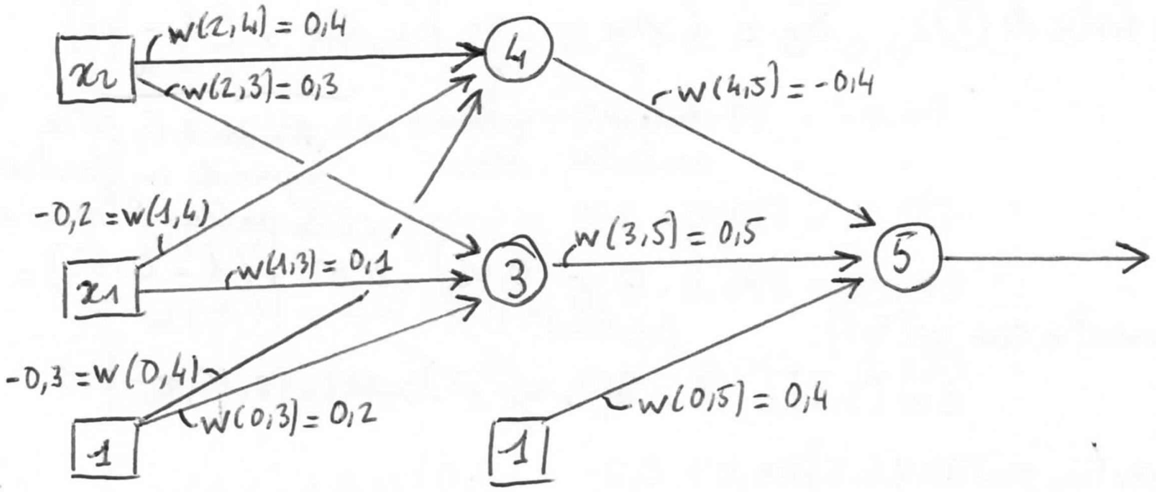
Réseau de neurones - Exemple de backpropagation

• Soit le réseau de neurones suivant :



I Prediction

• Prenons $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

↓ Alors en sortie des neurone 3 et 4

$$\sigma_3 = 1 \times 0,2 + x_1 \times 0,1 + x_2 \times 0,3 = 0,6 \Rightarrow \frac{1}{1+e^{-0,6}} \simeq 0,65 = y_3$$

$$\sigma_4 = 1 \times (-0,3) + x_1 \times (-0,2) + x_2 (0,4) = -0,1 \Rightarrow \frac{1}{1+e^{0,1}} \simeq 0,48 = y_4$$

↓ Donc en sortie du neurone 5 :

$$\sigma_5 = 1 \times (0,4) + 0,5\, y_3 - 0,4 \times y_4 = 0,4 + 0,65 \times 0,5 - 0,4 \times 0,48 = 0,53$$

$$\Rightarrow \frac{1}{1+e^{-0,53}} \simeq 0,63 = y_5$$

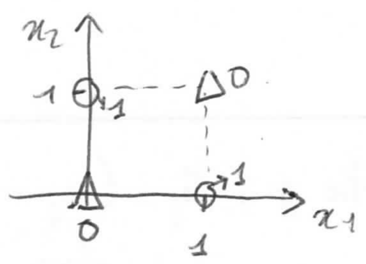
• Si on prend $\vec{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ alors : $\left[\begin{matrix} \sigma_3 = -0,5\ , & y_3 = 0 \\ \sigma_4 = -1,5\ , & y_4 = 0 \end{matrix}\right\}$ $\sigma_5 = -0,5,\ y_5 = 0$

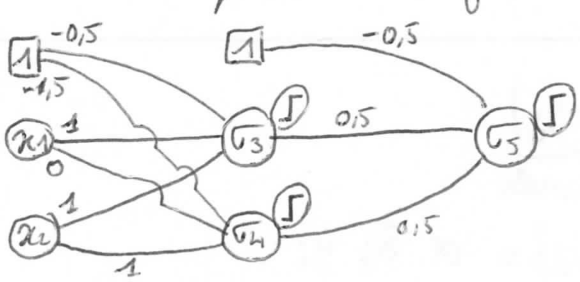
. Si on prend $\vec{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ alors : $\left[\begin{matrix} \sigma_3 = +0,5\ , & y_3 = 1 \\ \sigma_4 = -0,5\ , & y_4 = 1 \end{matrix}\right\}$ $\sigma_5 = +0,5,\ y_5 = 1$

. Si on prend $\vec{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ alors : $\left[\begin{matrix} \sigma_3 = 0,5, & y_3 = 1 \\ \sigma_4 = -0,5, & y_4 = 0 \end{matrix}\right\}$ $\sigma_5 = +0,5,\ y_5 = 1$

. On a vu que si $\vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ alors : $\left[\begin{matrix} \sigma_3 = 1,5\ , & y_3 = 1 \\ \sigma_4 = 0,5\ , & y_3 = 1 \end{matrix}\right\}$ $\sigma_5 = -0,5,\ y_5 = 0$

Donc ce réseau implémente la fonction XOR :





II Correction des poids $w(i,j)$

Imaginons que pour $\vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, on souhaite en sortie de 5 que $y_5 = 0$.

- Erreur en sortie de 5 :

$$\delta_5 = (\underset{\substack{\text{valeur}\\\text{souhaitée}}}{u_5} - \underset{\substack{\text{valeur}\\\text{obtenue}}}{y_5}) \times \underbrace{y_5 \times (1 - y_5)}_{\substack{\text{dérivée de la fction}\\\text{sigmoïde en fonction de}\\\text{la valeur obtenue en sortie de 5}}}$$

$$= (0 - 0{,}63) \times 0{,}63 \times (1 - 0{,}63) = -0{,}147$$

d'où la correction sur $w(3,5)$:

$$\Delta w(3,5) = \overset{\text{(hypothèse)}}{1} \times \delta_5 \times y_3 = -0{,}147 \times 0{,}65 \approx -0{,}1.$$

et la correction sur $w(4,5)$:

$$\Delta w(4,5) = 1 \times \delta_5 \times y_4 = -0{,}147 \times 0{,}48 \approx -0{,}07$$

puis celle sur $w(0,5)$:

$$\Delta w(0,5) = 1 \times \delta_5 \times y_0 = -0{,}147 \times 1 = -0{,}147$$

AVANT correction

- Erreur de sortie de 4 :

$$\delta_4 = \underbrace{\delta_5}_{\substack{\text{erreur issue}\\\text{de 5}}} \times \underbrace{y_4 \times (1 - y_4)}_{\text{sigmoïde}} \times \underbrace{w(4,5)}_{\substack{\text{pondération}\\\text{par le lien entre}\\\text{4 et 5}}}$$

$$= -0{,}147 \times 0{,}48 \times (1 - 0{,}48) \times (-0{,}4)$$

$$= 0{,}015$$

donc : $\Delta w(1,4) = \alpha \delta_4 y_1 = 1 \times 0{,}015 \times 1 = 0{,}015$ (car y_1 = sortie du neurone d'entrée $\boxed{x_1}$)

$\Delta w(2,4) = \alpha \delta_4 y_2 =$ " " " $= 0{,}015$ (car y_2 = sortie du neurone d'entrée $\boxed{x_2}$)

$\Delta w(0,4) = \alpha \delta_4 y_0 =$ " " " $= 0{,}015$ (car y_0 = sortie du neurone de biais $\boxed{1}$)

AVANT correction

- Erreur de sortie de 3 :

$$\delta_3 = \delta_5 \times y_3 \times (1 - y_3) \times w(3,5) = -0{,}147 \times 0{,}65\,(1 - 0{,}65) \times 0{,}5$$

$$= -0{,}0167 \approx -0{,}017.$$

donc : $\Delta w(1,3) = \alpha \delta_3 y_1 = -0{,}017 \times 1 = -0{,}017$

$\Delta w(2,3) = \alpha \delta_3 y_2 =$ " " $= -0{,}017$

$\Delta w(0,3) = \alpha \delta_3 y_0 =$ " " $= -0{,}017$.

Rq : en sortie j, $\delta_j = (\underset{\text{souhaité}}{u_j} - \underset{\text{calculé}}{y_j})\, \underbrace{y_j (1 - y_j)}_{\text{sigmoïde}}$

entre neurone i caché et sortie j : $\Delta w(i,j) = \alpha\, \delta_j\, y_i$

- On effectue seulement maintenant la correction des poids $w(i,j)$:

$$\begin{cases} w(0,5) + \Delta w(0,5) = 0,4 - 0,147 = 0,25 \\ w(3,5) + \Delta w(3,5) = 0,5 - 0,1 = 0,4 \\ w(4,5) + \Delta w(4,5) = -0,4 - 0,07 = -0,47 \end{cases}$$

$$\begin{cases} w(0,3) + \Delta w(0,3) = 0,2 - 0,017 = 0,183 \\ w(1,3) + \Delta w(1,3) = 0,1 - 0,017 = 0,083 \\ w(2,3) + \Delta w(2,3) = 0,3 - 0,017 = 0,283 \end{cases}$$

$$\begin{cases} w(0,4) + \Delta w(0,4) = -0,3 + 0,015 = -0,285 \\ w(1,4) + \Delta w(1,4) = -0,2 + 0,015 = -0,185 \\ w(2,4) + \Delta w(2,4) = 0,4 + 0,015 = 0,415 \end{cases}$$

Avec ces nouvelles valeurs, on peut calculer les prédictions suivantes avec $\vec{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$:

3 $\sigma_3 = 0,183 + 0,083 + 0,283 = 0,55 \quad \frac{1}{1+e^{-0,55}} = 0,634$

4 $\sigma_4 = -0,285 - 0,185 + 0,415 = -0,055 \quad \frac{1}{1+e^{-0,055}} = 0,51$

5 $\sigma_5 = 0,25 + 0,4 \times 0,634 - 0,47 \times 0,51 = 0,2639 \quad \frac{1}{1+e^{-0,26}} = 0,56$

Après cette itération corrective, la prédiction passe de 0,63 à 0,56. Il faudrait encore itérer pour faire tendre la prédiction vers 0.