

## Chapitre 3

# Apprentissage des réseaux de neurones

---

### 1- Introduction

La propriété qui est d'une importance primaire pour un réseau de neurones est sa capacité d'apprendre de son environnement, et d'améliorer ses performances par **l'apprentissage**.

Il y a plusieurs termes que nous devons définir :

- Apprentissage
- processus d'apprentissage
- concept de l'apprentissage
- Paradigme d'apprentissage
- Règle d'apprentissage
- Algorithme d'apprentissage

#### a. Définition de l'apprentissage

- *L'Apprentissage est un processus par lequel les paramètres libres du réseau de neurones sont adaptés.*
- *L'adaptation se fait par le moyen d'un processus de stimulation exercé par l'environnement dans lequel le réseau est plongé.*

Les paramètres libres sont les poids synaptiques

#### b. La définition du processus d'apprentissage implique la séquence d'événements suivante :

1. Le réseau de neurones est **stimulé** par un environnement.
2. Le réseau de neurones **subit des changements** dans ses paramètres libres comme résultat de cette stimulation.
3. Le réseau de neurones **répond d'une nouvelle manière** à l'environnement à cause des changements qui se sont produits dans sa structure interne.

#### c. Un réseau de neurones apprend de son environnement à travers un processus interactif d'ajustements appliqués à ses **poids synaptiques**.

La procédure pour exécuter le processus d'apprentissage est appelé **algorithme d'apprentissage**, dont la fonction est de modifier les poids synaptiques du réseau dans une façon ordonnée pour atteindre un objectif de conception désiré.

#### d. Définition d'un algorithme d'apprentissage

- Un ensemble fixé de règles bien définies pour la solution d'un problème d'apprentissage est appelé un **algorithme d'apprentissage**.
- Les algorithmes d'apprentissage diffèrent les uns des autres par la façon dans laquelle sont formulés les ajustements des poids synaptiques.

#### e. Etapes de l'apprentissage

- L'expérience laisse une empreinte.
- Dans les réseaux de neurones, cette expérience se matérialise dans les valeurs des poids synaptiques.
- L'aboutissement de l'apprentissage est un vecteur déterminé.
- C'est le vecteur final des poids synaptiques.
- Lorsque nous aboutissons à ce vecteur, l'apprentissage est terminé.

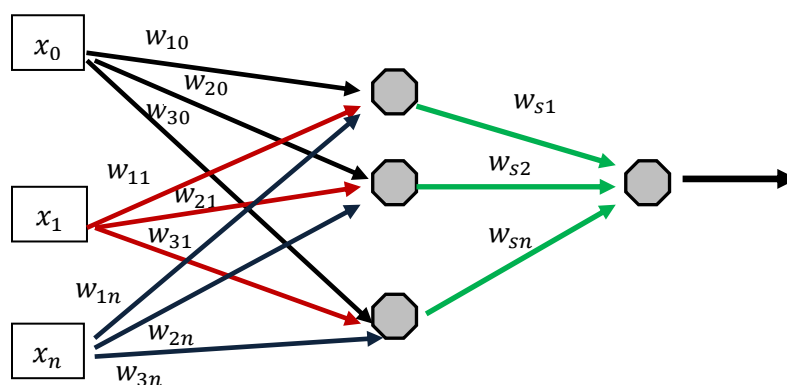


Figure 3.1 : Réseau avant entraînement

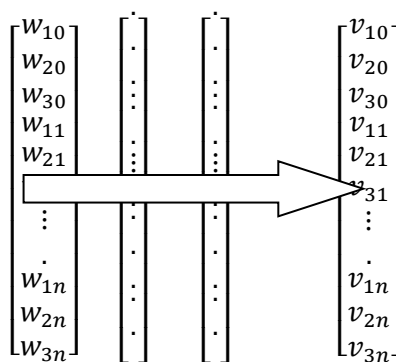


Figure 3.2 : Evolution du vecteur des paramètres

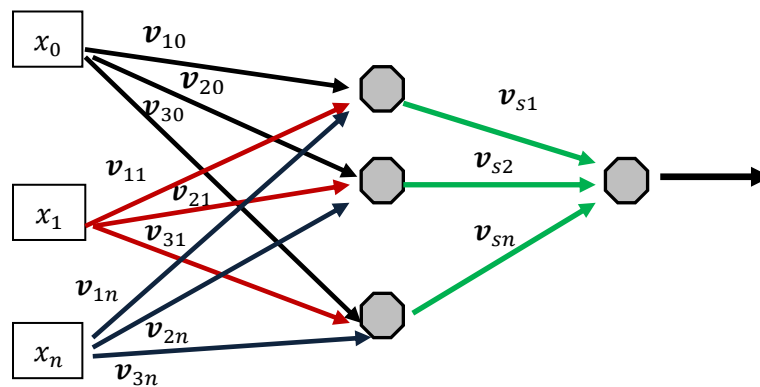
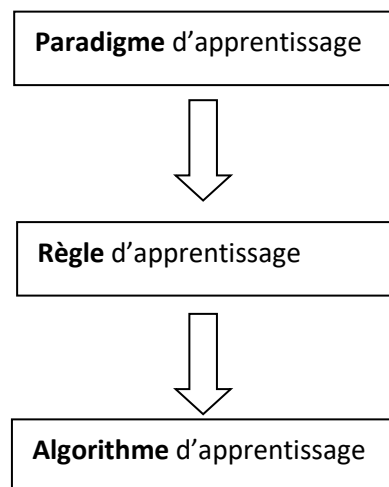


Figure 3.3 : Réseau après entraînement

#### e) Nouvelle définition de l'apprentissage

C'est la variation des valeurs du vecteur des poids synaptiques, à partir des valeurs initiales jusqu'aux valeurs finales.

#### f) Structuration du concept de l'apprentissage



L'algorithme d'apprentissage est la dernière étape de la matérialisation du concept d'apprentissage.

#### g) Paradigmes d'apprentissage (modes d'apprentissage)

- Apprentissage supervisé
- Apprentissage non supervisé
- Apprentissage par renforcement ou par essai-erreur.
- Apprentissage avec maître distale.

Il existe plusieurs règles d'apprentissage pour chaque type d'apprentissage.

## h) Règles d'apprentissage

- Apprentissage par correction d'erreur
- Apprentissage basé sur la mémoire
- Apprentissage de Hebb
- Apprentissage compétitif
- Apprentissage de Boltzmann.

## 2- Paradigmes d'apprentissage

### 2.1 Apprentissage supervisé

L'apprentissage supervisé est utilisé quand nous avons une base d'exemples qui contient en même temps les énoncés du problème et **les réponses**.

On parle d'apprentissage supervisé quand le réseau est alimenté avec les résultats **corrects** pour les exemples d'entrées donnés.

Le réseau a alors comme but d'approximer ces exemples aussi bien que possible et de développer à la fois la bonne représentation mathématique qui lui permet de généraliser ces exemples pour ensuite traiter de **nouvelles situations** (qui n'étaient pas présentes dans les exemples).

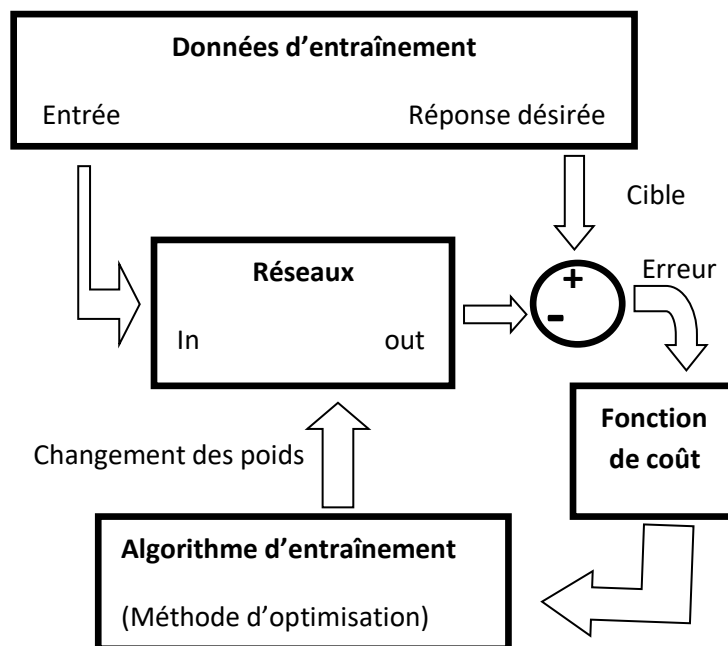


Figure 3.4 : Schéma du processus d'apprentissage par correction d'erreur

Après l'apprentissage, le réseau est testé en lui donnant seulement les valeurs d'entrée mais pas les sorties désirées, et en regardant si le résultat obtenu est proche du résultat désiré.

Ainsi, l'apprentissage est dit supervisé lorsque les exemples sont constitués de couples de valeurs du type : (valeur d'entrée, valeur de sortie désirée). Tout le problème de l'apprentissage supervisé consiste, étant donné un ensemble d'apprentissage E de N couples (entrée - sortie désirée)

$(x_i, y_i) \quad i = 1, 2, \dots, n$ , à déterminer le vecteur des poids  $w$  d'un réseau  $Fw$  capable de mettre ces informations en correspondance, c'est à dire un réseau tel que :

$$Fw(x_i) = y_i$$

avec  $i = 1, 2, \dots, n$

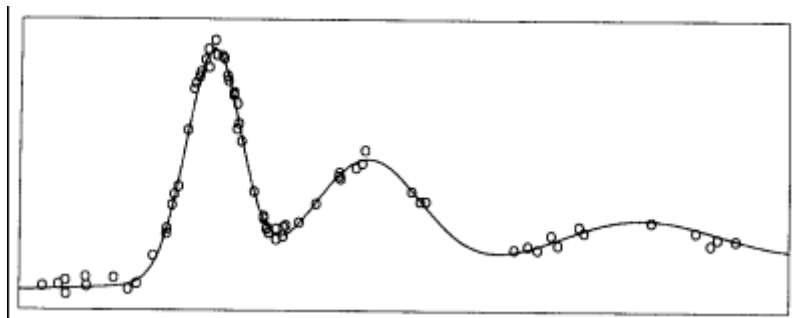


Figure 3.5 : Superposition de la courbe de réponses sur les observations d'apprentissage

**Correspondance avec les techniques traditionnelles de la statistique :**

- Régression
- Modélisation
- Séries chronologiques
- Classification

## 2.2 Apprentissage non supervisé

L'apprentissage est qualifié de non supervisé lorsque seules les valeurs d'entrée sont disponibles. Dans ce cas, les exemples présentés à l'entrée provoquent une auto-adaptation du réseau afin d'atteindre une configuration idéale par rapport aux exemples introduits.

**Correspondance avec les techniques traditionnelles de la statistique :**

- Analyses factorielles
- Classification (clustering)

## 2.3 Apprentissage par renforcement

C'est un apprentissage où le réseau donne une solution et est seulement alimenté avec une information indiquant si la réponse était correcte ou si elle était au moins meilleure que la dernière fois.

## 2.4 Apprentissage avec un maître distale

Lorsque seulement le résultat d'un ensemble d'opérations est donné au réseau.

La réponse peut ne pas être directe. Par exemple la position du bras d'un robot pour saisir un objet.

### 3- Apprentissage par correction d'erreur

Nous avons un neurone k constitué d'un seul nœud de calcul dans la couche de sortie d'un réseau de neurones à propagation avant, tel que décrit dans la figure suivante (2.1a.)

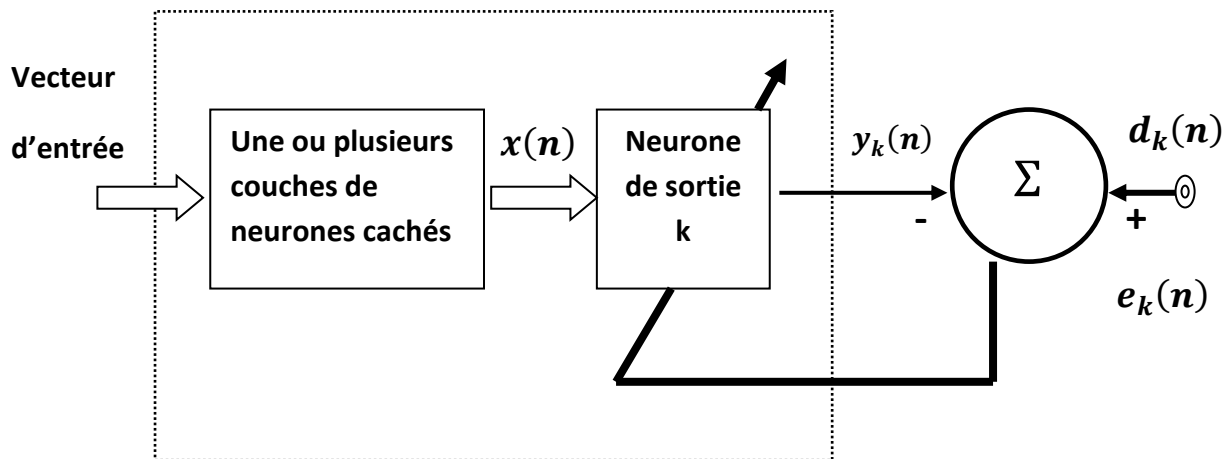


Figure 3.6 : schéma fonctionnel d'un réseau de neurones ayant un seul neurone de sortie

1. Le neurone k est dirigé par un vecteur signal  $x(n)$  produit par une ou plusieurs couches de neurones cachés.
2. Les neurones cachés sont eux-mêmes dirigés par un vecteur d'entrée appliqué aux nœuds sources (i.e., la couche d'entrée) du réseau de neurones.
3. L'argument n dénote le temps discret, ou plus précisément, l'étape de temps d'un processus itératif impliqué dans l'ajustement des poids synaptiques du neurone k.
4. Le signal de sortie du neurone k est noté par  $y_k(n)$ .
5. Ce signal de sortie, représentant la seule sortie du réseau de neurones, est comparé à la réponse désirée ou sortie cible, notée par  $d_k(n)$ .

En conséquence, un signal d'erreur, noté  $e_k(n)$ , est produit. par définition, nous avons alors

$$e_k(n) = d_k(n) - y_k(n)$$

6. Le signal d'erreur  $e_k(n)$  enclenche un mécanisme de commande, dont le but est d'appliquer une séquence d'ajustements correctifs aux poids synaptiques du neurone k.
7. les ajustements correctifs sont conçus pour faire que le signal de sortie  $y_k(n)$  se rapproche, pas à pas, de la réponse désirée  $d_k(n)$ .
8. Cet objectif est atteint en minimisant une fonction de coût ou index de performance,  $\mathcal{E}(n)$ , définie en termes de signal d'erreur  $e_k(n)$  comme :

$$\mathcal{E}(n) = \frac{1}{2} e_k^2(n)$$

9.  $\mathcal{E}(n)$  est dite la valeur instantanée de l'énergie d'erreur. Les ajustements pas –à-pas des poids synaptiques du neurone  $k$  se poursuivent jusqu'à ce que le système atteigne un état d'équilibre (i.e. les poids synaptiques se stabilisent). En ce point le processus d'apprentissage se termine.

#### 4- Règle delta

La minimisation de la fonction de coût  $\mathcal{E}(n)$  conduit à une règle d'apprentissage communément appelée **règle delta ou règle de Widrow-Hoff** :

Soit  $w_{kj}(n)$  la valeur du poids synaptique  $w_{kj}$  du neurone  $k$  excité par l'élément  $x_j(n)$  du vecteur signal  $\mathbf{x}(n)$  à l'étape de temps  $n$ .

conformément à la règle delta, l'ajustement  $\Delta w_{kj}(n)$  appliqué au poids synaptiques  $w_{kj}$  au temps  $n$  est défini par

$$\Delta w_{kj}(n) = \eta e_k(n) x_j(n)$$

Où  $\eta$  est une constante positive qui détermine le taux d'apprentissage lorsque nous passons d'une étape à une autre dans le processus d'apprentissage.

Nous appelons  $\eta$  **paramètre du taux d'apprentissage**.

La règle delta peut être définie comme suit :

***L'ajustement fait à un poids synaptique d'un neurone est proportionnel au produit du signal d'erreur et du signal d'entrée de la synapse en question.***

Ayant calculé l'ajustement synaptique  $\Delta w_{kj}(n)$ , la valeur mise à jour du poids synaptique  $w_{kj}$  est déterminée par

$$w_{kj}(n+1) = w_{kj}(n) + \Delta w_{kj}(n)$$

En effet,  $w_{kj}(n)$  et  $w_{kj}(n+1)$  peuvent être vues respectivement comme l'ancienne et la nouvelle valeur du poids synaptique  $w_{kj}$ .

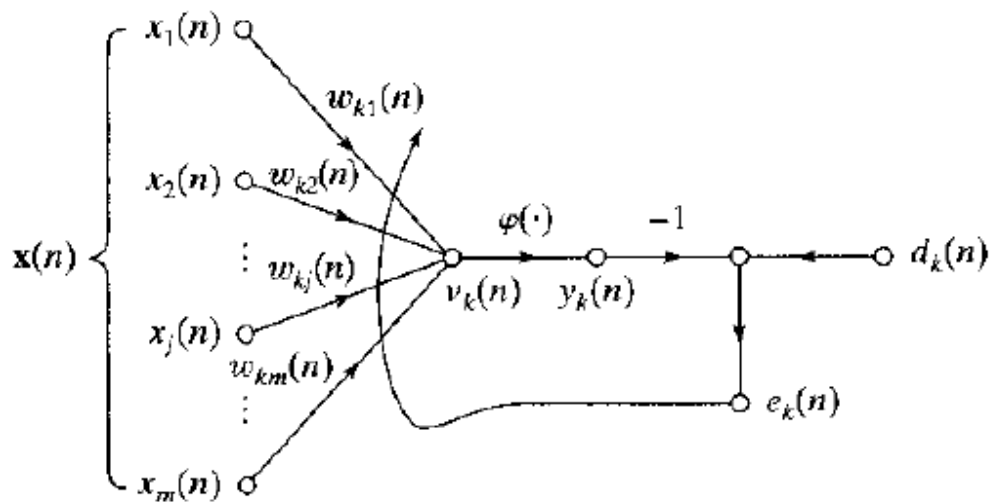


Figure 3.7 : Graphique du flux de signal du neurone de sortie

La figure 3.7 montre une représentation du graphique du flux de signal du processus d'apprentissage par correction d'erreur, focalisant sur l'activité entourant le neurone  $k$ .

Le signal d'entrée  $x_j$  et le potentiel  $v_k$  du neurone  $k$  sont respectivement appelé signaux pré-synaptique et post-synaptique de la  $j^{\text{ème}}$  synapse du neurone  $k$ .

Il est important que  $\eta$  soit choisi avec précaution pour assurer la stabilité ou la convergence du processus itératif d'apprentissage. Le choix de  $\eta$  a aussi une profonde influence sur l'exactitude et d'autre aspects du processus d'apprentissage. Le paramètre du taux d'apprentissage joue un rôle clé dans la performance en pratique de l'apprentissage par correction d'erreur.