

II.1. INTRODUCTION

Les réseaux de neurones artificiels connaissent depuis quelques années un succès croissant dans divers domaines des sciences de l'ingénieur. Ils mènent à élaborer une technique de traitement de données qui fera bientôt partie de la boîte à outils de tout ingénieur préoccupé de tirer le maximum d'informations pertinentes des données qu'il possède : faire des prévisions, élaborer des modèles, reconnaître des formes ou des signaux, etc.

L'objectif principal de la recherche sur réseaux de neurones était d'accroître nos connaissances sur le mécanisme cérébral via l'élaboration de systèmes artificiels capables de reproduire des calculs complexes, similaires à ceux qu'effectue le cerveau humain, donc l'étude approfondie des concepts de base de ces réseaux est toujours nécessaire pour qu'on puisse aboutir des solutions à des applications différentes.

II.2. HISTORIQUE

L'origine de l'inspiration des réseaux de neurones artificiels remonte à 1890 où W. James, célèbre psychologue américain, introduit le concept de mémoire associative. Il propose ce qui deviendra une loi de fonctionnement pour l'apprentissage des réseaux de neurones, connue plus tard sous le nom de loi de Hebb. Quelques années plus tard, en 1949, J. McCulloch et W. Pitts donnent leurs noms à une modélisation du neurone biologique (un neurone au comportement binaire). Ce sont les premiers à montrer que des réseaux de neurones formels simples peuvent réaliser des fonctions logiques, arithmétiques et symboliques complexes. C'est ensuite que D. Hebbel, physiologiste américain, présente en 1949 les propriétés des neurones par le conditionnement chez l'animal. Ainsi, un conditionnement de type pavlovien tel que, nourrir tous les jours à la même heure un chien, entraîne chez cet animal la sécrétion de salive à cette heure précise même en l'absence de nourriture. La loi de modification des propriétés des connexions entre neurones qu'il propose, explique en partie ce type de résultats expérimentaux.

Les premiers succès de cette discipline remontent à 1957, lorsque F. Rosenblatt développe le modèle du Perceptron. Il construit le premier neuro-ordinateur basé sur ce modèle et l'applique au domaine de la reconnaissance des formes. Notons qu'à cette époque les moyens à sa disposition étaient limités et c'était une prouesse technologique que de réussir à faire fonctionner correctement cette machine plus de quelques minutes. C'est alors qu'en 1960, l'automaticien Widrow développe le modèle Adaline (Adaptive Linear Element). Dans sa structure, le modèle ressemble au Perceptron, cependant la loi d'apprentissage est différente. Celle-ci est à l'origine de l'algorithme de rétro-propagation

de gradient très utilisé aujourd'hui. Avec les Perceptrons Multi Couches. M. Minsk et S. Papet publient ensuite en 1969 un ouvrage qui met en évidence les limitations théoriques du Perceptron. Ces limitations concernent l'impossibilité de traiter des problèmes non linéaires en utilisant ce modèle.

Quelques années d'ombre se sont en suites succédées de 1967 à 1982. Le renouveau de cette discipline reprend en 1982 grâce à J. J. Hopfield, un physicien reconnu. Au travers d'un article court, clair et bien écrit, il présente une théorie du fonctionnement et des possibilités des réseaux de neurones. Il faut remarquer la présentation anticonformiste de son article. Alors que les auteurs s'acharnent jusqu'alors à proposer une structure et une loi d'apprentissage, puis à étudier les propriétés émergentes, J. J. Hopfield fixe préalablement le comportement à atteindre par son modèle et construit, à partir de là la structure et la loi d'apprentissage correspondant au résultat escompté. Ce modèle est aujourd'hui encore très utilisé pour des problèmes d'optimisation. On peut citer encore la Machine de Boltzmann en 1983 qui était le premier modèle connu, apte à traiter de manière satisfaisante la limitation recensée dans le cas du Perceptron. Mais l'utilisation pratique s'avère difficile, la convergence de l'algorithme étant extrêmement longue (les temps de calcul sont considérables). C'est ensuite qu'en 1985 la rétro-propagation de gradient apparaît. C'est un algorithme d'apprentissage adapté au Perceptron Multi Couches. Sa découverte est réalisée par trois groupes de chercheurs indépendants. Dès cette découverte, nous avons la possibilité de réaliser une fonction non linéaire d'entrée/sortie sur un réseau, en décomposant cette fonction en une suite d'étapes linéairement séparables.

Enfin, en 1989 Moody et Darren exploitent quelques résultats de l'interpolation multi variables pour proposer le Réseau à Fonctions de base Radiales (RFR), connu sous l'appellation anglophone 'Radial Basis Function network'(RBF). Ce type de réseau se distingue des autres types de réseaux de neurones par sa représentation locale. [25]

II.3. LE MODELE NEUROPHYSIOLOGIQUE

L'élément de base du système nerveux central est le neurone. Le cerveau se compose d'environ mille milliards de neurones, avec 1000 à 10000 synapses (connexions) par neurone. Le neurone est une cellule composée d'un corps cellulaire et d'un noyau (Figure II.1). Le corps cellulaire se ramifie pour former ce que l'on nomme les dendrites. Celles-ci sont parfois si nombreuses que l'on parle alors de chevelure dendritique ou d'arborisation dendritique. C'est par les dendrites que l'information est acheminée de l'extérieur vers le soma (corps du neurone). L'information est traitée alors par le corps cellulaire. Les signaux transmis par les neurones cheminent des fibres nerveuses appelées Axones. La

transmission entre deux neurones n'est pas directe. En fait, il existe un espace intercellulaire de quelques dizaines d'Angströms entre l'axone du neurone afférent et les dendrites du neurone efférent. La jonction entre deux neurones est appelée synapse.

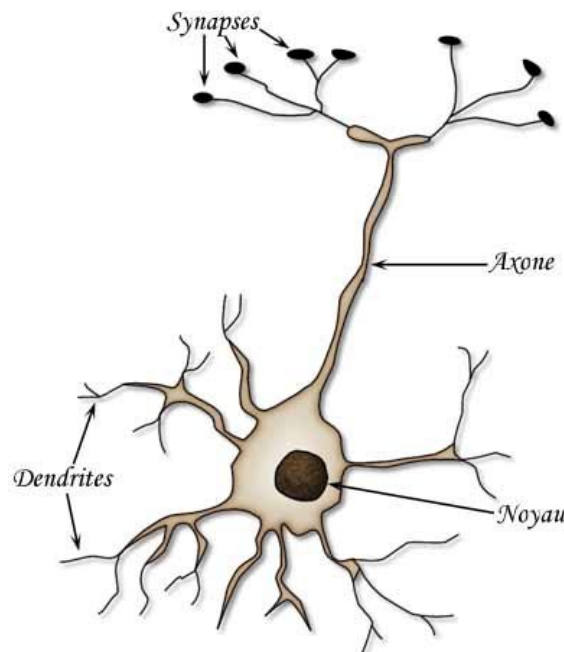


Figure II.1:le neurone biologique

Chaque neurone reçoit par ses dendrites un grand nombre de signaux excitateurs ou inhibiteurs en provenance des axones des neurones. Ces signaux pondérés sont combinés dans le corps cellulaire. Le potentiel qui en résulte est immédiatement comparé au seuil interne s'il est supérieur à celui, le neurone provoque l'émission dans l'axone d'un train d'impulsion vers un grand nombre de synapses, dans le cas contraire il reste inactif. [17][25]

II.4. LE MODEL MATHEMATIQUE(NEURAL ARTIFICIEL)

Un neurone est une fonction non linéaire, paramétrée, à valeurs bornées. Chaque neurone artificiel est un processeur élémentaire ; Il reçoit un nombre variable d'entrées en provenance de neurones amont ou à chacune de ces entrées est associée un poids w , représentatif de la force de la connexion. Chaque processeur élémentaire est doté d'une sortie unique, qui se ramifie ensuite pour alimenter un nombre variable de neurones aval. La Figure (II.2) montre la structure d'un neurone artificiel. [15][23][25]

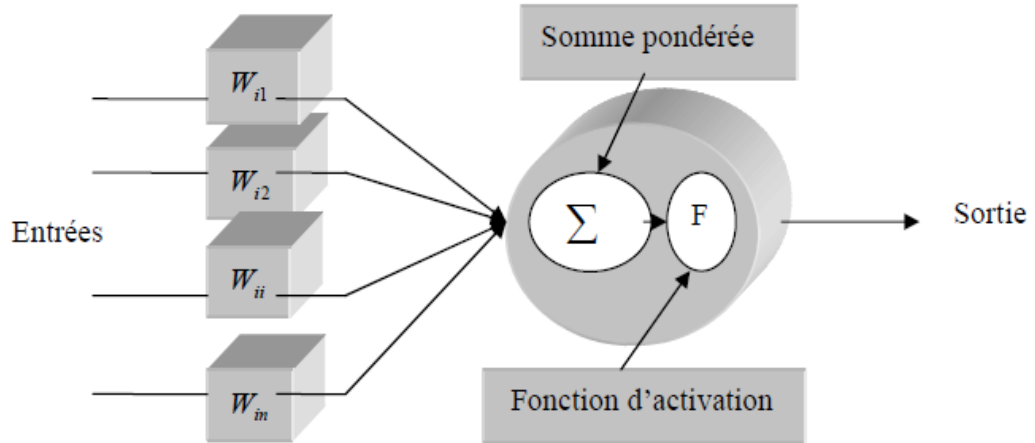


Figure II.2: La structure du neurone artificiel

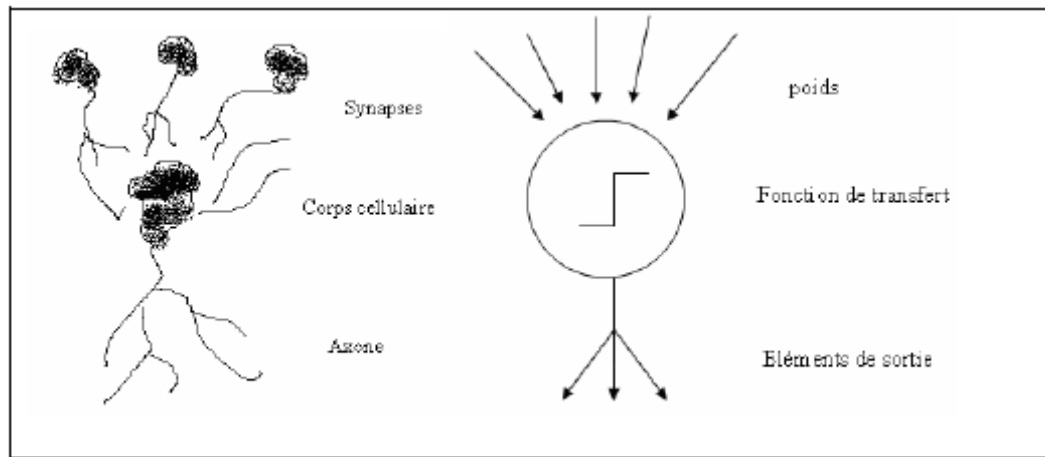


Figure II.3: Mise en correspondance du neurone biologique /et du neurone artificiel

II.4.1 Principe de fonctionnement

Un neurone formel est un élément de traitement qui fait une sommation pondérée de n entrées $x_1, \dots, x_n$. Si cette somme dépasse un certain seuil (fonction de seuillage F), le neurone est activé et transmet une réponse dont la valeur est celle de son activation. Si le neurone n'est pas activé, il ne transmet rien. [28][31]

$$Y_i = F(S_i) \quad (I.1)$$

$$S_i = \sum_{j=1}^n (W_{ij} X_j) \quad (I.2)$$

W_{ij} : coefficient synaptique ou poids associé à la $i^{\text{ème}}$ entrée du neurone j .

Parfois, il y a un terme additionnel V_j représentant le seuil interne du neurone, ce terme est considéré comme un poids $j W_0$ associé à une entrée constante, l'expression (I-1) devient donc :

$$Y_i = F\left(\sum_{j=1}^n (W_{ij} X_j) + V_j\right) \quad (I.3)$$

II.4.2 La fonction d'activation ou de seuillage

Elle établie une correspondance entre les valeurs présentées en entrée du neurone et la valeur de sortie, elle est présentée généralement par une non linéarité.

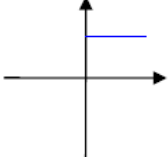
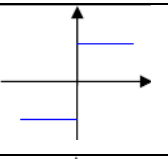
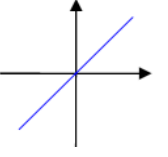
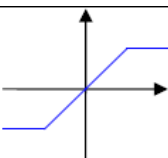
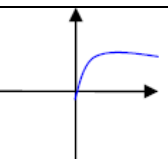
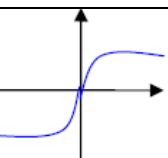
Catégorie	Type	Equation	Allure	Dérivée
Seuil	Heaviside	$f(x)=1 \text{ si } x>0$ $f(x)=0 \text{ si } x \leq 0$		/
	Signe	$f(x)=1 \text{ si } x>0$ $f(x)=-1 \text{ si } x \leq 0$		/
Linéaire	Identité	$f(x)=x$		$f'(x)=1$
	Saturation	$f(k, x) = 1$ $\text{si } x < -1/k$ $f(k, x) = 1$ $\text{si } x \geq 1/k$ $f(k, x) = 1 \text{ sinon}$		$f'(k, x) = 0 \text{ si } x < -1/k$ $f'(k, x) = 0 \text{ si } x \geq 1/k$ $f'(k, x) = k \text{ sinon}$
	Positive (logistique)	$f(x, y) = \frac{1}{1 + e^{-kx}}$		$f'(x, y) = \frac{k}{2 + e^{-kx} + e^{kx}}$
	Symétrique (type tanh)	$f(x, y) = \frac{1 - e^{-kx}}{1 + e^{-kx}}$		$f'(x, y) = \frac{k}{2 + e^{-kx} + e^{kx}}$

Tableau II.1: les fonction d'activation les plus utilisées

Toutes les fonctions d'activation utilisées doivent être différentiables, car l'architecture du réseau de neurones l'impose pour que l'apprentissage soit possible.

II.4.3 Les réseaux de neurones artificiels

Un réseau de neurones artificiel peut être défini comme un ensemble de petites unités(neurones) de calculs reliées entre elles par des liens de communication. L'information transportée par ces connexions est de type numérique ; elle peut être codée de diverses manières. Chaque unité, susceptible de posséder localement une mémoire de faible capacité, réalise un calcul à partir de données issues de ses connexions et de données locales.

Le réseau de neurone artificiel est sensé d'interagir avec les objets du monde réel de la même manière que les systèmes nerveux biologiques [15] [18] [23].

II.4.4 Mise en oeuvre des réseaux neuronaux

La procédure de développement d'un réseau neuronal peut être décrite en quatre étapes principales : [18] [29]

Etape 1 : fixer le nombre de couches cachées

Mis à part les couches d'entrée et de sortie, l'analyste doit décider le nombre de couches Intermédiaires ou cachées. Sans couche cachée, le réseau n'offre que de faibles possibilités d'adaptation ; avec une couche cachée, il est capable, avec un nombre suffisant de neurones, d'approximer toute fonction continue. Une seconde couche cachée prend en compte les discontinuités éventuelles.

Etape 2 : déterminer le nombre de neurones par couches cachées

Chaque neurone supplémentaire permet de prendre en compte des profils spécifiques des neurones d'entrée. Un nombre plus important permet donc de mieux coller aux données présentées mais diminue la capacité de généralisation du réseau. Il n'existe pas de règle générale mais des règles empiriques. La taille de la couche cachée doit être :

Soit égale à celle de la couche d'entrée.

Soit égale à 75% de celle-ci.

Soit égale à la racine carrée du produit de nombre de neurones dans la couche d'entrée et de sortie.

Etape 3 : choisir la fonction d'activation

Nous considérerons la fonction sigmoïde pour le passage de la couche d'entrée à la couche cachée. Le passage de cette dernière à la couche de sortie sera soit linéaire , soit sigmoïde (logistique) selon nos types de variables.

Etape 4 : choisir l'apprentissage

L'apprentissage par rétro-propagation nécessite la détermination du paramètre d'ajustement des poids synaptiques à chaque itération. La détermination du critère d'arrêt est aussi cruciale dans la mesure où la convergence peut passer par des minimums locaux.

II.4.5 Les propriétés des réseaux de neurones

L'intérêt porté aujourd'hui aux réseaux de neurones tient sa justification dans les quelques propriétés intéressantes qu'ils possèdent et qui devraient permettre de dépasser les Limitations de l'informatique traditionnelle, tant au niveau programmation qu'au niveau machine.[15][25][28]

II.4.5.1 Les réseaux de neurones sont des approximateurs universels

Toute fonction bornée suffisamment régulière peut être approchée uniformément avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurone cachée en nombre fini, possédant tout la même fonction d'activation et un neurone de sortie linéaire.

II.4.5.2 Sous apprentissage, généralisation et sur apprentissage

La capacité de généralisation est l'une des raisons qui motivent l'étude et le développement des réseaux de neurones artificiels. Elle peut être définie par la capacité d'élargir les connaissances acquises après apprentissage à des données nouvellement rencontrées par le réseau de neurones. C'est de cette façon que les réseaux de neurones sont capables d'approximer une fonction uniquement à partir d'une partie des données. La capacité de généralisation est très liée à la notion de sur apprentissage. Ces deux caractéristiques sont complètement antagonistes. On parle de sur apprentissage quand le réseau a trop parfaitement appris les exemples proposés. Il sera donc incapable de généraliser, comme il s'agit d'un sous apprentissage dans le cas contraire. La Figure (II.5) illustre clairement ce compromis entre sur apprentissage, sous apprentissage et bonne généralisation.

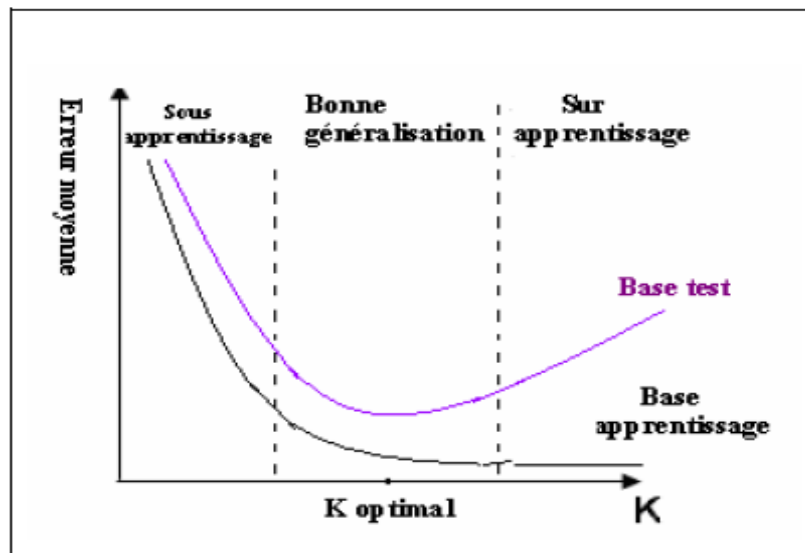


Figure II.4: propriété d'apprentissage

II.4.5.3 Le parallélisme

Cette notion se situe à la base de l'architecture des réseaux de neurones considérés comme ensemble d'entités élémentaires travaillant simultanément. Par l'étude du fonctionnement des réseaux de neurones, on pourrait aboutir à des nouvelles techniques de formalisation de problème qui permettraient de les traiter en parallèle.

II.4.5.4 La capacité d'adaptation

L'une des caractéristiques les plus complexes du fonctionnement de notre cerveau est bien la phase d'apprentissage. C'est une phase au bout de laquelle certaines modifications s'opèrent entre les connexions des neurones : certaines sont renforcées et d'autres affaiblies ou carrément inhibitrices. Le cerveau converge alors vers un comportement souhaité. Ceci nous emmène à la notion de mémoire qui donne au cerveau la capacité de retrouver des expériences passées. Celle-ci se manifeste par la capacité d'apprentissage qui permet de tenir compte des nouvelles contraintes ou de nouvelles données du monde extérieur. Certains réseaux se caractérisent aussi par leur capacité d'auto organisation qui assure leur stabilité en tant que systèmes dynamiques capables de tenir compte de situations non encore connues.

II.4.5.5 La mémoire distribuée

Dans les réseaux de neurones, la mémoire correspond à une carte d'activation de neurones. Cette carte est en quelque sorte un codage du fait mémorisé ce qui attribue à ces réseaux l'avantage de résister aux bruits car la perte d'un élément ne correspond pas à la perte d'un fait mémorisé.

II.5. LES ARCHITECTURES NEURONALES

L'intérêt des neurones réside dans les propriétés qui résultent de leur association en réseaux, c'est-à-dire de la composition des fonctions non linéaires réalisées par chacun des neurones. La plupart des réseaux de neurones ont une topologie définie sous forme de couches. Il existe quelques exceptions lorsque le réseau n'est pas explicitement défini sur plusieurs couches (comme par exemple certaines mémoires associatives), mais elles peuvent alors être considérées comme n'ayant qu'une seule couche. L'architecture du réseau peut alors être décrite par le nombre de couches, le nombre de neurones dans chaque couche et par les connexions entre les neurones. On distingue alors deux types des réseaux de neurones : les réseaux non bouclés et les réseaux bouclés.[15][17]

II.5.1 Les réseaux de neurones non bouclés

Dans les réseaux de neurones non bouclés, dit aussi statiques, non récurrents ou réseaux à couches, l'information circule des entrées vers les sorties dans un seul sens, c'est-à-dire la sortie d'un neurone ne peut pas être injectée ni directement à son entrée ni indirectement à travers d'autres neurones, donc la sortie courante n'a aucune influence sur les sorties futures (réseaux acycliques). Il existe des réseaux monocouches tels que le Perceptron et l'Adaline et des réseaux à plusieurs couches dits les MLP. [24][19]

II.5.1.1 LE PERCEPTRON

Le Perceptron est le premier modèle de réseau de neurones à été introduit par McCulloch et Pitt en 1949. Les Perceptrons sont des réseaux monocouches qui contiennent en entrée des neurones entièrement connectés à d'autres neurones organisés en sortie par une couche de connexions modifiable de poids. Les neurones de ce modèle sont binaires. Ils reçoivent des signaux excitateurs et inhibiteurs provenant des neurones en amont afin d'effectuer une sommation. Si cette somme est supérieure à un seuil, le neurone est dans un état actif et émet un signal vers les autres neurones. Si ce n'est pas le cas, il est dans un état inactif et n'émet aucun signal.[15][28]

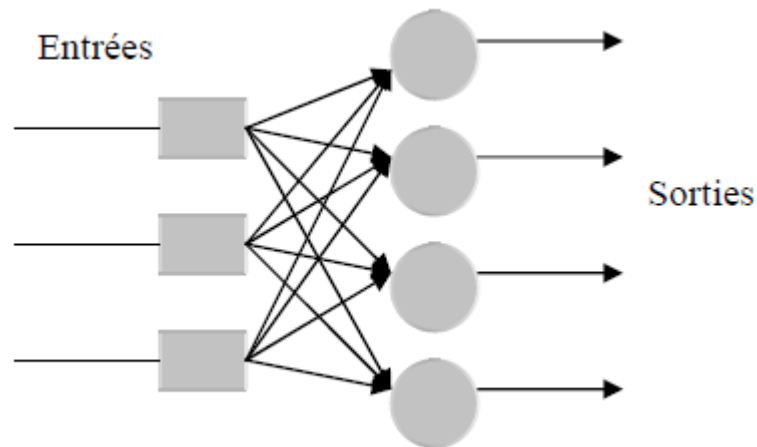


Figure II.5: Le Perceptron monocouche

II.5.1.2 L'ADALINE (ADAPtative LINear Element)

Ce modèle est composé d'un seul neurone de type McCulloch dont la fonction d'activation est linéaire. La principale différence entre les deux modèles (perceptron et Adaline) se situe au niveau de la règle d'apprentissage car l'erreur dans le premier est calculée sur la sortie non linéaire alors que dans le deuxième, elle est évaluée sur la sortie linéaire.[31]

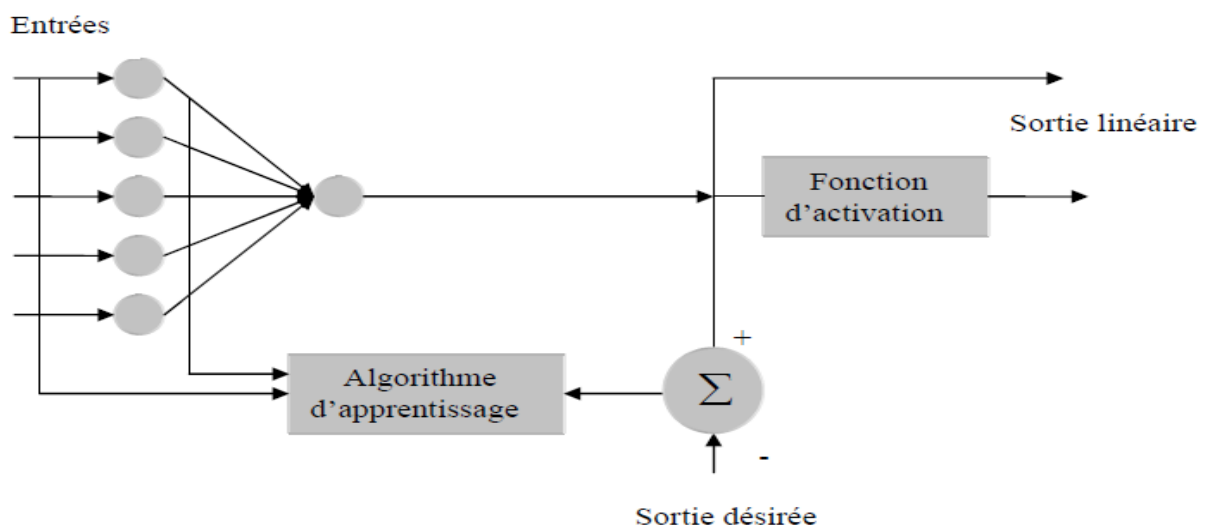


Figure II.6: Représentation de l'Adaline

II.5.1.3 LE PERCEPTRON MULTI-COUCHES (MLP)

Le Perceptron multicouches (noté MLP pour Multi Layer Perceptron en anglais) est directement inspiré du Perceptron monocouche, et cela afin de résoudre les problèmes de classification non linéaire du Perceptron et dépasser les limites principales de celui-ci. Il est organisé en couches tel que :

La couche d'entrée: Elle recevra les données source que l'on veut utiliser pour l'analyse. Sa taille est donc directement déterminée par le nombre de variables d'entrées.

Les couches cachées : Le MLP peut comporter une ou plusieurs couches cachées, en ce sens qu'elle n'a qu'une utilité intrinsèque pour le réseau de neurones et n'a pas de contact direct avec l'extérieur. Les fonctions d'activations sont en général non linéaires sur cette couche mais il n'y a pas de règle à respecter. Le choix de sa taille n'est pas implicite et doit être ajusté.

La couche de sortie : Elle donne le résultat obtenu après compilation par le réseau des données entrées dans la première couche. Sa taille est directement déterminée par le nombre de variables qu'on veut en sortie. [18][29][31]

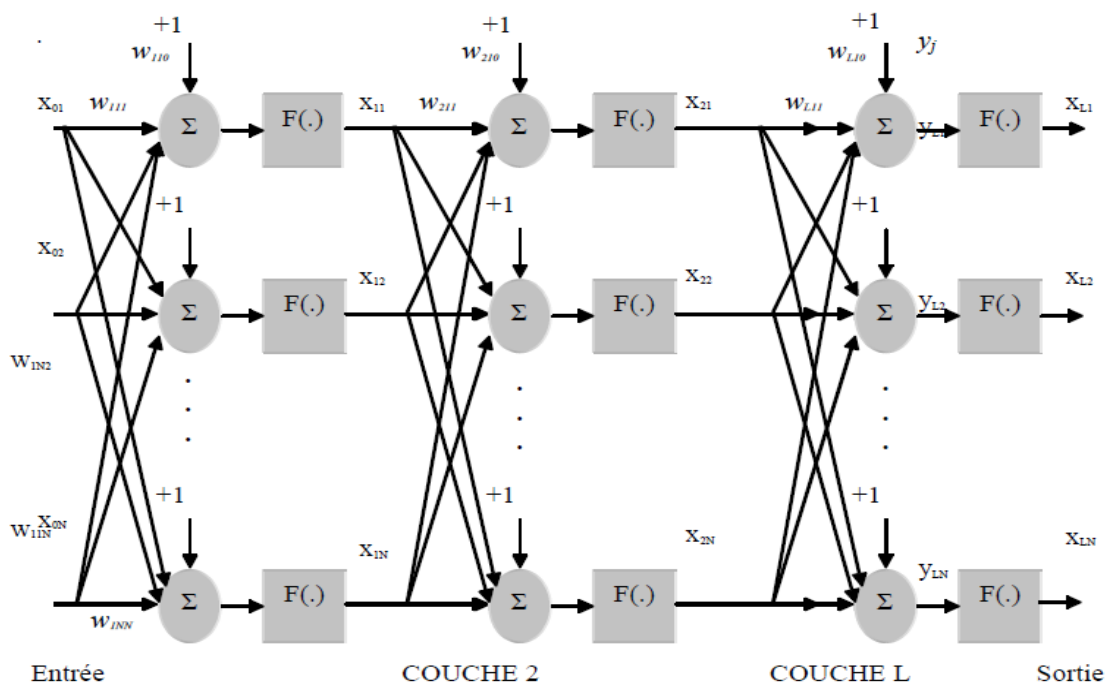


Figure II.7: Architecture générale d'un réseau de neurone MLP

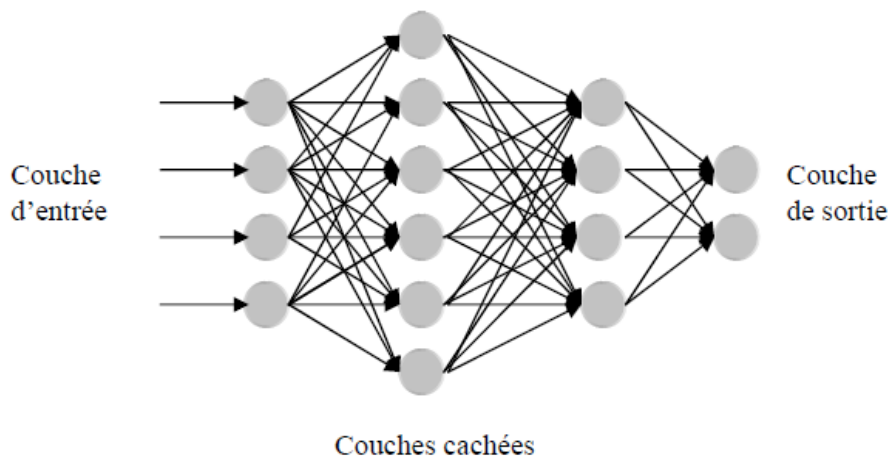


Figure II.8: Exemple d'une structure d'un MLP

II.5.1.4 LES RESEAUX RBF (FONCTIONS RADIALES DE BASE)

des Le réseau RBF (noté RBF pour Radial Basis Function) est un réseau de neurone supervisé, ou il s'agit d'une spécialisation d'un MLP. Il est constitué uniquement de trois couches (voir figure II.10) :

La couche d'entrée: Elle retransmet les inputs sans distorsion.

La couche RBF : Couche cachée qui contient des neurones RBF.

La couche de sortie : Simple couche qui contient une fonction linéaire.

Chaque neurone RBF contient une gaussienne qui est centrée sur un point de l'espace d'entrée. La fonction gaussienne permet aux neurones de ne répondre qu'à une petite région de l'espace d'entrée, région sur laquelle la gaussienne est centrée.[20]

La sortie du réseau RBF est simplement une combinaison linéaire des sorties des neurones RBF multipliées par le poids de leur connexion. Il y'a quatre paramètres principaux dans un réseau RBF :

Le nombre de neurones RBF (dans l'unique couche cachée).

La position des centres des gaussiennes de chacun neurones.

La largeur de ces gaussiennes.

Les poids des connexions entre les neurones RBF et les neurones de sortie.

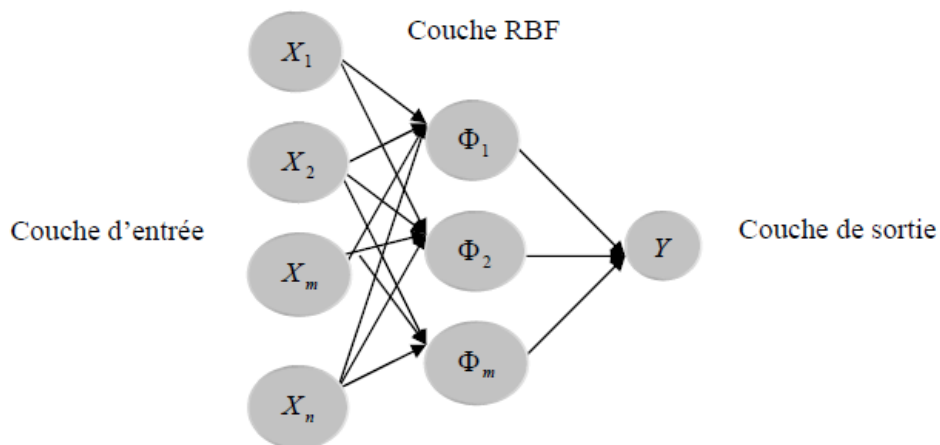


Figure II.9: Représentation d'un RBF

II.5.2 Les réseaux de neurones bouclés

Les réseaux récurrents peuvent contenir des chemins bouclés, passant plusieurs fois par un même neurone. Ils présentent donc une dynamique complexe due aux multiples rétroactions internes. Ce type de réseau a donc des capacités théoriques supérieures à celles des réseaux non récurrents. Il existe plusieurs types de réseaux récurrents tels que le

modèle de HOPFIELD. Un exemple de réseaux bouclés est donné par La Figure (II.11). [15][17]

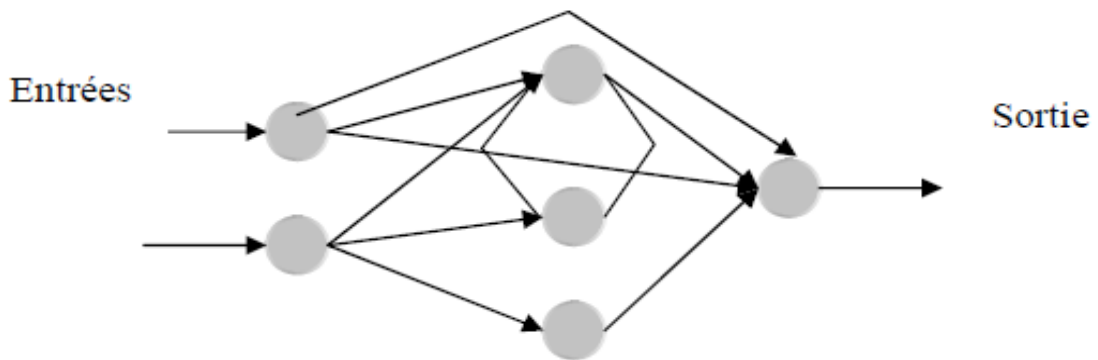


Figure II.10: Exemple d'un réseau bouclé

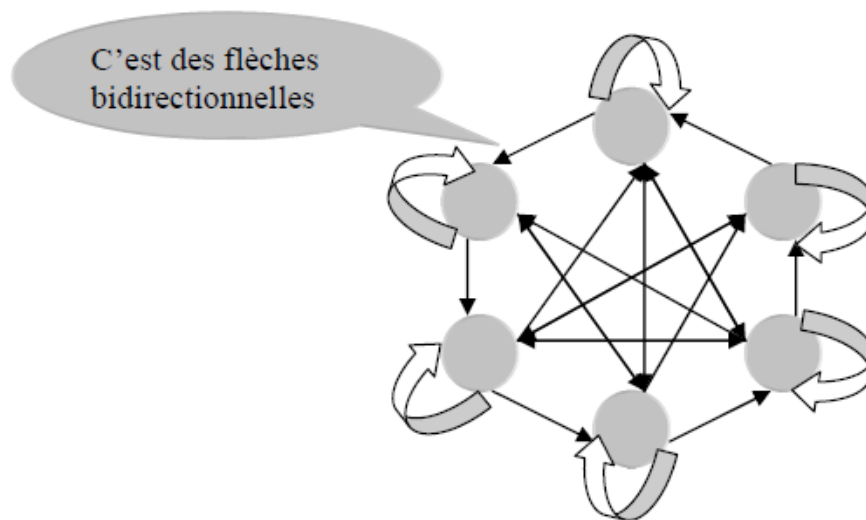


Figure II.11: Réseau entièrement bouclé

La Figure (II.11) montre un Réseau à connexion complète : c'est la structure d'interconnexion la plus générale. Chaque neurone est connecté à tous les neurones du réseau(et à lui-même).

II.6. L'APPRENTISSAGE DES RESEAUX DE NEURONES

L'une des caractéristiques les plus complexes du fonctionnement de notre cerveau est bien la phase d'apprentissage. C'est une phase au bout de laquelle certaines modifications s'opèrent entre les connexions des neurones : certaines sont renforcées et d'autres affaiblies ou carrément inhibitrices. [24][22]

On appelle apprentissage des réseaux de neurones la procédure qui consiste à estimer les paramètres des neurones du réseau, afin que celui-ci remplisse la tâche qui lui est affectée. Dans le cadre de cette définition on peut distinguer deux types d'apprentissage: apprentissage supervisé et apprentissage non supervisé. [15]

II.6.1 L'apprentissage non supervisé

Un apprentissage est non supervisé ou auto organisé quand l'adaptation des poids dépend uniquement de critères internes au réseau, et n'est pas basée sur la comparaison avec une certaine sortie désirée. Aucun signal d'erreur ou de critique ne provient de l'extérieur, seul le signal d'entrée détermine la réponse et l'adaptation du réseau. Les poids reflètent alors certaines caractéristiques statistiques des données d'entrée.[27]

II.7. DOMAINES D'APPLICATION DES RESEAUX DE NEURONES

Les propriétés d'apprentissage et d'approximation des réseaux de neurones ont permis une importante application de ces derniers dans les différents domaines pratiques, notamment, les domaines suivants : [30]

II.7.1 Identification des processus

Identifier un processus, c'est essayer de donner un modèle plus proche au processus réel qui permet de reproduire avec la précision voulue, la sortie en fonction des entrées des processus, c'est typiquement un problème d'approximation qui peut être résolu par un réseau de neurones.

II.7.2 Contrôle des processus

Contrôler un système c'est donner un ensemble d'entrées adéquats qui permettent d'avoir la sortie désirée (position, vitesse, etc...), et comme étant la fonction liant la sortie du système avec l'ensemble de commande est non linéaire, cette fonction peut être approchée avantageusement par un réseau de neurones.

II.7.3 Classification des signaux

Classer des signaux dans des catégories différents en fonction des caractéristiques de forme (amplitude, fréquence, phase, etc.), peut être réalisé par l'utilisation d'un réseau de neurones. Supposant qu'on a deux catégories de signaux A et B et qu'on souhaite classer des formes dans ces deux catégories en fonction des caractéristiques de ces formes. On peut simplement définir la fonction non linéaire F qui vaut +1 pour les formes de la classe A et -1 pour celle de la classe B, le problème de classification revient à l'approximation de la fonction non linéaire F qui peut être résolu par un réseau de neurones qui possède des entrées et une seule sortie pour fournir la décision.

II.7.4 Mémoires associatives

Dans ce type de mémoires, l'accès à une donnée se fait par l'association d'une autre donnée (mémoire adressable par contenu), cette mémoire peut être réalisée par un modèle de HOPFIELD, dans lequel chaque neurone reçoit et communique de et à tous les neurones. La sortie correspondante sera utilisée comme une donnée suivante pour le réseau et ainsi de suite. le fonctionnement de ce type de mémoires est complètement différent de celui des mémoires classiques (RAM, ROM, etc.).

II.8. CONCLUSION

A partir du comportement du cerveau humain et d'un modèle neuronal biologique simple, les chercheurs ont arrivé à construire des modèles neuronaux artificiels plus complexes. Les réseaux de neurones présentent donc une très grande diversité. En effet un type de réseau neuronal est défini par sa topologie, sa structure interne et, son algorithme d'apprentissage. On se restreindra par la suite aux réseaux non bouclés (spécialement les Mlns), car ils forment le modèle qui a le plus d'applications, et dont les techniques d'apprentissage sont les mieux connues.

Dans le prochain chapitre nous allons présenter les cartes à puces qui vont d'être utilisé dans notre travail.