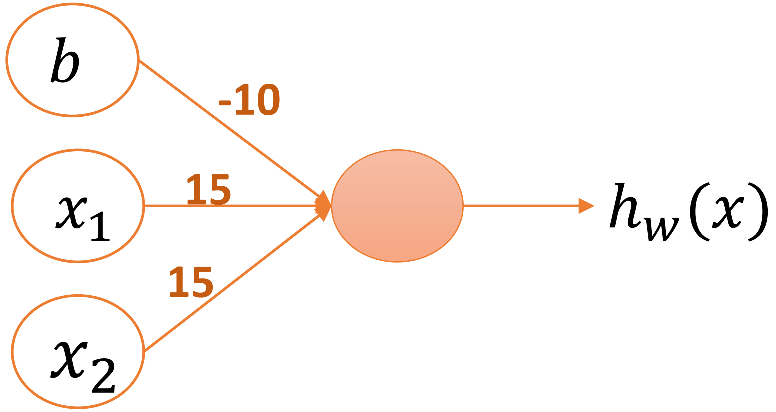
Perceptron multicouches



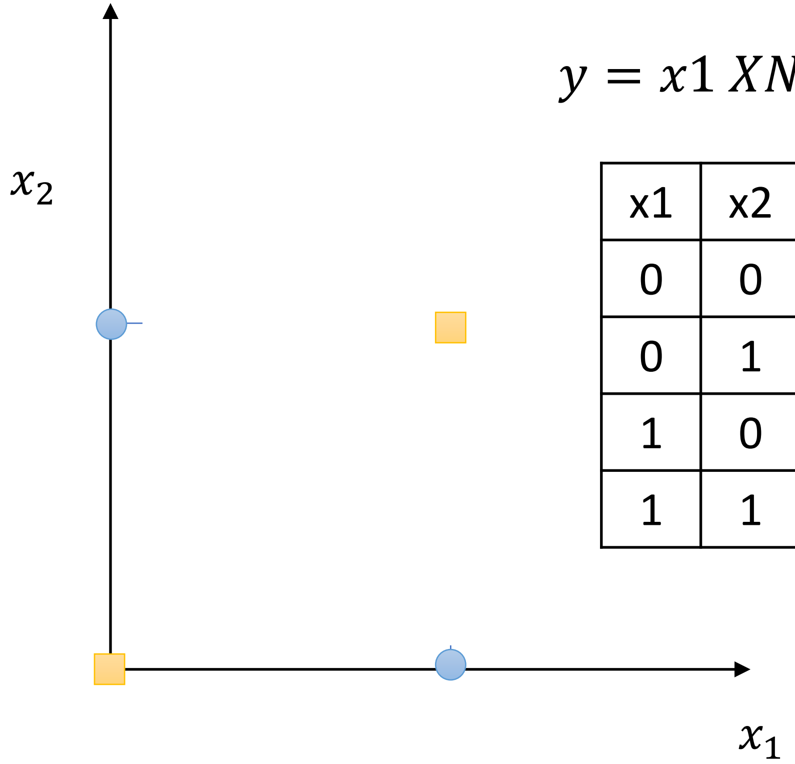
$y = x1\ et\ x2$



$y = non\ (x1) et\ non(x2)$



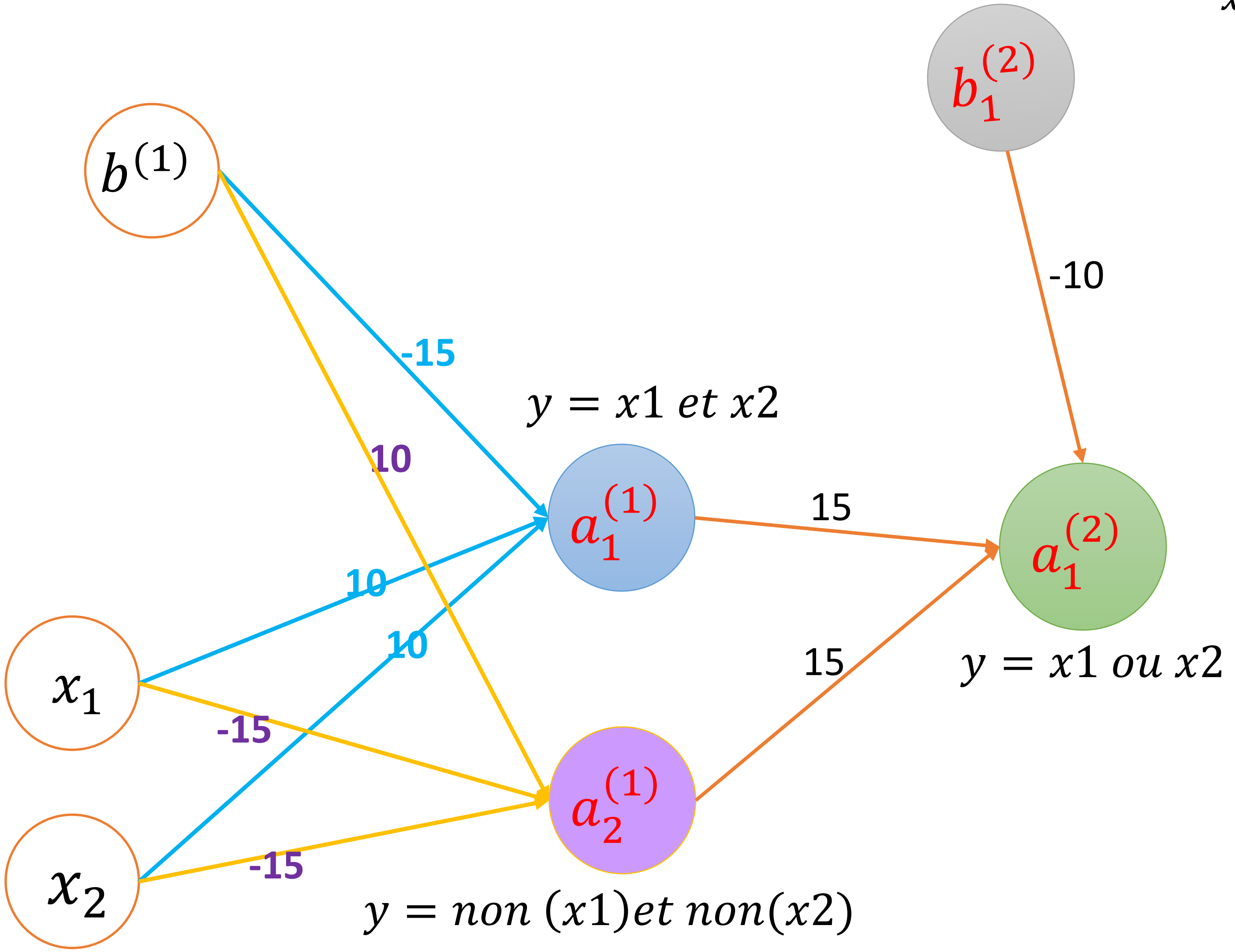
$y = x1\ ou\ x2$

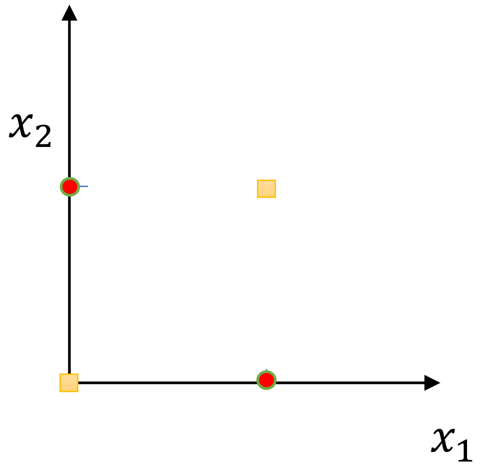


$y = x1\ XNOR\ x2$

x1	x2	Y
0	0	1
0	1	0
1	0	0
1	1	1

Perceptron multicouches : propagation



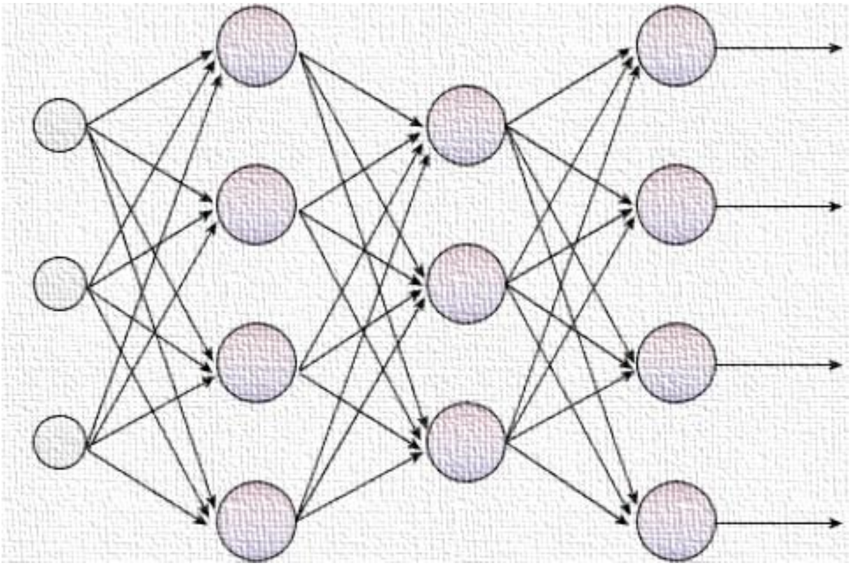


$y = x1\ XNOR\ x2$

x1	x2	Y
0	0	1
0	1	0
1	0	0
1	1	1

x_1	x_2	$a_1^{(1)}$	$a_2^{(1)}$	$a_1^{(2)}$
0	0	0	1	1
0	1	0	0	0
1	0	0	0	0
1	1	1	0	1

Sortie désirée



A

$$y = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

B

$$y = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

C

$$y = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

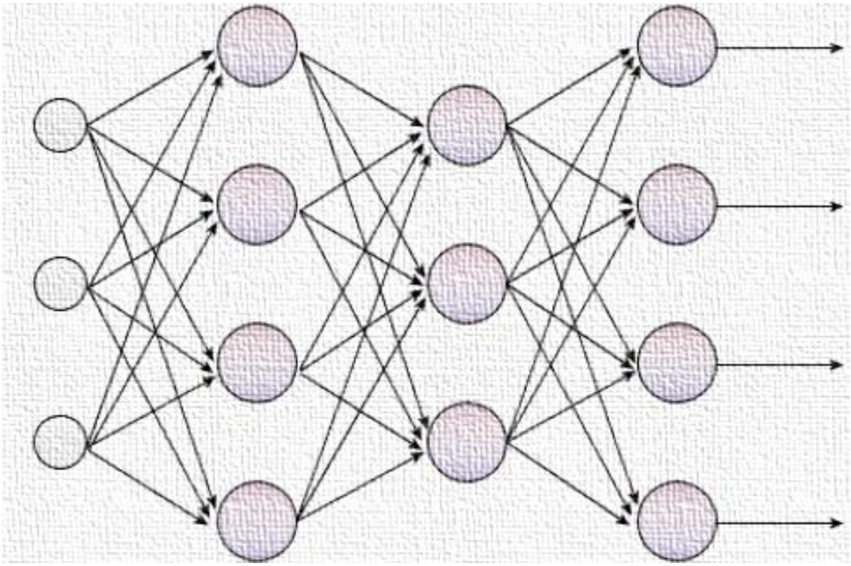
D

$$y = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

C

$$y = \begin{bmatrix} 0{,}2 \\ 0{,}5 \\ 0{,}8 \\ 0{,}1 \end{bmatrix}$$

Fonction cout

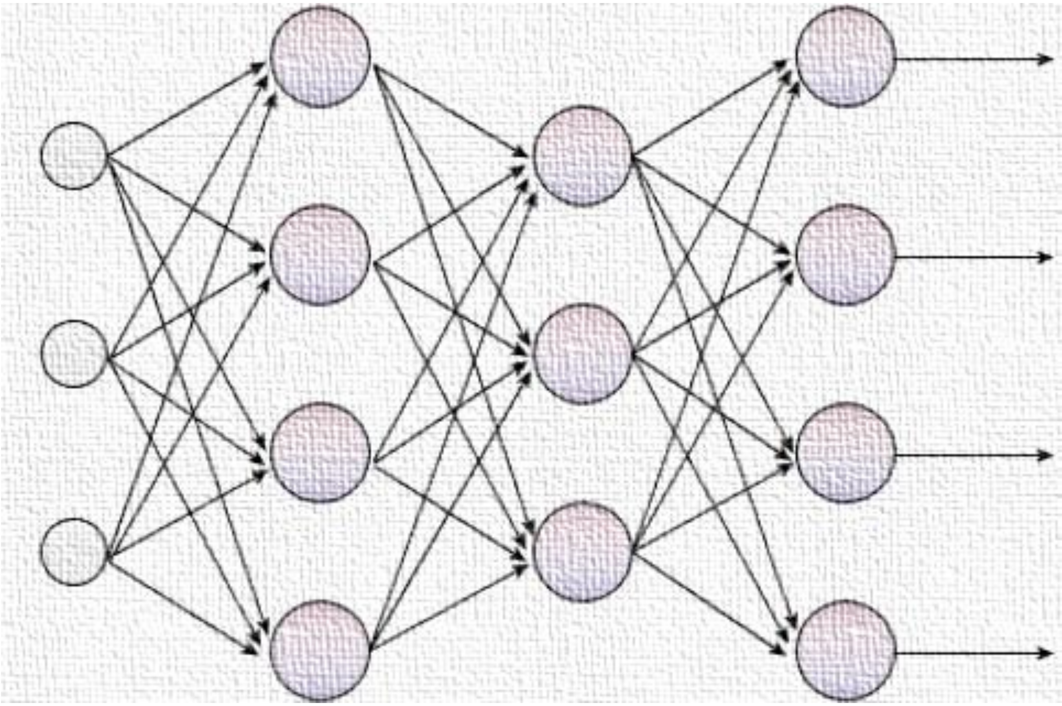


$$J(W,b) = -\frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{S_L} y_j^{(i)}\log\left(\left(h_w\left(x^{(i)}\right)\right)_j\right) + \left(1 - y_j^{(i)}\right)\log\left(1 - \left(h_w\left(x^{(i)}\right)\right)_j\right)$$

S_L est le nombre de neurones de la couche de sortie

$\left(h_w\left(x^{(i)}\right)\right)_j$ est la sortie du neurone j de la couche de sortie

Fonction cout



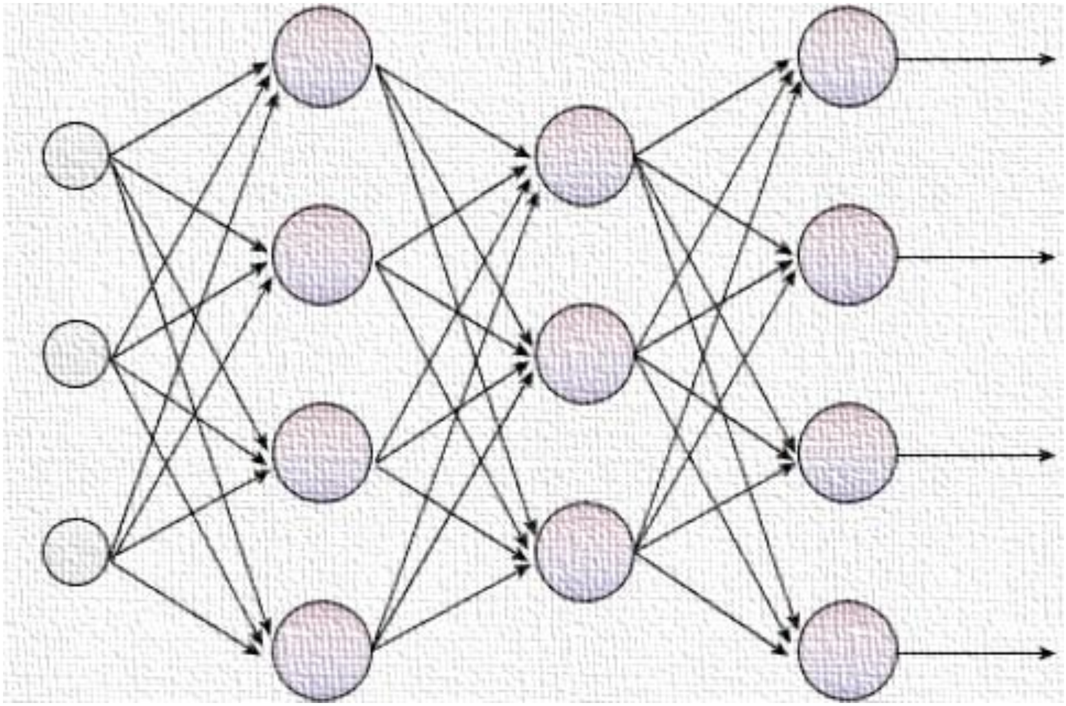
$$J(W,b) = -\frac{1}{2m}\sum_{i=1}^{m}\left(h_{w,b}\left(x^{(i)}\right)_j - y_j\right)^2$$

$$\min_{W} J(W,b)$$

Réseaux de neurones artificiels (séance 2)

Aziz KHAMJANE

Apprentissage du réseau de neurones

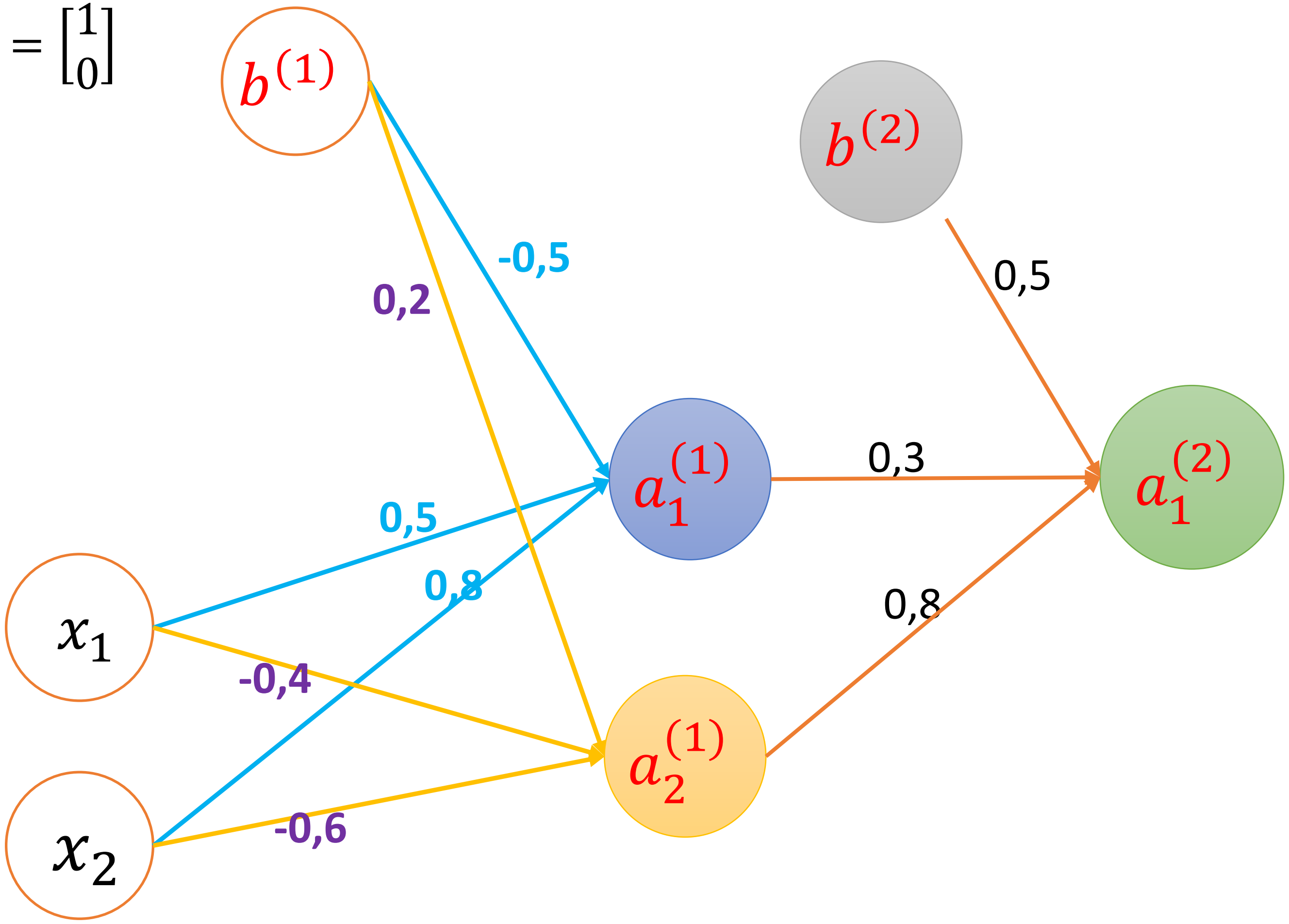


$$J(W,b) = -\frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{k}\left(h_w\left(x^{(i)}\right)_j - y_j\right)^2$$

$$\min_{W,b} J(W,b)$$

Calcul de l'erreur

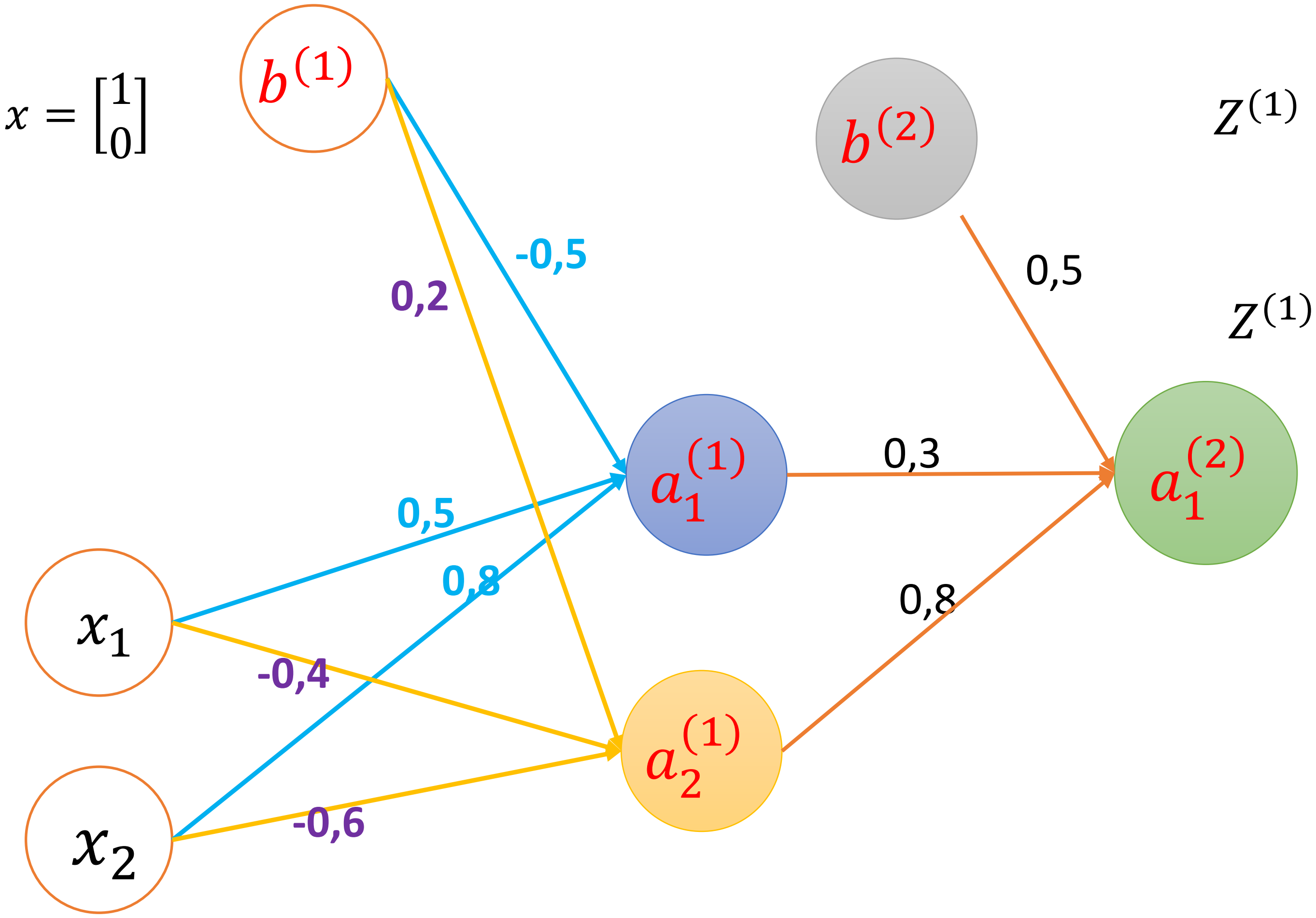
$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$



Calculer la sortie de ce réseau sachant que la fonction d'activation utilisée est la fonction sigmoide.

Calcul de l'erreur

$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$



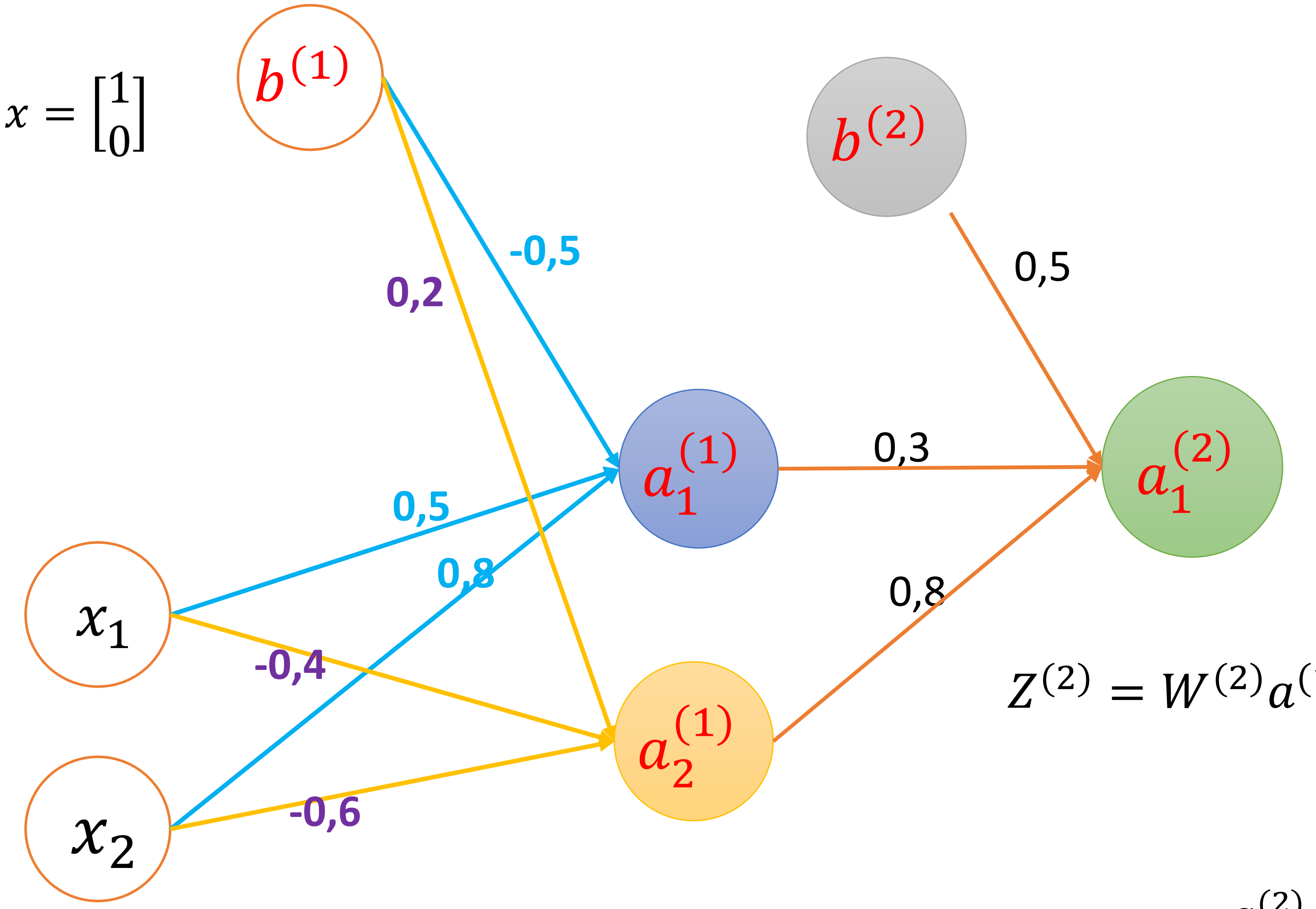
$$Z^{(1)} = W^{(1)}a^{(0)} + b^{(1)} = \begin{bmatrix} 0{,}5 & 0{,}8 \\ -0{,}4 & 0{,}6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -0{,}5 \\ 0{,}2 \end{bmatrix}$$

$$Z^{(1)} = \begin{bmatrix} 0 \\ -0{,}2 \end{bmatrix} \Rightarrow \begin{bmatrix} \sigma(0) \\ \sigma(-0{,}2) \end{bmatrix} = \begin{bmatrix} 0{,}5 \\ 0{,}45 \end{bmatrix}$$

$$a^{(1)} = \begin{bmatrix} 0{,}5 \\ 0{,}45 \end{bmatrix}$$

Calcul de l'erreur

$x = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

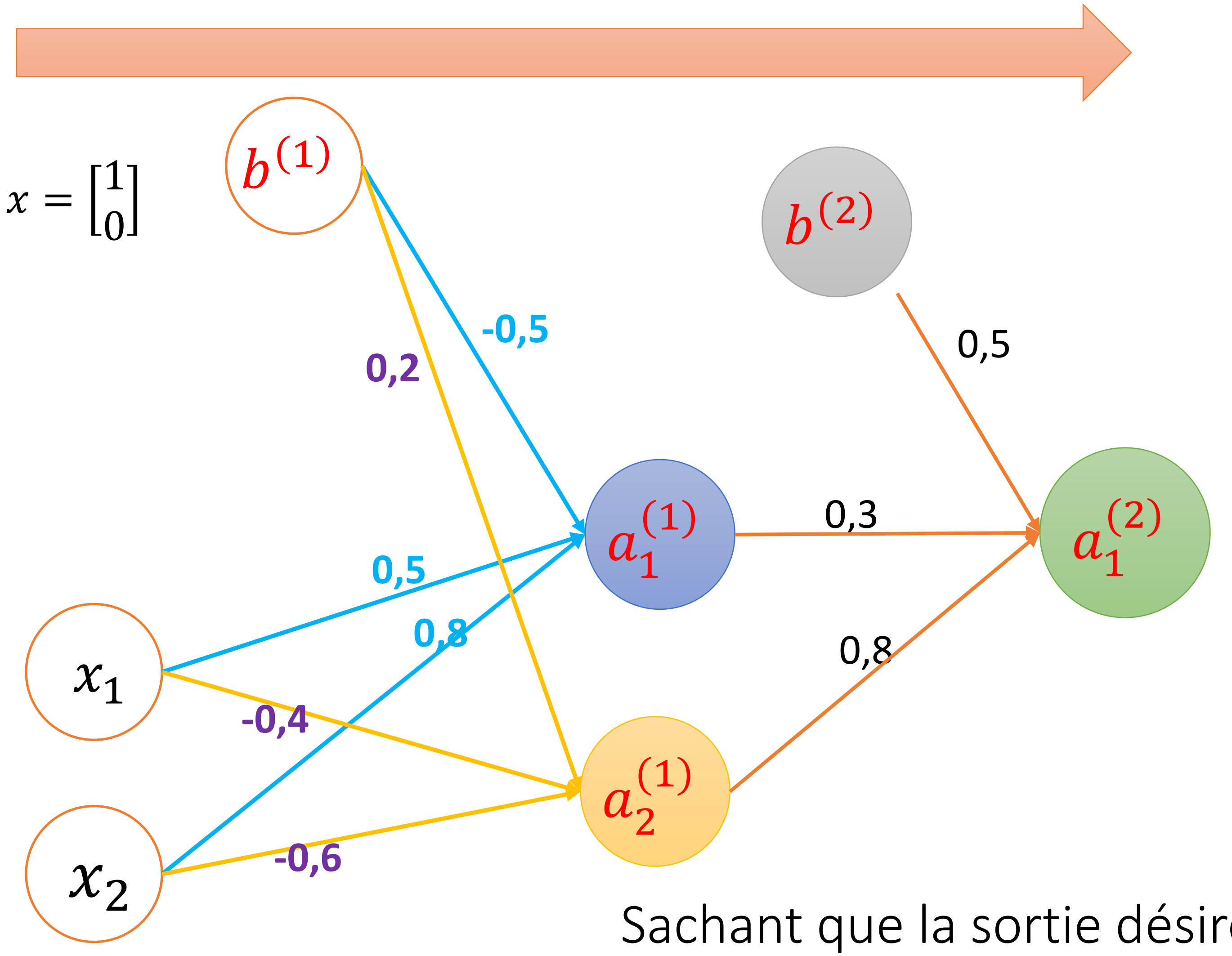


$$a^{(1)} = \begin{bmatrix} 0,5 \\ 0,45 \end{bmatrix}$$

$$Z^{(2)} = W^{(2)} a^{(1)} + b^{(2)} = \begin{bmatrix} 0,3 & 0,8 \end{bmatrix} \begin{bmatrix} 0,5 \\ 0,45 \end{bmatrix} + [0,5] = 1,035$$

$$a^{(2)} = \sigma(1,035) = 0,73$$

Calcul de l'erreur

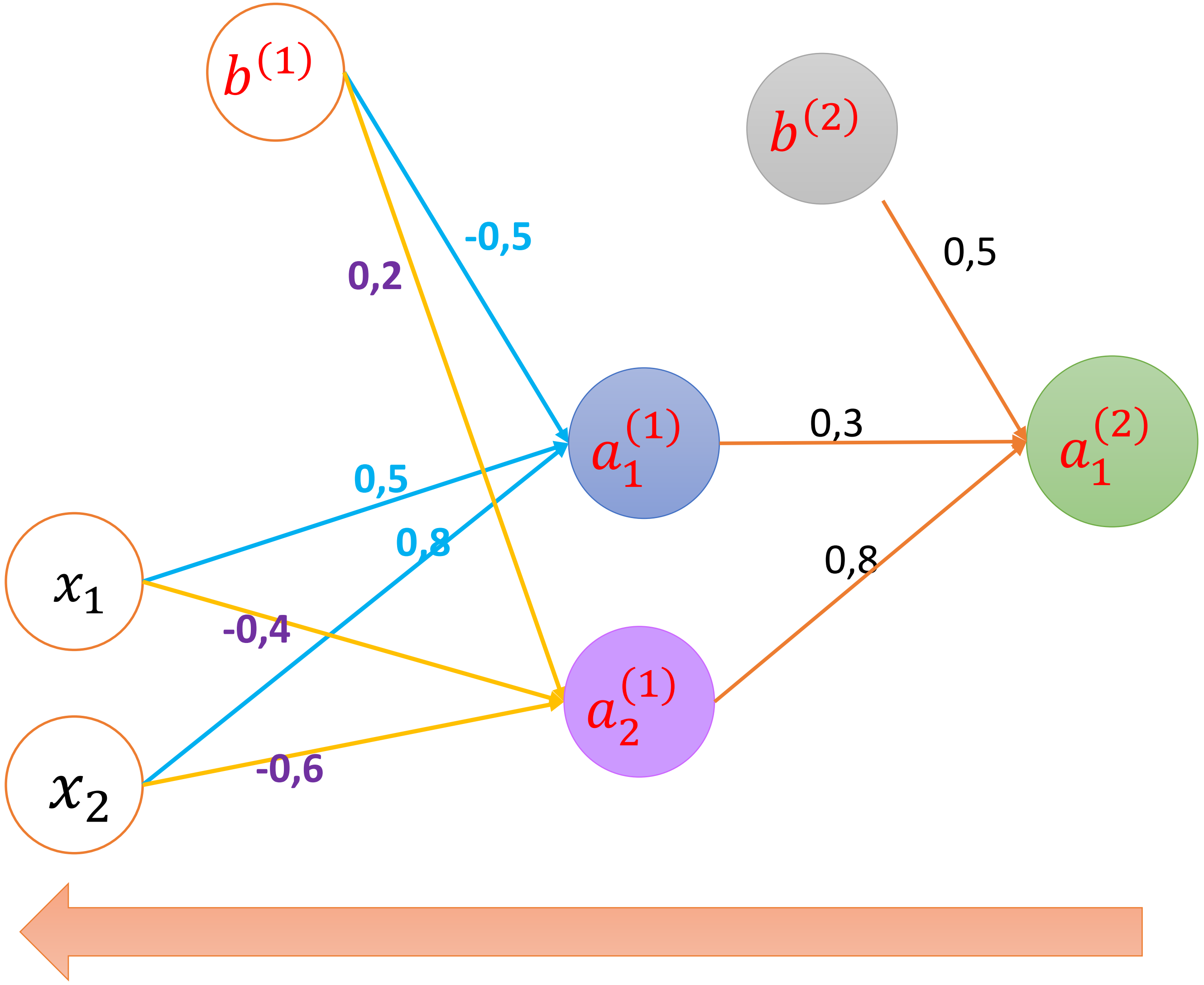


$$a^{(2)} = \sigma(1{,}01) = 0{,}73$$

Sachant que la sortie désirée est 0, alors l'erreur est :

$$\mathrm{J}(W,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(h_w\left(x^{(i)}\right) - y_j\right)^2 = \frac{1}{2}\left(a_1^{(2)} - y\right)^2 = \frac{1}{2}(0{,}73 - 0)^2 = 0{,}266$$

Règle de mise à jour des poids

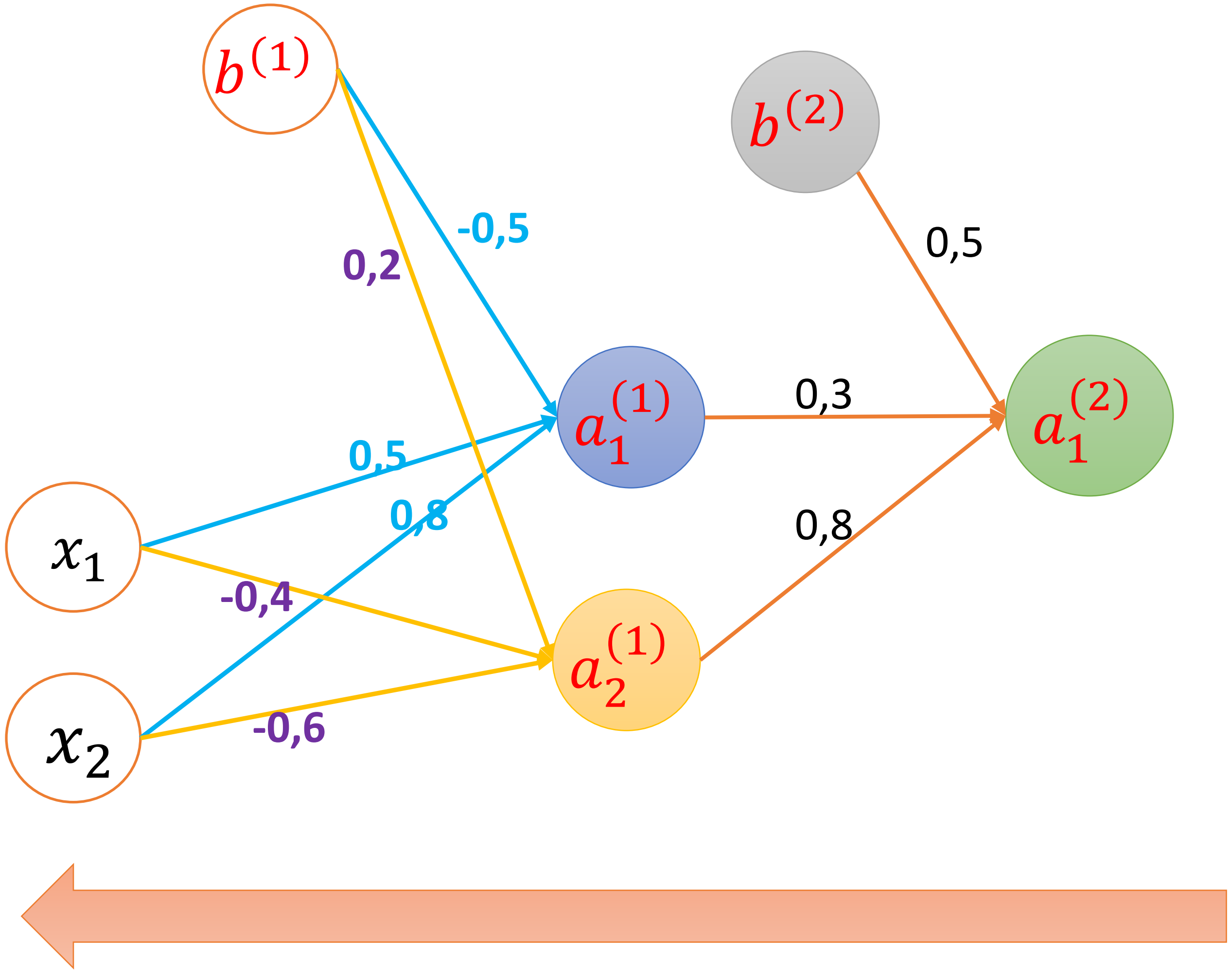


$$w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha \frac{\partial J(W,b)}{\partial w_{ij}^{(l)}}$$

$$w_{11}^{(2)} = w_{11}^{(2)} - \alpha \frac{\partial J(W,b)}{\partial w_{11}^{(2)}}$$

$$b_{j}^{(l)} = b_{j}^{(l)} - \alpha \frac{\partial J(W,b)}{\partial b_{j}^{(l)}}$$

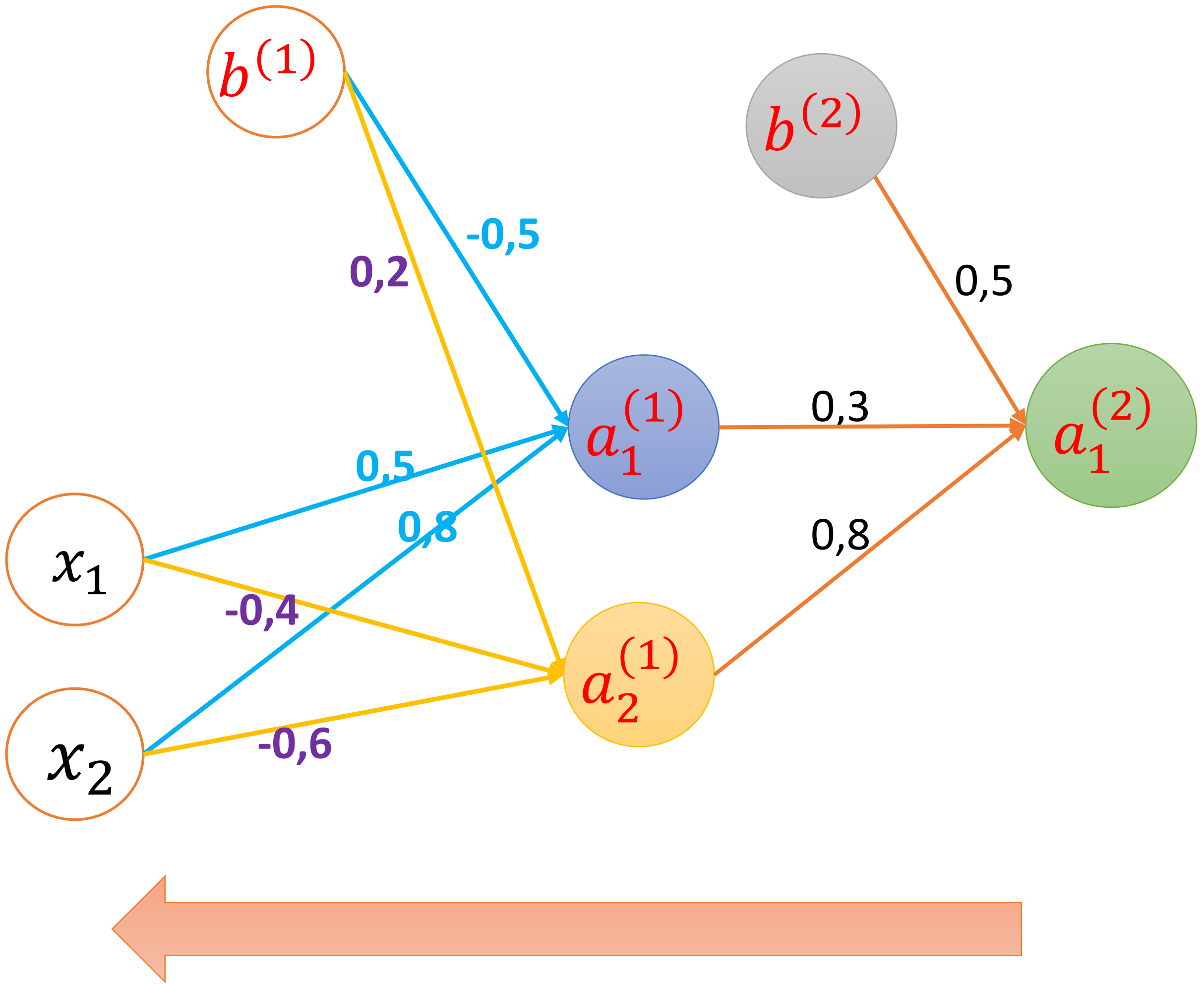
Calcul des des dérivées partielles



$$w_{11}^{(2)} = w_{11}^{(2)} - \alpha \frac{\partial J(W)}{\partial w_{11}^{(2)}}$$

$$\frac{\partial J(W)}{\partial w_{11}^{(2)}} = \frac{\partial J(W)}{\partial a_1^{(2)}} \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \frac{\partial z_1^{(2)}}{w_{11}^{(2)}}$$

Calcul des dérivés partielles



$$\frac{\partial\, J(W)}{\partial w_{11}^{(2)}} = \frac{\partial J(W)}{\partial a_1^{(2)}} \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \frac{\partial z_1^{(2)}}{w_{11}^{(2)}}$$

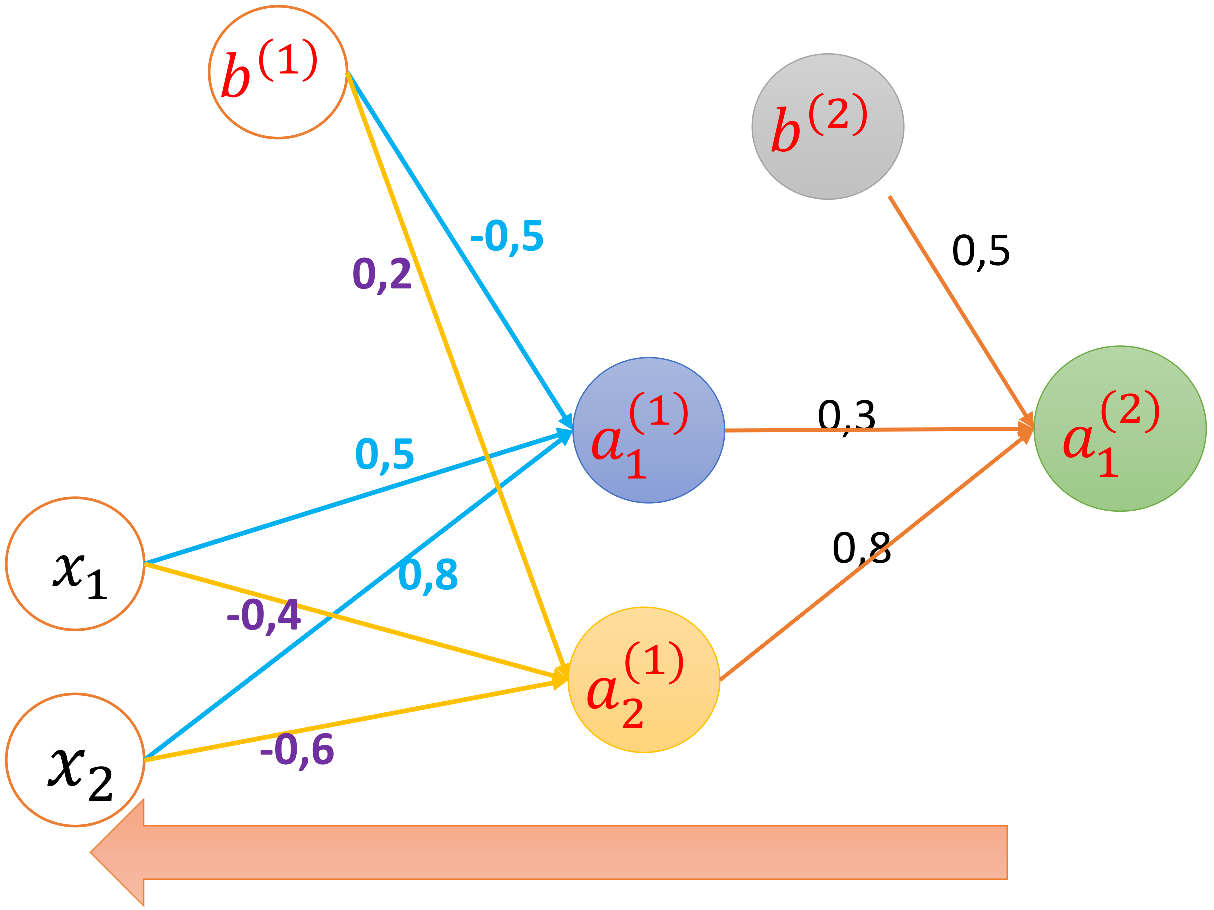
$$\frac{\partial J(W)}{\partial a_1^{(2)}} = \frac{\partial}{\partial a_1^{(2)}} \left(\frac{1}{2}\left(a_1^{(2)} - y\right)^2\right) = \left(a_1^{(2)} - y\right)$$

$$\frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} = \frac{\partial \sigma(z_1^{(2)})}{\partial z_1^{(2)}} = \frac{\partial \left(\frac{1}{1 + e^{-z_1^{(2)}}}\right)}{\partial z_1^{(2)}} = a_1^{(2)}(1 - a_1^{(2)})$$

$$\frac{\partial z_1^{(2)}}{\partial w_{11}^{(2)}} = \frac{\partial (w_{11}^{(2)} a_1^{(1)} + w_{12}^{(2)} a_2^{(1)} + b_1^{(2)})}{\partial w_{11}^{(2)}} = a_1^{(1)}$$

$$\frac{\partial J(W)}{\partial w_{11}^{(2)}} = \left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)}) a_1^{(1)}$$

Calcul des dérivées partielles



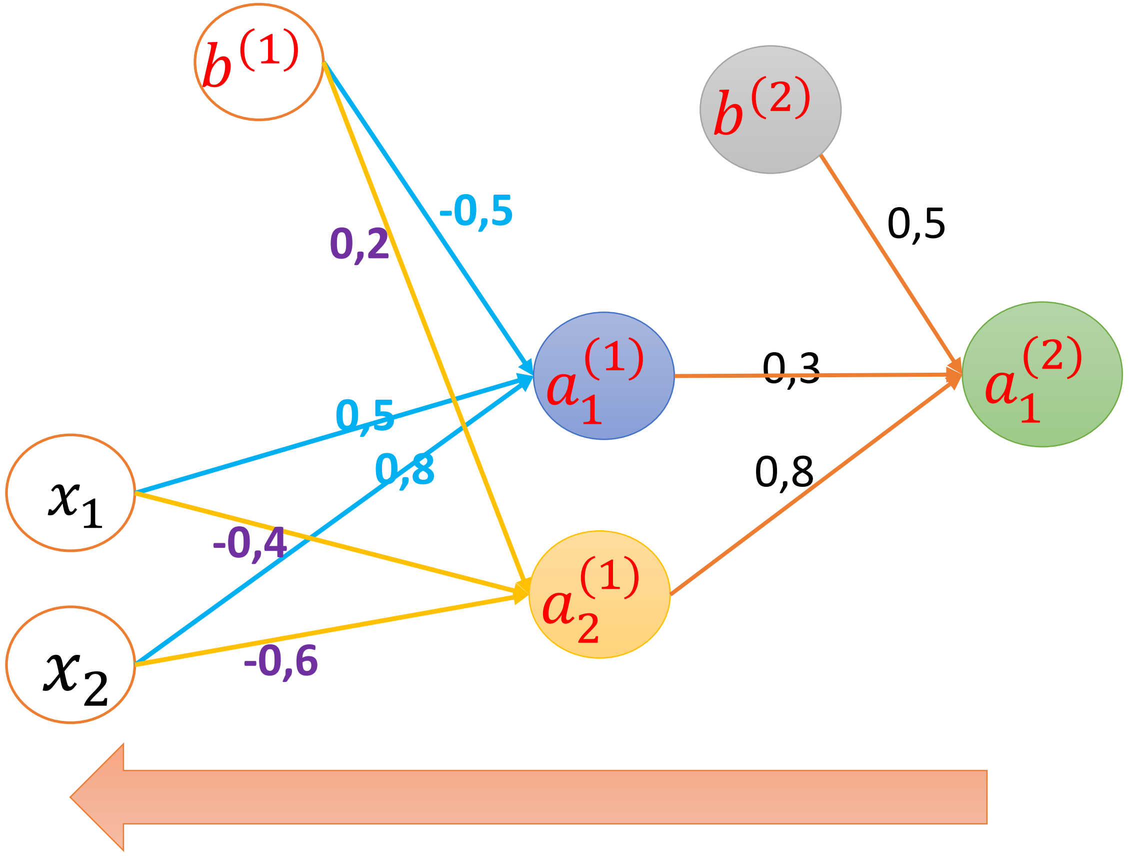
$$\frac{\partial J(W)}{\partial w_{11}^{(2)}} = \left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)}) a_1^{(1)}$$

$$\frac{\partial J(W)}{\partial w_{12}^{(2)}} = \left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)}) a_2^{(1)}$$

$$\frac{\partial J(W)}{\partial b_1^{(2)}} = \left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)}) \times 1$$

Mise à jour des poids $w_{ij}^{(2)}$

$x = a^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$



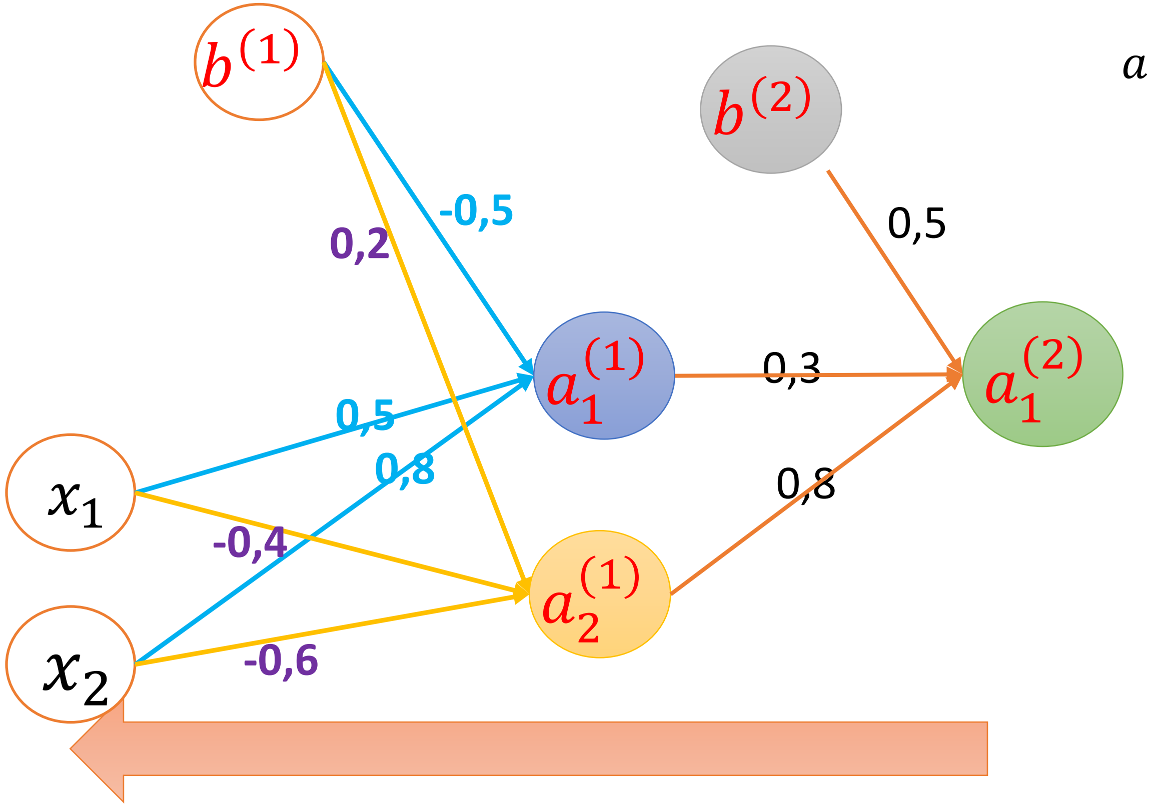
$$b_1^{(2)} = b_1^{(2)} - \alpha\left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)})$$

$$w_{12}^{(2)} = w_{12}^{(2)} - \alpha\left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)})a_1^{(1)}$$

$$w_{11}^{(2)} = w_{11}^{(2)} - \alpha\left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)})a_2^{(1)}$$

Mise à jour des poids $w_{ij}^{(2)}$

$x = a^{(0)} = \begin{bmatrix}1\\0\end{bmatrix}$ $\quad y = 0$



$$b_1^{(2)} = b_1^{(2)} - \alpha\left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)})$$

$a^{(1)} = \begin{bmatrix}0,5\\0,45\end{bmatrix}$ $\quad a_1^{(2)} = 0,73$ $\quad w^{(2)} = [0,3 \quad 0,8]$ $\quad b^{(2)} = [0,5]$

$$b_1^{(2)} = 0,5 - 0,1 \times (0,73 - 0) \times 0,73 \times (1 - 0,73)$$

$$b_1^{(2)} = 0,5 - 0,014 = 0,485$$

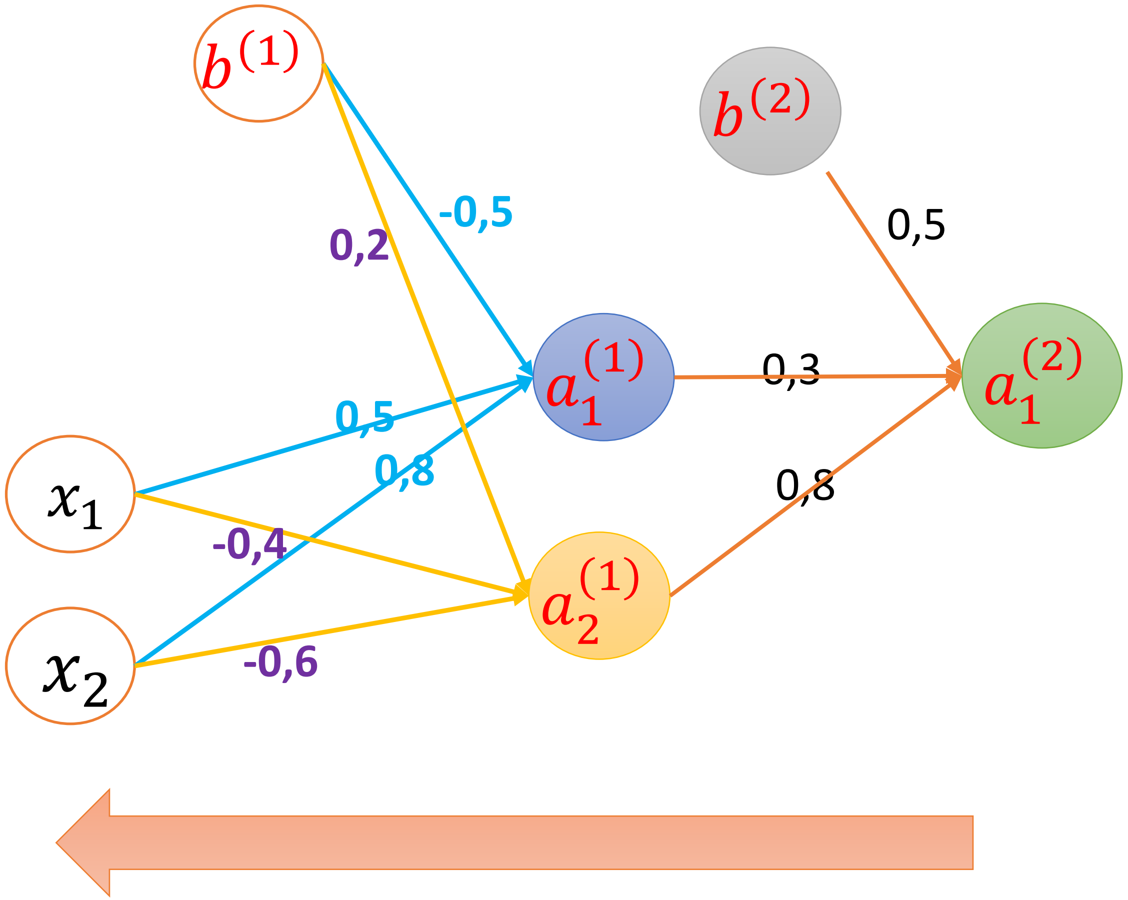
$$w_{11}^{(2)} = 0,3 - 0,1 \times (0,73 - 0) \times 0,73 \times (1 - 0,73) \times 0,5$$

$$w_{11}^{(2)} = 0,3 - 0,0071 = 0,292$$

$$w_{12}^{(2)} = 0,8 - 0,1 \times (0,73 - 0) \times 0,73 \times (1 - 0,73) \times 0,45$$

$$w_{12}^{(2)} = 0,8 - 0,0064 = 0,793$$

Calcul des dérivées partielles de la fonction cout par rapport aux poids $w_{ij}^{(1)}$

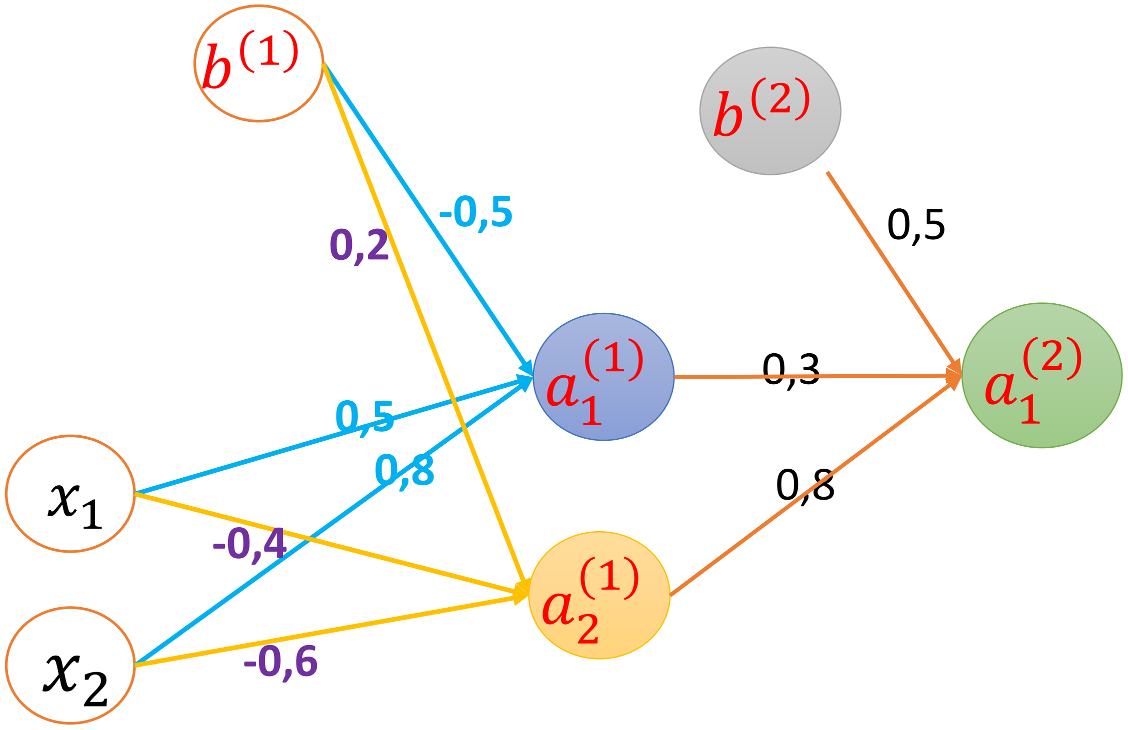


$$\frac{\partial J(W)}{\partial w_{ij}^{(1)}}$$

$$\frac{\partial J(W)}{\partial b_1^{(1)}} = \frac{\partial J(W)}{\partial a_1^{(2)}} \times \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \times \frac{\partial z_1^{(2)}}{\partial a_1^{(1)}} \times \frac{\partial a_1^{(1)}}{\partial z_1^{(1)}} \times \frac{\partial z_1^{(1)}}{\partial b_1^{(1)}}$$

$$\frac{\partial J(W)}{\partial w_{11}^{(1)}} = \frac{\partial J(W)}{\partial a_1^{(2)}} \times \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \times \frac{\partial z_1^{(2)}}{\partial a_1^{(1)}} \times \frac{\partial a_1^{(1)}}{\partial z_1^{(1)}} \times \frac{\partial z_1^{(1)}}{\partial w_{11}^{(1)}}$$

Calcul des dérivées partielles de la fonction cout par rapport aux poids $w_{ij}^{(1)}$



$$\frac{\partial J(W)}{\partial b_1^{(1)}} = \frac{\partial J(W)}{\partial a_1^{(2)}} \times \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \times \frac{\partial z_1^{(1)}}{\partial a_1^{(1)}} \times \frac{\partial a_1^{(1)}}{\partial z_1^{(1)}} \times \frac{\partial z_1^{(2)}}{\partial b_1^{(1)}}$$

$$\frac{\partial J(W)}{\partial a_1^{(2)}} \times \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} = \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right)$$

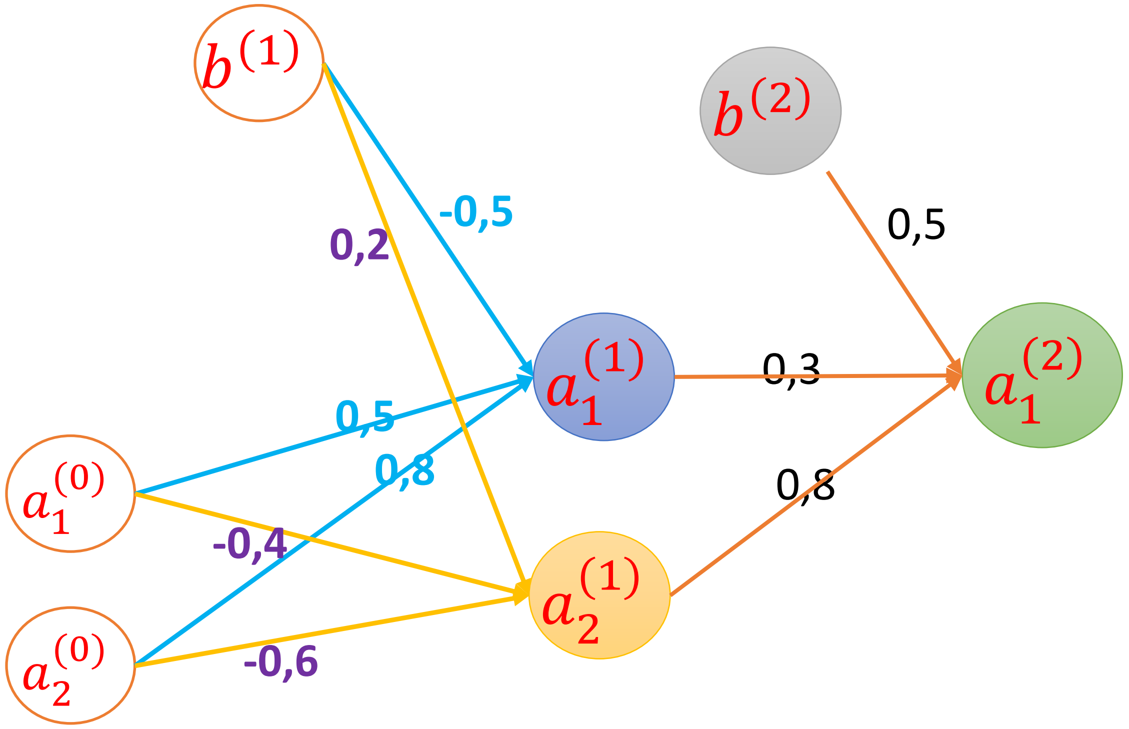
$$\frac{\partial z_1^{(2)}}{\partial a_1^{(1)}} = \frac{\partial (w_{11}^{(2)} a_1^{(1)} + w_{12}^{(2)} a_2^{(1)} + b_1^{(2)})}{\partial a_1^{(2)}} = w_{11}^{(2)}$$

$$\frac{\partial a_1^{(1)}}{\partial z_1^{(1)}} = a_1^{(1)}(1 - a_1^{(1)})$$

$$\frac{\partial z_1^{(1)}}{\partial b_1^{(1)}} = \frac{\partial (w_{11}^{(1)} a_1^{(0)} + w_{12}^{(1)} a_2^{(0)} + b_1^{(2)})}{\partial b_1^{(1)}} = 1$$

$$\frac{\partial J(W)}{\partial b_1^{(1)}} = \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right) w_{11}^{(2)} a_1^{(1)} \left(1 - a_1^{(1)}\right) 1$$

Calcul des dérivées partielles de la fonction cout par rapport aux poids $w_{ij}^{(1)}$



$$\frac{\partial J(W)}{\partial b_1^{(1)}} = \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right) w_{11}^{(2)} a_1^{(1)} \left(1 - a_1^{(1)}\right) 1$$

$$\frac{\partial J(W)}{\partial w_{11}^{(1)}} = \left(a_1^{(2)} - y\right) a_1^{(2)} (1 - a_1^{(2)}) w_{11}^{(2)} a_1^{(1)} \left(1 - a_1^{(1)}\right) a_1^{(0)}$$

$$\frac{\partial J(W)}{\partial w_{12}^{(1)}} = \left(a_1^{(2)} - y\right) a_1^{(2)} (1 - a_1^{(2)}) w_{11}^{(2)} a_1^{(1)} \left(1 - a_1^{(1)}\right) a_2^{(0)}$$

$$\frac{\partial J(W)}{\partial b_2^{(1)}} = \left(a_1^{(2)} - y\right) a_1^{(2)} (1 - a_1^{2}) w_{12}^{(2)} a_2^{(1)} \left(1 - a_2^{(1)}\right) 1$$

$$\frac{\partial J(W)}{\partial w_{21}^{(1)}} = \left(a_1^{(2)} - y\right) a_1^{(2)} (1 - a_1^{(2)}) w_{12}^{(2)} a_2^{(1)} \left(1 - a_2^{(1)}\right) a_1^{(0)}$$

$$\frac{\partial J(W)}{\partial w_{22}^{(1)}} = \left(a_1^{(2)} - y\right) a_1^{(2)} (1 - a_1^{(2)}) w_{12}^{(2)} a_2^{(1)} \left(1 - a_2^{(1)}\right) a_2^{(0)}$$

Rétropropagation

La fonction cout utilisée est l'erreur quadratique moyenne (sans utilisation du terme de régularisation)

L'erreur sur le neurone j de la couche de sortie est définie par :

$$\delta_j^{(L)} = \left(a_j^{(L)} - y_j\right) g'\left(z_j^{(L)}\right)$$

$$\delta^{(L)} = \left(a^{(L)} - y\right) .* \, g'\left(z^{(L)}\right)$$

Vecteur des $\delta_j^{(L)}$

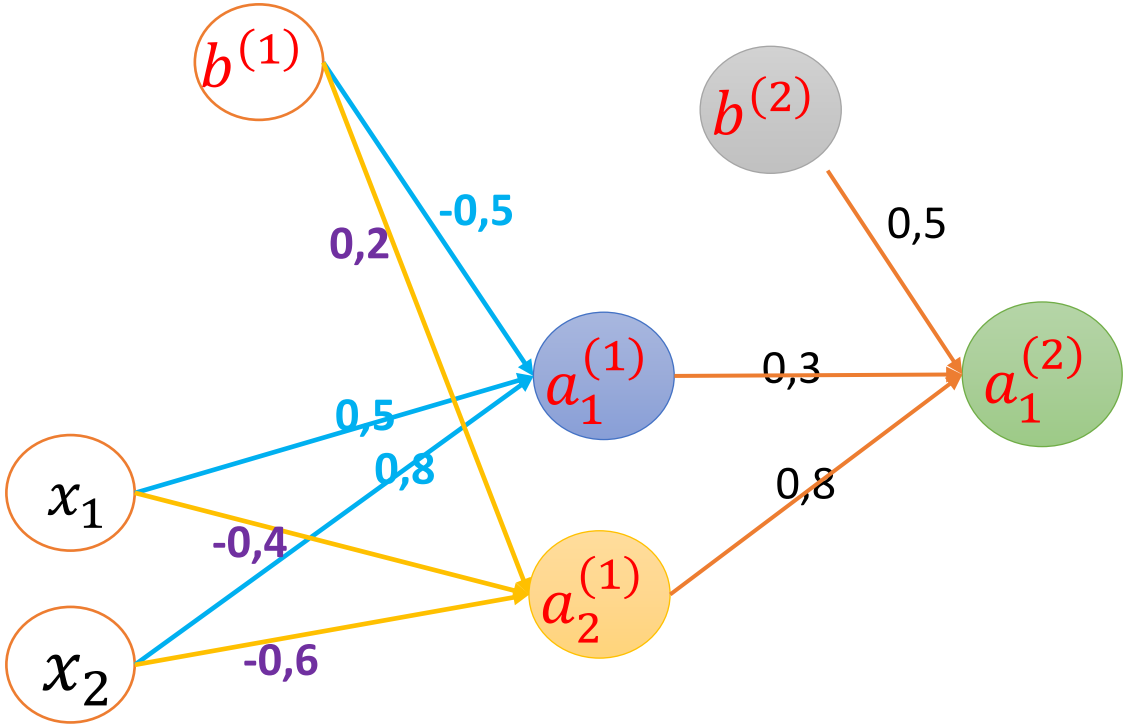
L'erreur sur les neurones des couches cachées peut être définie par :

$$\delta^{(l)} = \left(w^{(l+1)}\right)^T \delta^{(l+1)} .* \, g'\left(z^{(l)}\right)$$

Les dérivées partielles peuvent être définies par :

$$\frac{\partial J(W)}{\partial w_{ij}^{(l)}} = a_j^{(l-1)} \delta_i^{(l)} \qquad \frac{\partial J(W)}{\partial b_j^{(l)}} = \delta_i^{(l)}$$

Rétropropagation

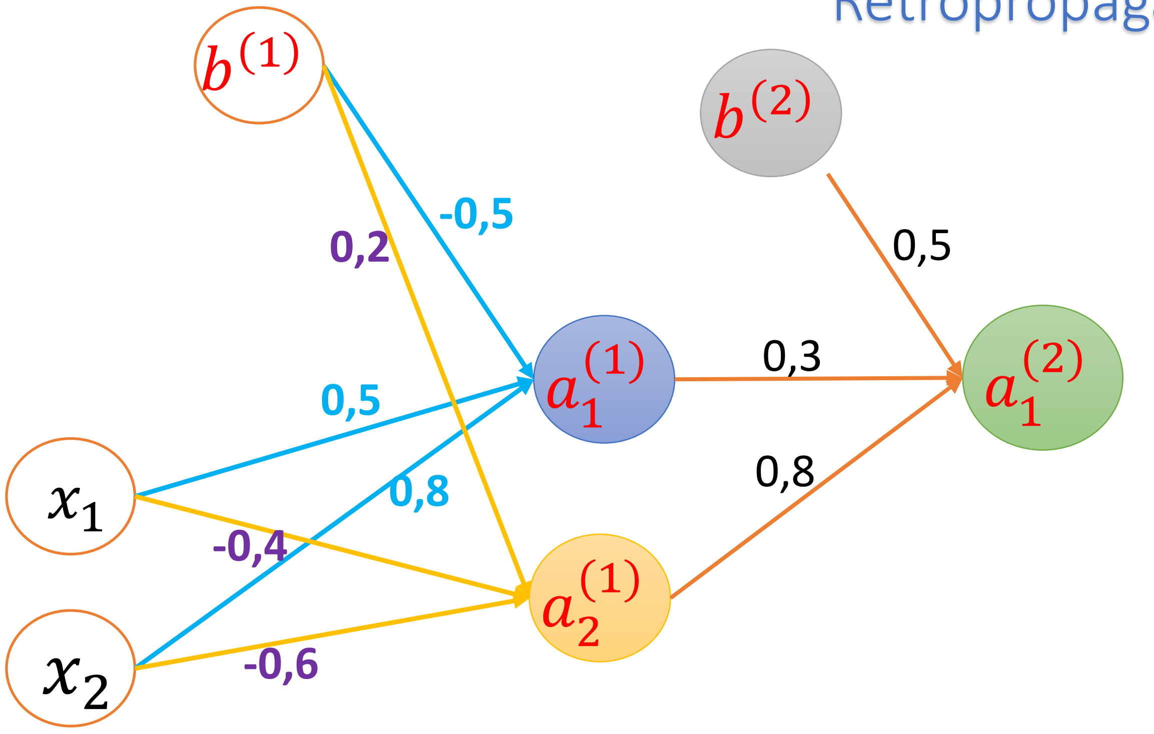


$$\delta^{(L)} = \left(a^{(L)} - y\right) .* \, g'\left(z^{(L)}\right)$$

$$\delta^{(2)} = \left(a^{(2)} - y\right) .* \, g'\left(z^{(2)}\right)$$

$$\delta_1^{(2)} = \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right)$$

Rétropropagation



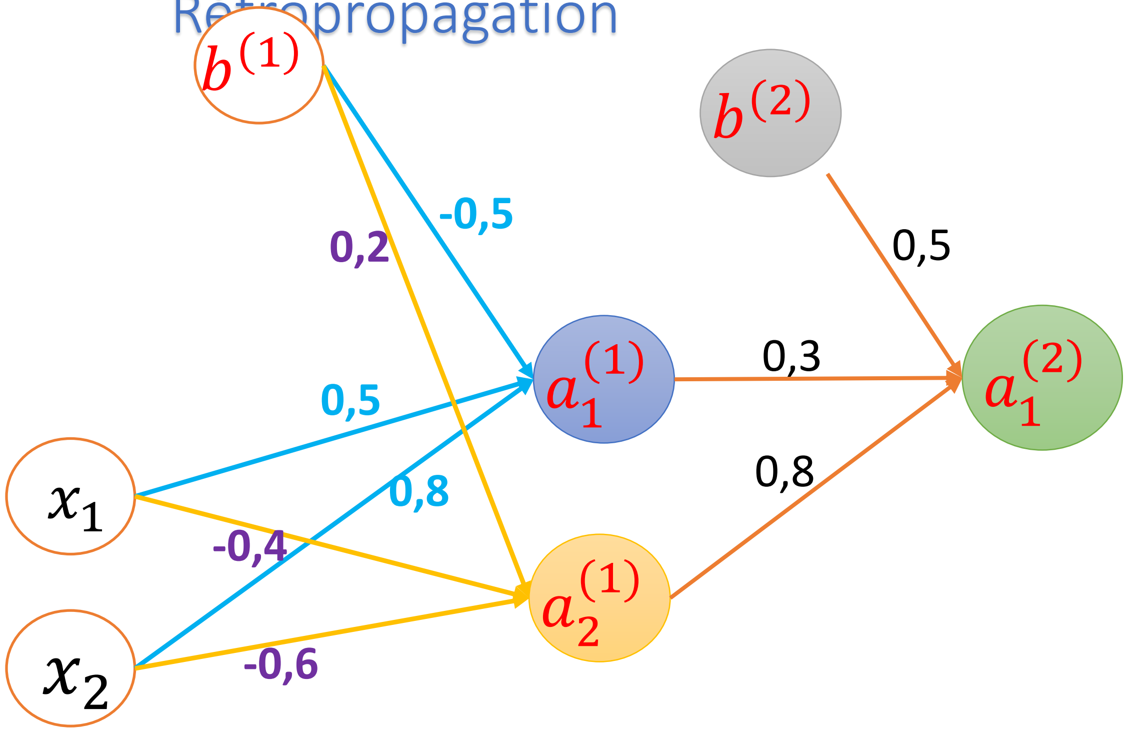
$$\frac{\partial J(W)}{\partial w_{11}^{(2)}} = \left(a_1^{(2)} - y\right) a_1^{(2)} (1 - a_1^{(2)}) a_1^{(1)}$$

$$\delta^{(2)} = \left(a^{(2)} - y\right) .* \, g'\left(z^{(2)}\right)$$

$$\delta_1^{(2)} = \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right)$$

$$\frac{\partial J(W)}{\partial w_{ij}^{(l)}} = a_j^{(l-1)} \delta_i^{(l)} \Rightarrow \frac{\partial J(W)}{\partial w_{11}^{(2)}} = a_1^{(1)} \delta_1^{(2)} = a_1^{(1)} \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right)$$

Rétropropagation



$$\delta^{(2)} = \left(a^{(2)} - y\right) .* \, g'\left(z^{(2)}\right)$$

$$\delta_1^{(2)} = \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right)$$

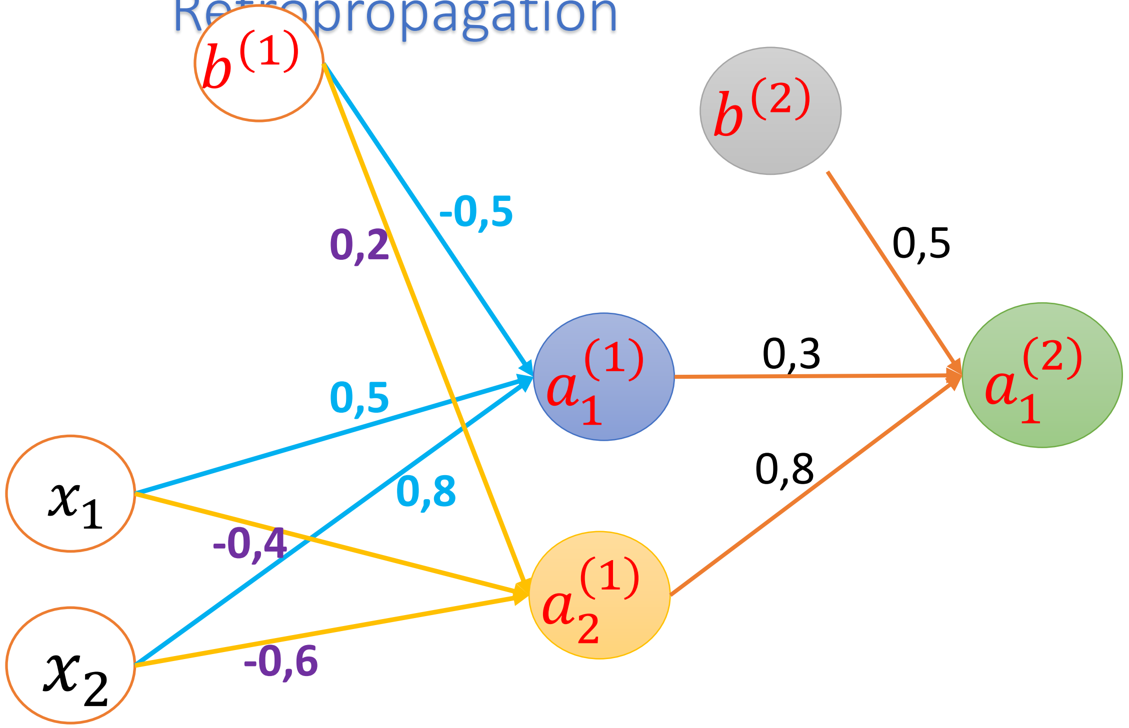
$$\delta^{(l)} = \left(w^{(l+1)}\right)^T \delta^{(l+1)} .* \, g'\left(z^{(l)}\right)$$

$$\delta^{(1)} = \left(w^{(2)}\right)^T \delta^{(2)} .* \, g'\left(z^{(1)}\right)$$

$$\delta^{(1)} = \begin{bmatrix} w_{11}^{(2)} \\ w_{12}^{(2)} \end{bmatrix} \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right) .* \begin{pmatrix} a_1^{(1)} \left(1 - a_1^{(1)}\right) \\ a_2^{(1)} \left(1 - a_2^{(1)}\right) \end{pmatrix}$$

$$\begin{bmatrix} \delta_1^{(1)} \\ \delta_2^{(1)} \end{bmatrix} = \begin{bmatrix} w_{11}^{(2)} \left(a_1^{(2)} - y\right) a_1^{(2)} (1 - a_1^{(2)}) a_1^{(1)} \left(1 - a_1^{(1)}\right) \\ w_{12}^{(2)} \left(a_1^{(2)} - y\right) a_1^{(2)} (1 - a_1^{(2)}) a_2^{(1)} \left(1 - a_2^{(1)}\right) \end{bmatrix}$$

Rétropropagation



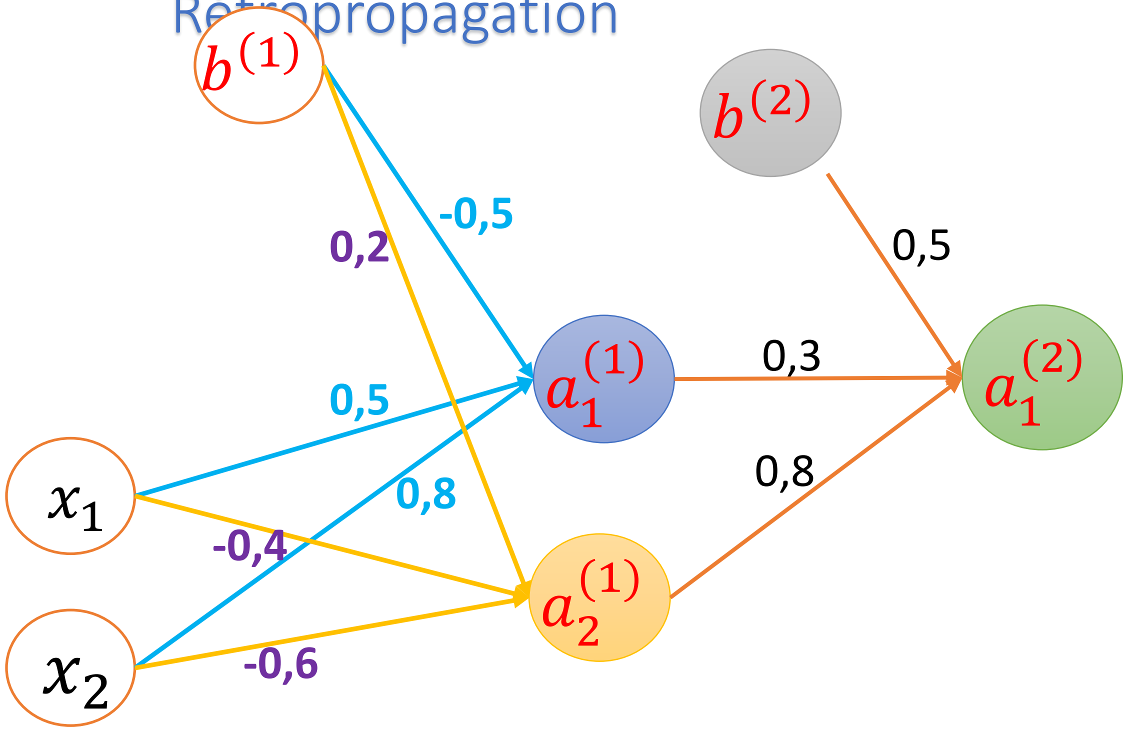
$$\begin{bmatrix} \delta_1^{(1)} \\ \delta_2^{(1)} \end{bmatrix} = \begin{bmatrix} w_{11}^{(2)} \left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)}) a_1^{(1)} \left(1 - a_1^{(1)}\right) \\ w_{12}^{(2)} \left(a_1^{(2)} - y\right) a_1^{(2)}(1 - a_1^{(2)}) a_2^{(1)} \left(1 - a_2^{(1)}\right) \end{bmatrix}$$

$$\frac{\partial J(W)}{\partial b_i^{(l)}} = \delta_i^{(l)} \quad \Rightarrow \quad \frac{\partial J(W)}{\partial b_1^{(1)}} = \delta_1^{(1)} = w_{11}^{(2)} \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right) a_1^{(1)} \left(1 - a_1^{(1)}\right)$$

$$\frac{\partial J(W)}{\partial w_{ij}^{(l)}} = \delta_i^{(l)} a_j^{(l-1)} \quad \Rightarrow \quad \frac{\partial J(W)}{\partial w_{11}^{(1)}} = \delta_1^{(1)} a_1^{(0)} = w_{11}^{(2)} \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right) a_1^{(1)} \left(1 - a_1^{(1)}\right) a_1^{(0)}$$

$$\frac{\partial J(W)}{\partial w_{12}^{(1)}} = \delta_1^{(1)} a_2^{(0)} = w_{11}^{(2)} \left(a_1^{(2)} - y\right) a_1^{(2)} \left(1 - a_1^{(2)}\right) a_1^{(1)} \left(1 - a_1^{(1)}\right) a_2^{(0)}$$

Rétropropagation



$$\begin{bmatrix}\delta_1^{(1)}\\ \delta_2^{(1)}\end{bmatrix} = \begin{bmatrix} w_{11}^{(2)}\left(a_1^{(2)}-y\right)a_1^{(2)}(1-a_1^{(2)})a_1^{(1)}\left(1-a_1^{(1)}\right) \\ w_{12}^{(2)}\left(a_1^{(2)}-y\right)a_1^{(2)}(1-a_1^{(2)})a_2^{(1)}\left(1-a_2^{(1)}\right)\end{bmatrix}$$

$$\frac{\partial J(W)}{\partial b_i^{(l)}} = \delta_i^{(l)}$$

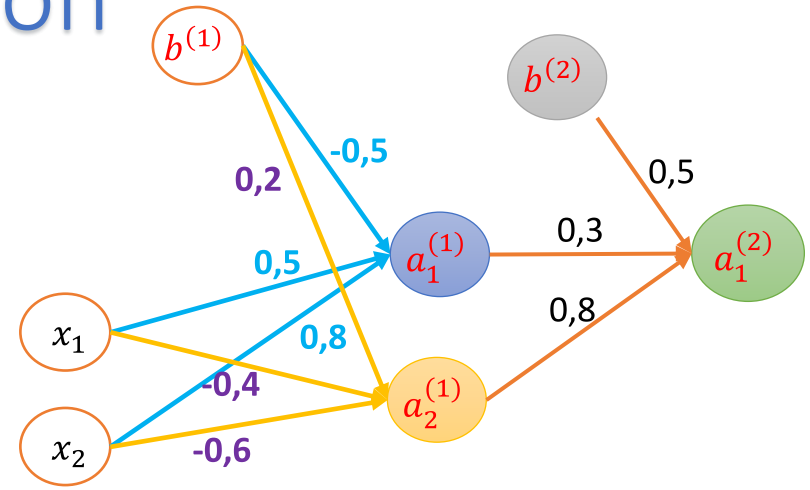
$$\frac{\partial J(W)}{\partial b_2^{(1)}} = \delta_2^{(1)} = w_{12}^{(2)}\left(a_1^{(2)}-y\right)a_1^{(2)}\left(1-a_1^{(2)}\right)a_2^{(1)}\left(1-a_2^{(1)}\right)$$

$$\frac{\partial J(W)}{\partial w_{ij}^{(l)}} = \delta_i^{(l)}a_j^{(l-1)}$$

$$\frac{\partial J(W)}{\partial w_{21}^{(1)}} = \delta_2^{(1)}a_1^{(0)} = w_{12}^{(2)}\left(a_1^{(2)}-y\right)a_1^{(2)}\left(1-a_1^{(2)}\right)a_2^{(1)}\left(1-a_2^{(1)}\right)a_1^{(0)}$$

$$\frac{\partial J(W)}{\partial w_{22}^{(1)}} = \delta_2^{(1)}a_2^{(0)} = w_{12}^{(2)}\left(a_1^{(2)}-y\right)a_1^{(2)}\left(1-a_1^{(2)}\right)a_2^{(1)}\left(1-a_2^{(1)}\right)a_2^{(0)}$$

Algorithme de rétropropagation



Pour i =1 jusqu'à m

$$\Delta_{ij}^{(l)} \leftarrow 0 \; pour\ tout\ i, j\ et\ l$$

$$a^{(0)} \leftarrow x$$

Effectuer la propagation pour calculer $a^{(1)}, a^{(2)}, \ldots . a^{(L)}$ (L est le nombre de couches -1)

En utilisant la sortie désirée $y^{(i)}$(la sortie désirée de l'exemple $x^{(i)}$) calculer $\delta^{(L)}$

Effectuer la rétropropagation pour calculer $\delta^{(L-1)}, \delta^{(L-2)}, \ldots . \delta^{(2)}$

$$\Delta_{ij}^{(l)} \leftarrow \Delta_{ij}^{(l)} + \delta_i^{(l)} a_j^{(l-1)}$$

$$D_{ij}^{(l)} \leftarrow \frac{1}{m} \Delta_{ij}^{(l)} \qquad \left(D_{ij}^{(l)} = \frac{\partial J(W)}{\partial w_{ij}^{(l)}} \right)$$

FIN