

Réseaux de Neurones Artificiels

UNIVERSITÉ ABDELMALEK ESSAADI
Ecole Nationale des Sciences Appliquées d'Al Hoceima
Cycle d'ingénieure : génie civil

Assuré par : EL HASSANI FATIMA ZAHRAE
: Pr. HADDOUCH Khalid

Année universitaire : 2020/2021

DÉFINITION

RÉSEAUX DE NEURONES ARTIFICIELS

Un réseau de neurones artificiel est un ensemble de neurones formels, connectés entre eux selon une certaine **architecture** / **topologie**.

Les activités se propagent de neurone en neurone selon une certaine **dynamique**.

Les poids des connexions sont modifiés / adaptés par une règle d'**apprentissage**, à partir d'exemples.

HISTORIQUE

LA PRÉHISTOIRE

James [1890] :

mémoire associative

McCulloch & Pitts [1943]

A logical calculus of the ideas immanent in nervous activities
neurone formel

- ⇒ *Les ordinateurs à codage binaire (Von Neumann)*
- ⇒ *L 'intelligence artificielle (calcul symbolique)*
- ⇒ *Les réseaux de neurones*

Hebb [1949]

Organisation of behavior

le conditionnement est une propriété des neurones
loi d'apprentissage

HISTORIQUE

LES PREMIERS SUCCÈS

Rosenblatt [1957] :

le perceptron, premier modèle opérationnel

reconnaissance d'une configuration apprise tolérance aux bruits

Widrow [1960] :

adaline, adaptive linear element

Minsky & Papert [1969] :

impossibilité de classer des configurations non linéairement séparables.

abandon (financier) des recherches sur les RNA

HISTORIQUE

LES PREMIERS SUCCÈS

[1967 - 1982] :

Mise en sommeil des recherches sur les RNA. Elles continuent sous le couvert de domaines divers.

Grossberg, Kohonen, Anderson, ...

Hopfield [1982] :

modèle des verres de spins

Boltzmann [1983] :

première réponse à Minsky et Papert

[1985] :

la rétro-propagation du gradient et le perceptron multicouche

Rumelhart, McClelland, ... [1985] :

le groupe *Parallel Distributed Processing*

FONDEMENTS BIOLOGIQUES

STRUCTURE DES NEURONES

Le système nerveux est composé de 10^{12} neurones interconnectés.

Bien qu'il existe une grande diversité de neurones, ils fonctionnent tous sur le même schéma.

Ils se décomposent en trois régions principales :

- *Le corps cellulaire*
- *Les dendrites*
- *L 'axone*

FONDEMENTS BIOLOGIQUES

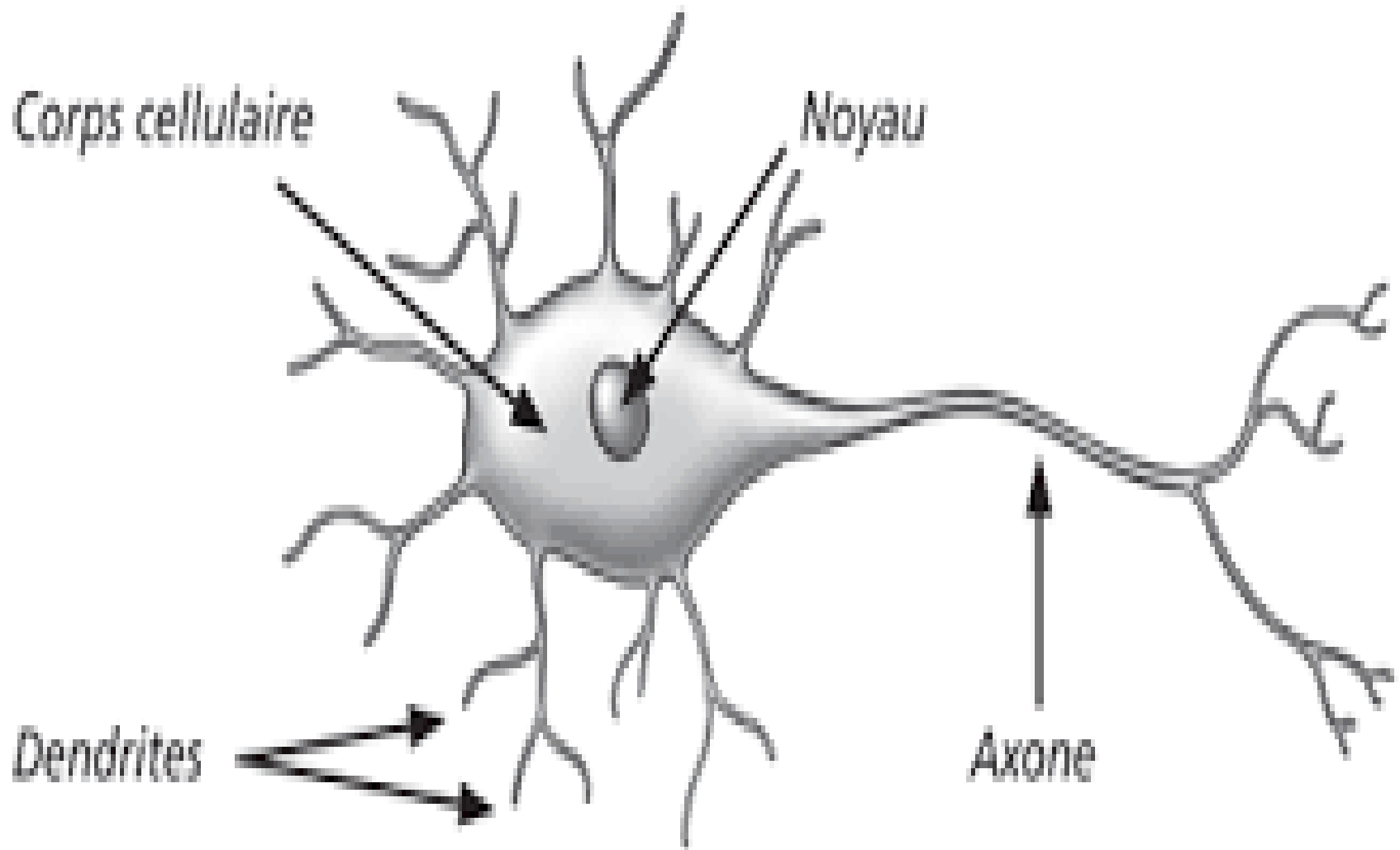
FONCTIONNEMENT DES NEURONES

L'influx nerveux est assimilable à un signal électrique se propageant comme ceci :

- Les dendrites reçoivent l'influx nerveux d'autres neurones.
- Le neurone évalue l'ensemble de la stimulation reçue.
- Si elle est suffisante, il est excité : il transmet un signal (0/1) le long de l'axone.
- L'excitation est propagée jusqu'aux autres neurones qui y sont connectés via les synapses.

FONDEMENTS BIOLOGIQUES

FONCTIONNEMENT DES NEURONES



APPRENTISSAGE

Définition

Apprentissage supervisé

Apprentissage Non supervisé

Règle d'apprentissage

Perceptron

Perceptron multicouche

DÉFINITION

L'apprentissage est une phase du développement d'un réseau de neurones durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré.

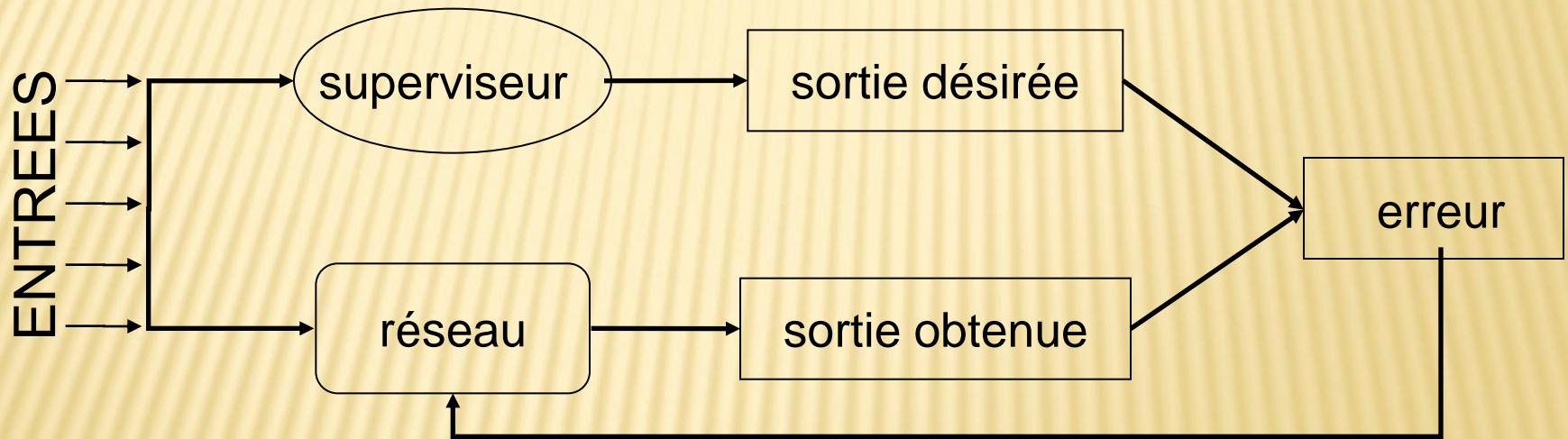
On distingue deux grandes classes d'algorithmes d'apprentissage :

- *L'apprentissage supervisé*
- *L'apprentissage non supervisé*

APPRENTISSAGE SUPERVISÉ

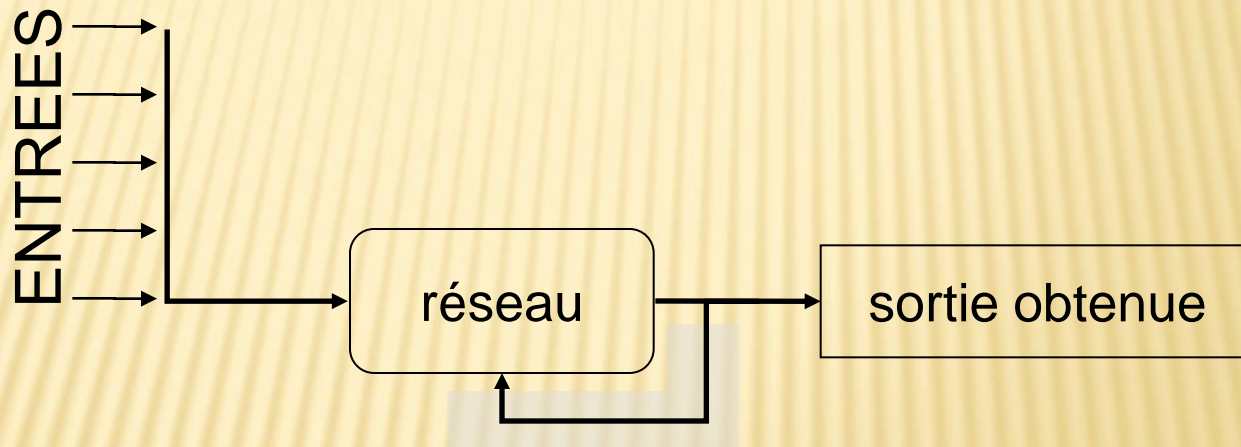
Nous utilisons un algorithme pour apprendre la fonction de mapping de l'entrée à la sortie.

$$Y = f(X)$$



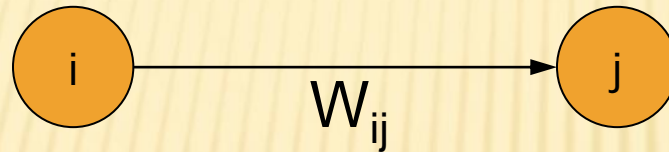
APPRENTISSAGE NON SUPERVISÉ

L'apprentissage non supervisé (Unsupervised Learning) consiste à ne disposer que de données d'entrée (X) et pas de variables de sortie correspondantes.



RÈGLES D'APPRENTISSAGE

L'apprentissage consiste à modifier le poids des connections entre les neurones.



Il existe plusieurs règles de modification :

- *Loi de Hebb* : $\Delta w_{ij} = R a_i a_j$
- *Règle de Widrow-Hoff (delta rule)* : $\Delta w_{ij} = R (d_i - a_i) a_j$
- *Règle de Grossberg* : $\Delta w_{ij} = R (a_j - w_{ij}) a_i$

RÈGLES D'APPRENTISSAGE

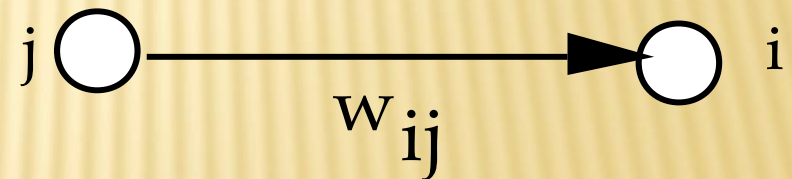
Loi de Hebb :

Si deux unités connectées sont actives simultanément, le poids de leur connexion est augmenté ou diminué. R est une constante positive qui représente la force d'apprentissage (learning rate).

$$a_i = -1 \quad a_i = 1$$

$$a_j = -1 \quad \Delta W_{ij} = R \quad \Delta W_{ij} = -R$$

$$a_j = 1 \quad \Delta W_{ij} = -R \quad \Delta W_{ij} = R$$



$$\Delta W_{ij} = R a_i a_j$$

RÈGLES D'APPRENTISSAGE

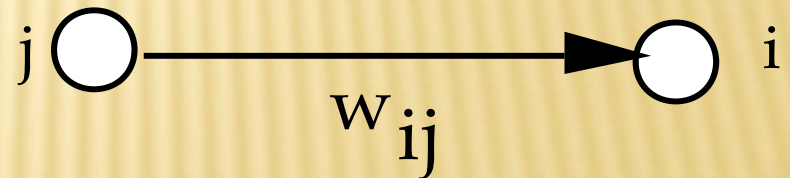
Loi de Widrow-Hoff (delta rule) :

a_i activation produite par le réseau

d_i réponse désirée par l'expert humain

Par exemple si la sortie est inférieure à la réponse désirée, il va falloir augmenter le poids de la connexion à condition bien sûr que l'unité **j** soit excitatrice (égale à 1). On est dans l'hypothèse d'unités booléennes $\{0,1\}$.

	$a_i = 0$	$a_i = 1$
$d_i = 0$	$\Delta W_{ij} = 0$	$\Delta W_{ij} = -R$
$d_i = 1$	$\Delta W_{ij} = R$	$\Delta W_{ij} = 0$

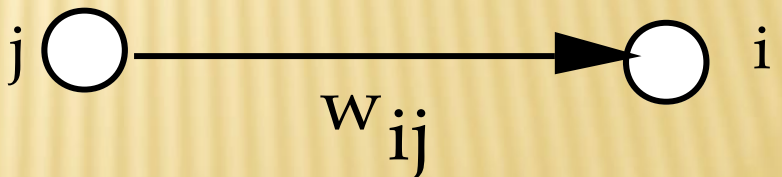


$$\Delta W_{ij} = R(d_i - a_i)a_j$$

RÈGLES D'APPRENTISSAGE

Loi de Grossberg :

On augmente les poids qui entrent sur l'unité gagnante a_i s'ils sont trop faibles, pour les rapprocher du vecteur d'entrée a_j . C'est la règle d'apprentissage utilisée dans les cartes auto-organisatrices de Kohonen

$$\Delta W_{ij} = R a_i (a_j - W_{ij})$$


The diagram shows two white circular nodes. The left node is labeled 'j' and the right node is labeled 'i'. A horizontal arrow points from node 'j' to node 'i'. Below the arrow is the label w_{ij} .

LE PERCEPTRON

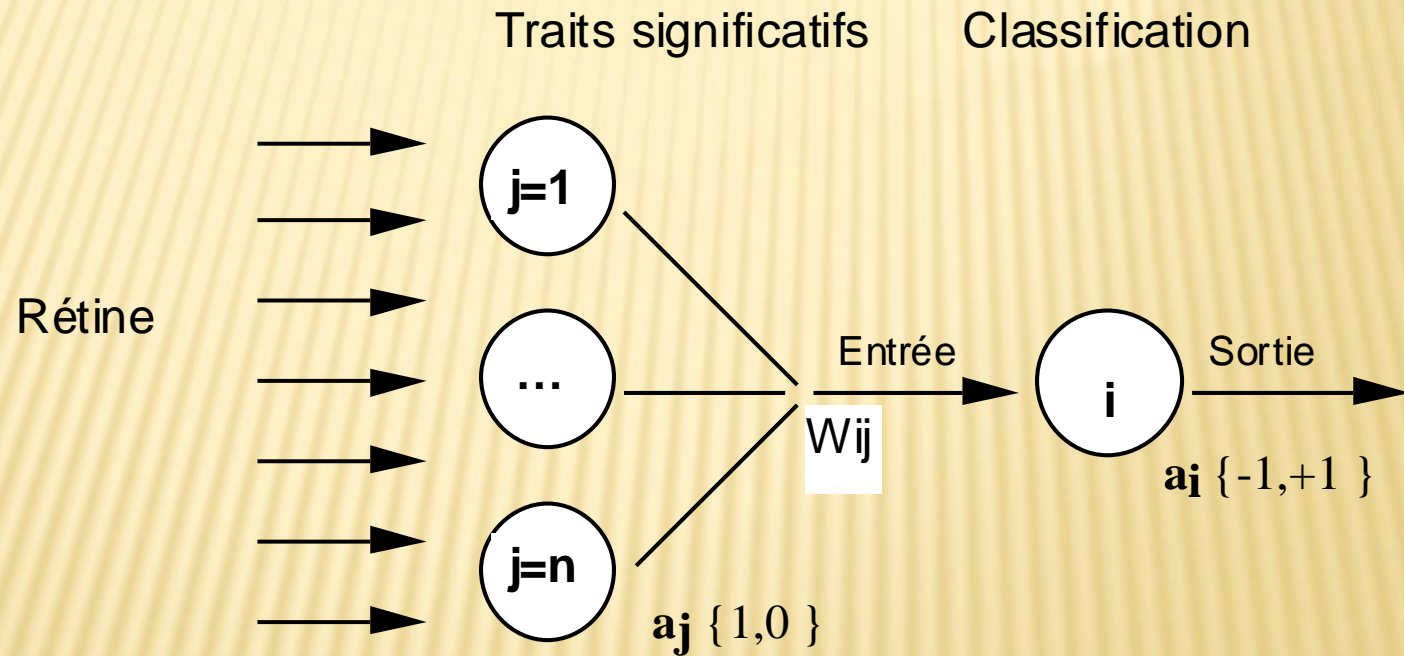
Le perceptron de Rosenblatt (1957) est le premier RNA opérationnel.

C'est un réseau à propagation avant avec seulement deux couches (entrée et sortie) entièrement interconnectées.

Il est composé de neurones à seuil.

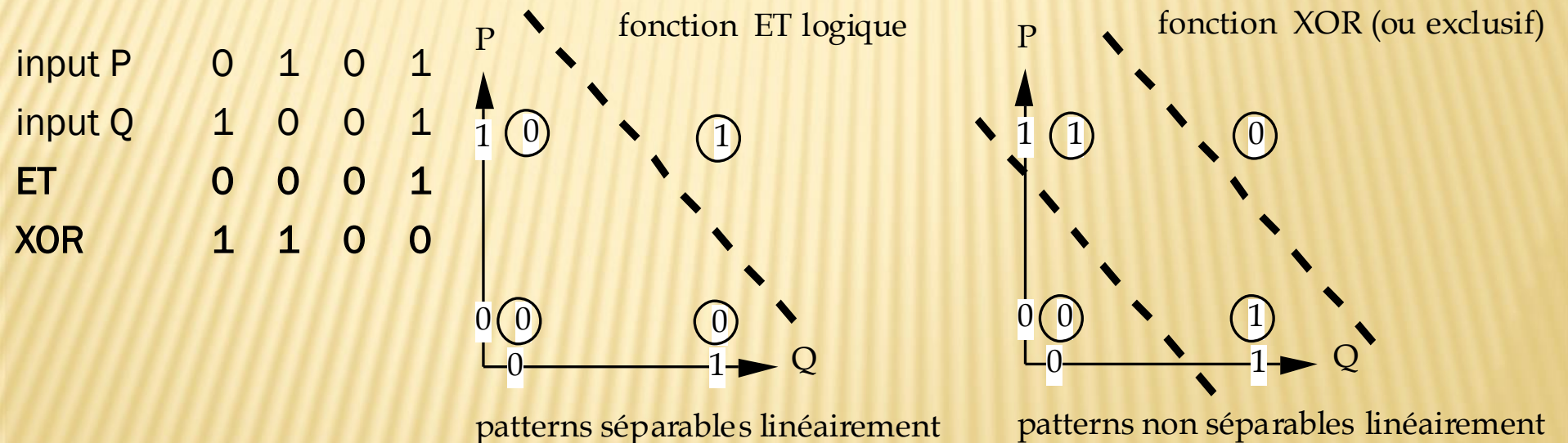
L'apprentissage est supervisé et les poids sont modifiés selon la règle delta.

LE PERCEPTRON



LIMITE DE LE PERCEPTRON

Le perceptron est incapable de distinguer les patterns non séparables linéairement [Minsky 69]



LE PERCEPTRON MULTICOUCHE

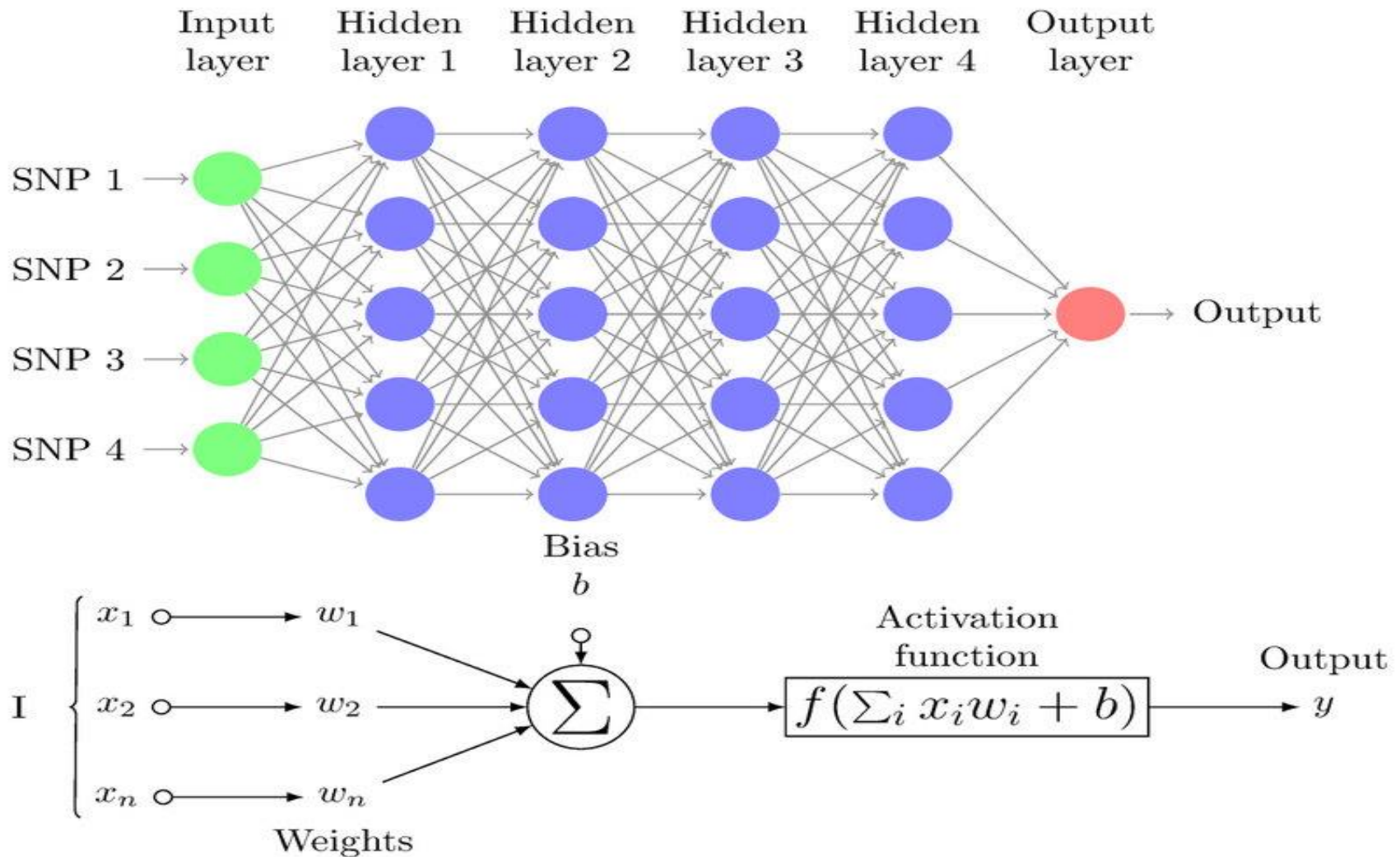
DÉFINITION

Le **perceptron multicouche** (multilayer **perceptron** MLP) est un type de réseau neuronal artificiel organisé en plusieurs couches au sein desquelles une information circule de la couche d'entrée vers la couche de sortie uniquement ; il s'agit donc d'un réseau à propagation directe (feedforward).

Chaque couche est constituée d'un nombre variable de neurones, les neurones de la dernière couche (dite « de sortie ») étant les sorties du système global.

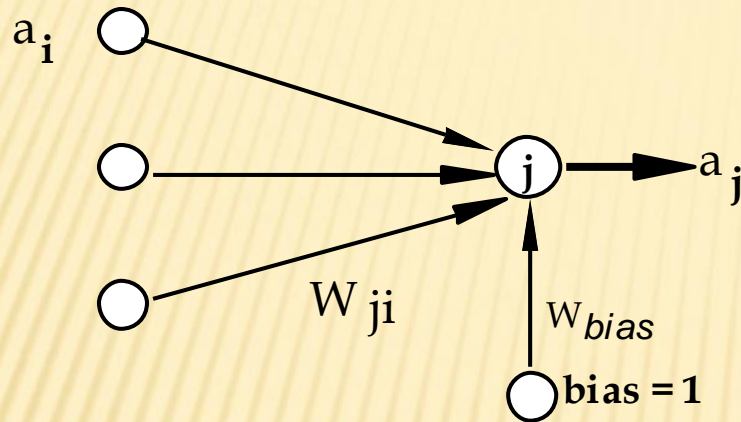
LE PERCEPTRON MULTICOUCHE

ARCHITECTURE



LE PERCEPTRON MULTICOUCHE

ACTIVATION



$$x_j = \sum w_{ji} a_i$$

$$a_j = f(x_j)$$

fonction sigmoïde

$$a = f(x) = \frac{1}{1 + e^{-x}}$$

$$f'(x) = f(x) \cdot (1 - f(x))$$

fonction tangente hyperbolique

$$a = f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$f'(x) = (1 + f(x)) \cdot (1 - f(x))$$

LE PERCEPTRON MULTICOUCHE

APPRENTISSAGE : RETROPROPAGATION DE L'ERREUR

❖ Ajustement des poids

En statistiques, la rétro-propagation du gradient est une méthode pour calculer le gradient de l'erreur pour chaque neurone d'un réseau, de la dernière couche vers la première. Cette technique consiste à corriger les erreurs pour rapprocher la sortie désirée de la sortie calculée.

- *si la sortie du réseau est égale à la sortie désirée, le poids de la connexion entre ce neurone et le neurone d'association qui lui est connecté n'est pas modifié.*
- *Dans le cas contraire les MLP utilisent pour modifier leurs poids par un algorithme d'apprentissage **Rétro-propagation**.*

LE PERCEPTRON MULTICOUCHE

APPRENTISSAGE : RETROPROPAGATION DE L'ERREUR

$$S_j = \sum_i a_i W_{ij}$$

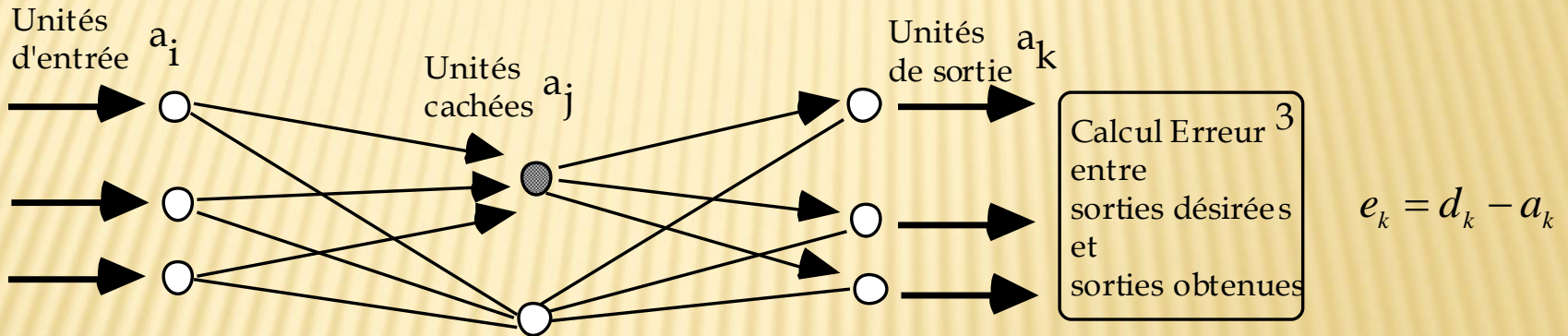
$$a_j = f(S_j)$$

1 Calcul activations
unités cachées

$$S_k = \sum_j a_j W_{jk}$$

$$a_k = f(S_k)$$

2 Calcul activations
unités de sortie



$$\delta_j = \left(\sum_k W_{jk} \delta_k \right) \cdot f'(S_j)$$

5 Calcul de l'erreur
sur les unités cachées

4 Calcul de l'erreur
sur les unités de sortie

$$\delta_k = e_k \cdot f'(S_k)$$

6 **Ajustement des poids**

Apprentissage des unités cachées	Apprentissage des unités de sortie
-------------------------------------	---------------------------------------

$$\Delta W_{ij} = \varepsilon \delta_j a_i$$

$$\Delta W_{jk} = \varepsilon \delta_k a_j$$

LE PERCEPTRON MULTICOUCHE

APPRENTISSAGE : RETROPROPAGATION DE L'ERREUR

Initialiser les poids à de petites valeurs aléatoires (habituellement entre -0,05 et 0,05)

Tand que la condition d'arrêt n'est pas atteinte

- pour chaque exemple d'entraînement :

- Calculer les sorties du réseau.
- Pour toutes les unités de sortie k , calculer l'erreur :

$$\Delta_k \leftarrow f'(S_k) \times (d_k - a_k)$$

- Pour toutes les unités cachées j , calculer l'erreur:

$$\Delta_j \leftarrow f'(S_j) \times \sum W_{jk} \Delta_k$$

- Mettre à jour tous les poids W_{jk}

$$W_{jk}(t + 1) \leftarrow W_{jk}(t) + \alpha \times a_j \times \Delta_k$$

LE PERCEPTRON MULTICOUCHE

CONDITION D'ARRÊT

- ❑ **Le nombre d'itération est important car:**
 - *Si trop faible, l'erreur n'est pas suffisamment réduite*
 - *Si trop grand, le réseau devient trop spécifique aux données d'entraînement.*

- ❑ **Il y'a plusieurs conditions d'arrêts possible**
 - *Après un certain nombre fixe d'itérations.*
 - *Lorsque les poids se stabilisent*
 - *Lorsque l'erreur dans les sorties des exemples d'entraînement descend en dessous d'une certaine borne.*

LE PERCEPTRON MULTICOUCHE

APPLICATIONS

- ❑ L'interprétation d'image .
- ❑ La reconnaissance vocal.
- ❑ La reconnaissance de mots écrits à la main .
- ❑ L'apprentissage de stratégies de contrôle pour les robots.
- ❑ Une des meilleures méthodes connues pour l'interprétation de données provenant de capteurs dans le monde réel.

LE PERCEPTRON MULTICOUCHE

APPLICATIONS

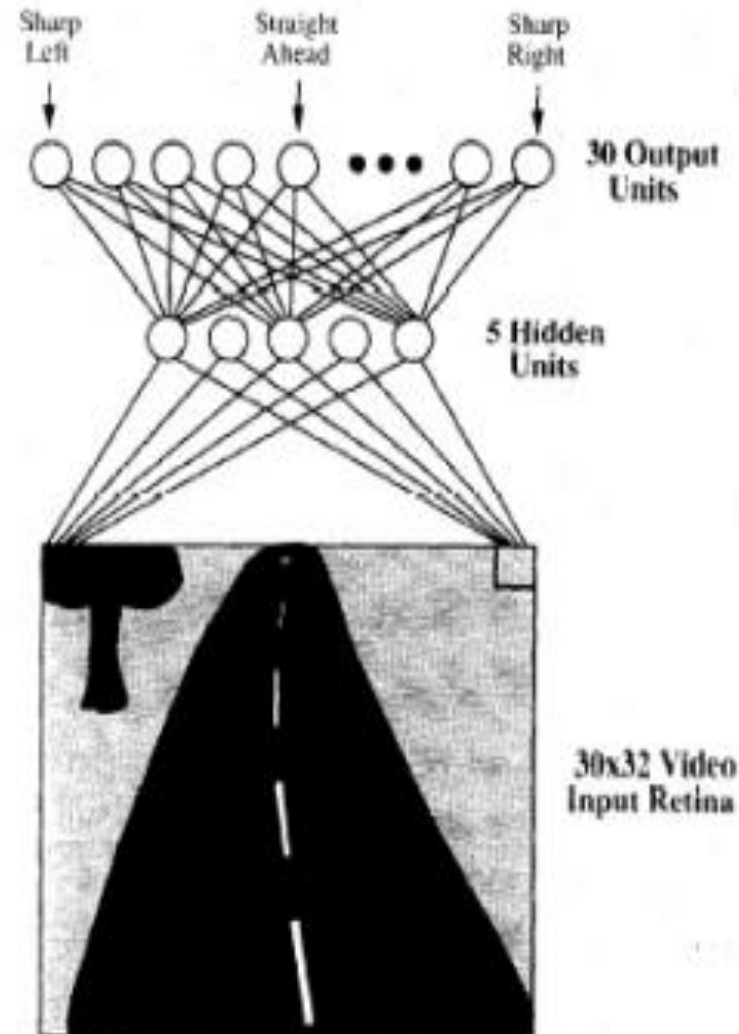
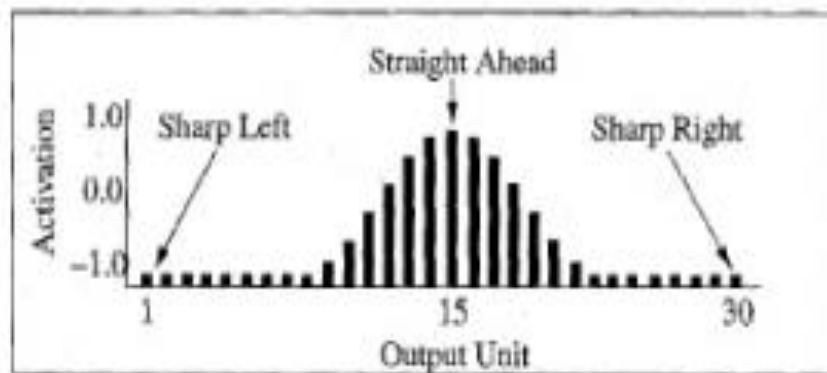
- Le système ALVINN a conduit une voiture à 55 milles/h pendant 90 miles sur une autoroute.



LE PERCEPTRON MULTICOUCHE

APPLICATIONS

- L'architecture d'ALVINN est un réseau de neurones à rétropropagation avec une couche cachée.
- La sortie est une distribution de Gauss centrée autour de la direction appropriée.



LE PERCEPTRON MULTICOUCHE

APPLICATIONS

Reconnaissance de chiffres

