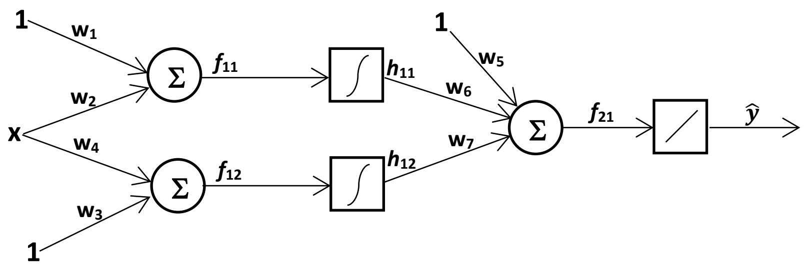
Exercices corrigés pour le chapitre réseaux de neurones

Exercice 1.

Soit le réseau de neurones multicouches décrit par le graphe suivant :



1- Donner les formules mathématiques qui déterminent les sorties intermédiaires f_{11}, f_{12}, h_{11}, h_{12}, f_{21} ainsi que la sortie finale $\hat{y}$.

2- Soit la fonction d'erreur : $\mathbf{E}(\mathbf{w}) = (\boldsymbol{y} - \hat{\boldsymbol{y}})^2$

En appliquant l'algorithme de propagation en arrière (backpropagation), trouver les expressions des mises à jour des paramètres Δw_j pour j = 1, …, 7.

Solution :

1- Propagation en avant (forward propagartion)

- $f_{11} = w_2 x + w_1$
- $f_{12} = w_4 x + w_3$
- $h_{11} = sigm(f_{11}) = \frac{1}{1+e^{-f_{11}}}$
- $h_{12} = sigm(f_{12}) = \frac{1}{1+e^{-f_{12}}}$
- $\hat{\boldsymbol{y}} = \boldsymbol{f_{21}} = \boldsymbol{w_6 h_{11}} + \boldsymbol{w_7 h_{12}} + \boldsymbol{w_5}$

2- Propagation en arrière (backpropagartion algorithm)

La fonction d'erreur est donnée par $\mathbf{E}(\mathbf{w}) = (\boldsymbol{y} - \hat{\boldsymbol{y}})^2$

Donc, on aura $\frac{\partial E(w)}{\partial w_j} = -\mathbf{2}(\boldsymbol{y} - \hat{\boldsymbol{y}}) \frac{\partial \hat{y}}{\partial w_j}$

D'après la propagation en avant, on a : $\hat{\boldsymbol{y}} = \boldsymbol{f_{21}} = \boldsymbol{w_6 h_{11}} + \boldsymbol{w_7 h_{12}} + \boldsymbol{w_5}$

Donc, les dérivées $\frac{\partial \hat{y}}{\partial w_j}$ peuvent être calculées par :

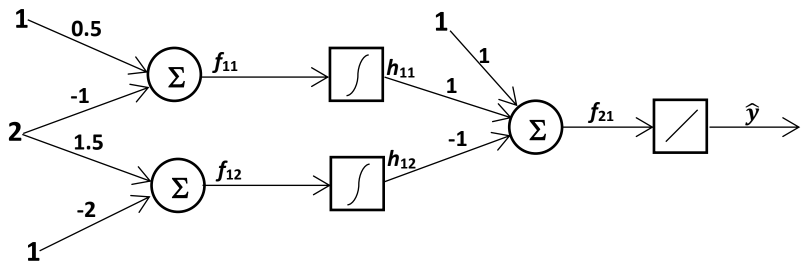
- $\frac{\partial \hat{y}}{\partial w_5} = 1$
- $\frac{\partial \hat{y}}{\partial w_6} = h_{11}$
- $\frac{\partial \hat{y}}{\partial w_7} = h_{12}$
- $\frac{\partial \hat{y}}{\partial w_1} = \frac{\partial \hat{y}}{\partial h_{11}} \frac{\partial h_{11}}{\partial f_{11}} \frac{\partial f_{11}}{\partial w_1} = w_6 h_{11}(1 - h_{11})$
- $\frac{\partial \hat{y}}{\partial w_2} = \frac{\partial \hat{y}}{\partial h_{11}} \frac{\partial h_{11}}{\partial f_{11}} \frac{\partial f_{11}}{\partial w_2} = w_6 h_{11}(1 - h_{11})x$
- $\frac{\partial \hat{y}}{\partial w_3} = \frac{\partial \hat{y}}{\partial h_{12}} \frac{\partial h_{12}}{\partial f_{12}} \frac{\partial f_{12}}{\partial w_3} = w_7 h_{12}(1 - h_{12})$
- $\frac{\partial \hat{y}}{\partial w_4} = \frac{\partial \hat{y}}{\partial h_{12}} \frac{\partial h_{12}}{\partial f_{12}} \frac{\partial f_{12}}{\partial w_4} = w_7 h_{12}(1 - h_{12})x$

En fin, la mise à jour de chaque paramétré est donnée par la formule :

$$\Delta \boldsymbol{w_j} = \boldsymbol{\alpha}(\boldsymbol{y} - \hat{\boldsymbol{y}}) \frac{\partial \hat{\boldsymbol{y}}}{\partial \boldsymbol{w_j}}$$

Exercice 2.

- Soit le réseau de neurones multicouches décrit par le graphe suivant :



- Soit la donnée **(x, y) = (2, 1)**

1- Calculer les sorties intermédiaires f_{11}, f_{12}, h_{11}, h_{12}, f_{21} ainsi que la sortie finale $\hat{y}$.

2- Calculer les paramètres Δw_j et w_j pour **j = 1, …, 7** après une itération de mise à jour (en considérant le paramètre d'apprentissage $\alpha = 0.1$).

Solution :

1- Calcul des sorties intermédiaires f_{11}, f_{12}, h_{11}, h_{12}, f_{21} :

- $f_{11} = 0.5 * 1 + -1*2 = -1.5$
- $f_{12} = 1.5 * 2 + -2*1 = 1$
- $h_{11} = sigm(f_{11}) = \frac{1}{1+e^{-f_{11}}} = \frac{1}{1+e^{1.5}}$ = 0.1824
- $h_{12} = sigm(f_{21}) = \frac{1}{1+e^{-f_{21}}} = \frac{1}{1+e^{-1}}$ = 0.7311
- f_{21} = 1*1 + 1*0.1824 + -1*0.7311 = 0.4513
- $\hat{y} = f_{21}$= 0.4513

2- Calculer les paramètres w_j pour **j = 1, …, 7.**

- Soit le paramètre d'apprentissage $\alpha = 0.1$:
- On a $w_j = w_j + \Delta w_j$ pour **j = 1, …, 7**:

2-1- Calcul des Δw_j pour **j = 1, …, 7**:

- $\Delta w_1 = \alpha(y - \hat{y})\frac{\partial \hat{y}}{\partial w_1}$ = **0.1*(1 - 0.4513)*1*0.1824*(1-0.1824) = 0.0082**
- $\Delta w_2 = \alpha(y - \hat{y})\frac{\partial \hat{y}}{\partial w_2}$ = **0.1*(1 - 0.4513)*2*0.1824*(1-0.1824) = 0.0164**
- $\Delta w_3 = \alpha(y - \hat{y})\frac{\partial \hat{y}}{\partial w_3}$ = **0.1*(1 - 0.4513)*1*0.7311*(1-0.7311) = 0.0108**
- $\Delta w_4 = \alpha(y - \hat{y})\frac{\partial \hat{y}}{\partial w_4}$ = **0.1*(1 - 0.4513)*2*0.7311*(1-0.7311) = 0.0216**
- $\Delta w_5 = \alpha(y - \hat{y})\frac{\partial \hat{y}}{\partial w_5}$ = **0.1*(1 - 0.4513)*1 = 0.0019 = 0.0549**
- $\Delta w_6 = \alpha(y - \hat{y})\frac{\partial \hat{y}}{\partial w_6}$ = **0.1*(1 - 0.4513)*0.1824 = 0.0100**
- $\Delta w_7 = \alpha(y - \hat{y})\frac{\partial \hat{y}}{\partial w_7}$ = **0.1*(1 - 0.4513)* 0.7311 = 0.0401**

2-2- Calcul des w_j pour **j = 1, …, 7**:

- $w_1 = w_1 + \Delta w_1 = 0.5 + 0.0082 = 0.5082$
- $w_2 = w_2 + \Delta w_2 = -1 + 0.0164 = -0.9836$
- $w_3 = w_3 + \Delta w_3 = -2 + 0.0108 = -1.9892$
- $w_4 = w_4 + \Delta w_4 = 1.5 + 0.0216 = 1.5216$
- $w_5 = w_5 + \Delta w_5 = 1 + 0.0549 = 1.0549$
- $w_6 = w_6 + \Delta w_6 = 1 + 0.0100 = 1.0100$
- $w_7 = w_7 + \Delta w_7 = -1 + 0.0401 = -0.9599$

Exercice 3.

Soit $\mathbf{A}$ et $\mathbf{B}$ deux variables booléennes.

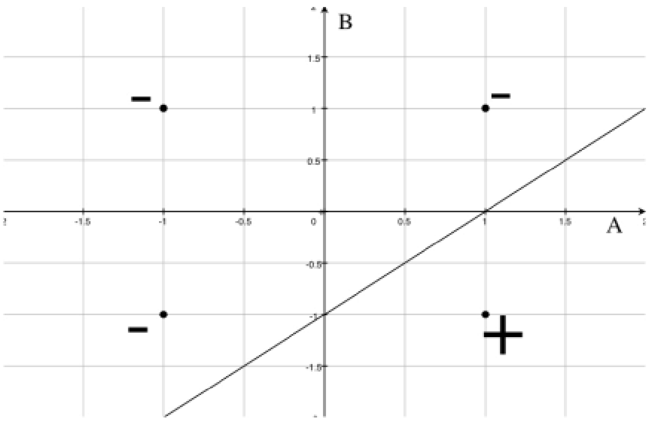
1) Concevoir un réseau de neurones à deux entrées permettant d'implémenter la fonction booléenne $\mathbf{A} \wedge \neg\mathbf{B}$.
2) Concevoir un réseau de neurones à deux couches implémentant la fonction booléenne $\mathbf{A\ XOR\ B}$.

Solution :

1) Le perceptron demandé a **3** entrées : $\mathbf{A}$, $\mathbf{B}$ et la constante $\mathbf{1}$. Les valeurs de $\mathbf{A}$ et $\mathbf{B}$ sont $\mathbf{1}$ (vrai) ou $-\mathbf{1}$ (faux). Le tableau suivant décrit la sortie $\boldsymbol{y}$ du perceptron :

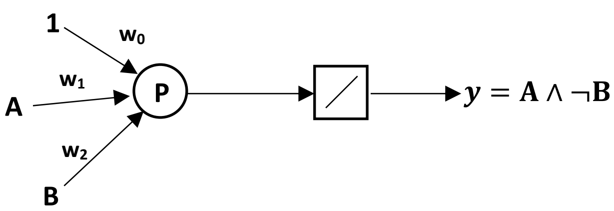
A	B	$\mathbf{y} = \mathbf{A} \wedge \neg\mathbf{B}$
-1	**-1**	**-1**
-1	**1**	**-1**
1	**-1**	**1**
1	**1**	**-1**

Une des surfaces de décision correctes (n'importe quelle droite séparant le point positif des points négatifs serait correcte) est affichée dans la figure suivante :



La droite traversant l'axe $\mathbf{A}$ en $\mathbf{1}$ et l'axe $\mathbf{B}$ en $-\mathbf{1}$. Donc l'équation de la droite est

$$\frac{\boldsymbol{A}-\mathbf{0}}{\mathbf{1}-\mathbf{0}} = \frac{\boldsymbol{B}-(-\mathbf{1})}{\mathbf{0}-(-\mathbf{1})} \Rightarrow \boldsymbol{A} = \boldsymbol{B}+\mathbf{1} \Rightarrow \mathbf{1}-\boldsymbol{A}+\boldsymbol{B} = \mathbf{0}$$

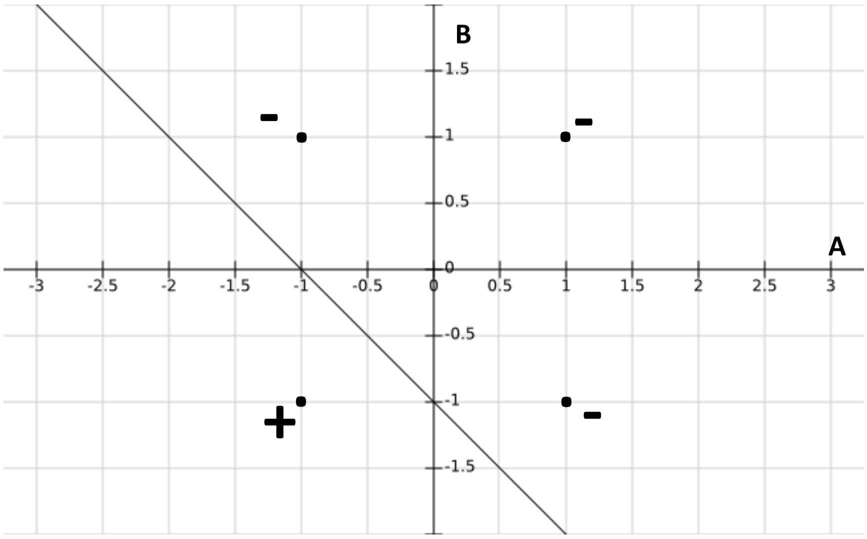


Les valeurs possibles pour les poids $\mathbf{w_0}$, $\mathbf{w_1}$ et $\mathbf{w_2}$ sont **1** et **-1**. En utilisant ces valeurs, la sortie du perceptron pour **A = 1, B = -1** est positive. Par conséquent, nous pouvons conclure que $\mathbf{w_0 = -1, w_1 = 1, w_2 = -1}$.

2) Le perceptron demandé a **3** entrées : $\mathbf{A}$, $\mathbf{B}$ et la constante $\mathbf{1}$. Les valeurs de $\mathbf{A}$ et $\mathbf{B}$ sont $\mathbf{1}$ (vrai) ou $-\mathbf{1}$ (faux). Le tableau suivant décrit la sortie $\boldsymbol{y}$ du perceptron :

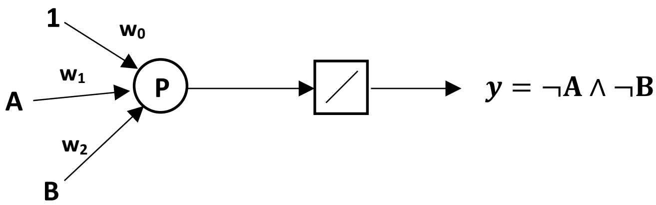
A	B	$y = \neg A \wedge \neg B$
-1	**-1**	**1**
-1	**1**	**-1**
1	**-1**	**-1**
1	**1**	**-1**

Une des surfaces de décision correctes (n’importe quelle droite séparant les points positifs des points négatifs serait correcte) est affichée dans la figure suivante :



La droite traversant l'axe $\mathbf{A}$ en $-\mathbf{1}$ et l'axe $\mathbf{B}$ en $-\mathbf{1}$. Donc l'équation de la droite est

$$\frac{\boldsymbol{A}-\mathbf{0}}{-\mathbf{1}-\mathbf{0}} = \frac{\boldsymbol{B}-(-\mathbf{1})}{\mathbf{0}-(-\mathbf{1})} \Rightarrow \boldsymbol{A} = -\boldsymbol{B}-\mathbf{1} \Rightarrow \mathbf{1}+\boldsymbol{A}+\boldsymbol{B} = \mathbf{0}$$



Les valeurs possibles pour les poids $\mathbf{w_0}$, $\mathbf{w_1}$ et $\mathbf{w_2}$ sont **1** et **-1**. En utilisant ces valeurs, la sortie du perceptron pour **A = -1, B = -1** est positive. Par conséquent, nous pouvons conclure que $\mathbf{w_0 = -0.7}$, $\mathbf{w_1 = -0.5}$, $\mathbf{w_2 = -0.5}$.

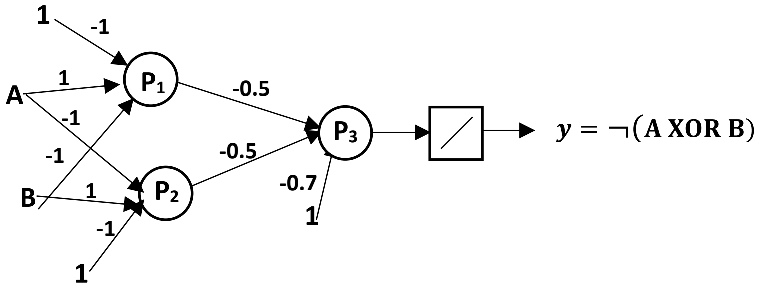
3) $\neg(\mathbf{A\ XOR\ B})$ ne peut pas être calculé par un perceptron unique, nous devons donc construire un réseau à deux couches de perceptrons. La structure du réseau peut être dérivée par :

- Exprimer $\neg(\mathbf{A\ XOR\ B})$ en fonction des composantes logiques :

$$\neg(\mathbf{A\ XOR\ B}) = \neg\big((\mathbf{A} \wedge \neg\mathbf{B}) \vee (\neg\mathbf{A} \wedge \mathbf{B})\big) = \neg(\mathbf{A} \wedge \neg\mathbf{B}) \wedge \neg(\neg\mathbf{A} \wedge \mathbf{B})$$

- Définir les perceptrons $\mathbf{P_1}$ et $\mathbf{P_2}$ pour $(\mathbf{A} \wedge \neg\mathbf{B})$ et $(\neg\mathbf{A} \wedge \mathbf{B})$
- Combiner les sorties de $\mathbf{P_1}$ et $\mathbf{P_2}$ dans $\mathbf{P_3}$ qui implémente $\neg f(\mathbf{P_1}) \wedge \neg f(\mathbf{P_2})$

En fin, le réseau demandé est donné dans la figure suivante :



Exemple :

Soit A = 1, B = -1

Résultat pour le perceptron P1 = 1 + 1 - 1 = 1

Résultat pour le perceptron P2 = -1 - 1 - 1 = -3

Résultat pour le perceptron P3 = -0.5*1 -0.5*-1 – 0.7 = -1.7 $\Rightarrow$ le résultat final est faux (y = -1).