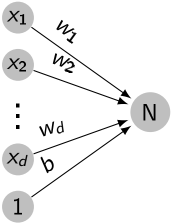
# Réseaux de neurones : expressivité et espaces implicites

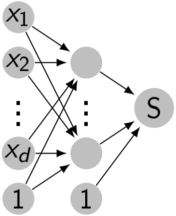
# Plan

$$g(\overrightarrow{w} \cdot \overrightarrow{x} + b)$$



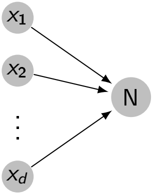

Introduction

$$\sum_{i=1}^{k} \beta_i g(A_i(\overrightarrow{x}))$$



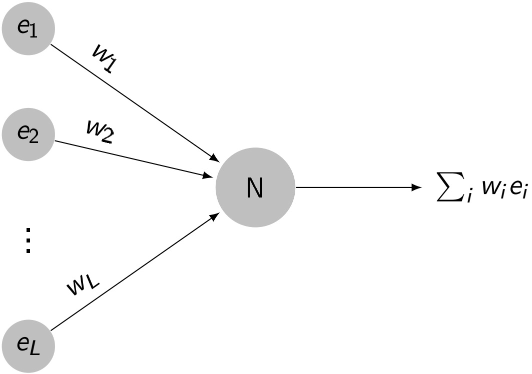

Densité

$$g(K(\overrightarrow{w}, \overrightarrow{x}) + b)$$




Dualité

# Neurone

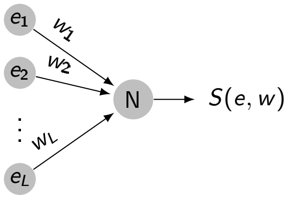



Fonction associée

$$
\begin{array}{ccccc}
\mathbb{R}^L & \to & \mathbb{R}^L & \to & \mathbb{R} \\
w & \mapsto & e & \mapsto & \mathbb{R}
\end{array}
$$

# Apprentissage - Descente de gradient

Apprendre $f : \mathbb{R}^L \to \mathbb{R}$.

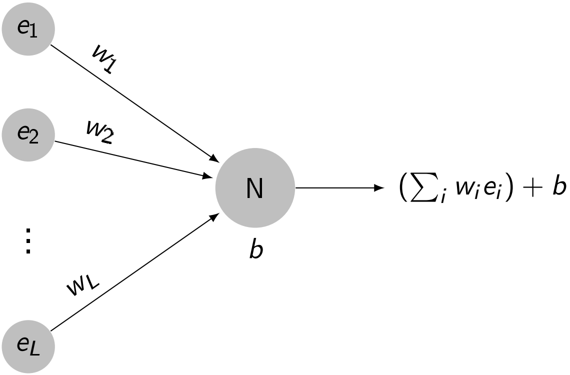



$$Err(e, w) = \frac{1}{2}(S(e, w) - f(e))^2$$

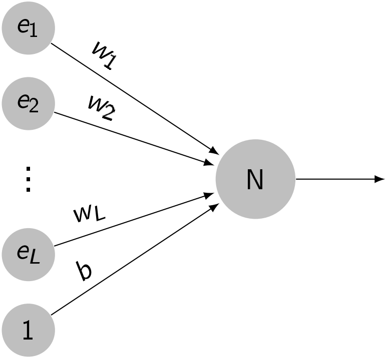
$$w_i \mapsto w_i - \eta \frac{\partial Err}{\partial w_i}$$

$$w \mapsto w - \eta \, \overrightarrow{grad_w} Err$$

# Biais



# Biais



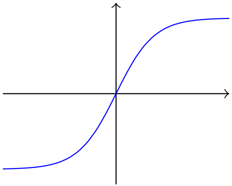
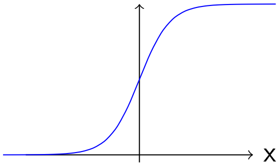

$$\left(\sum_i w_i e_i\right) + b$$

# Fonction d'activation

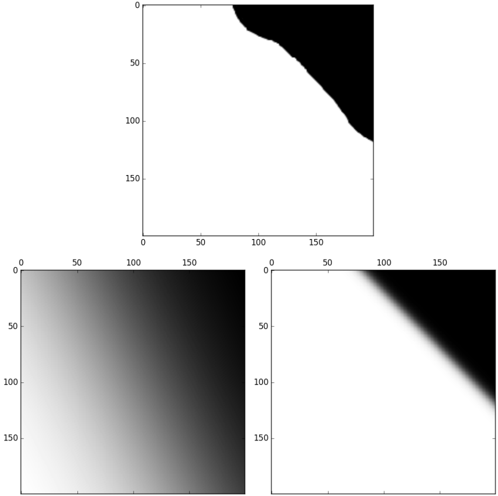
$g : \mathbb{R} \to \mathbb{R}$.

$$N(w, b) : e \mapsto g\left(\sum_i w_i e_i + b\right)$$

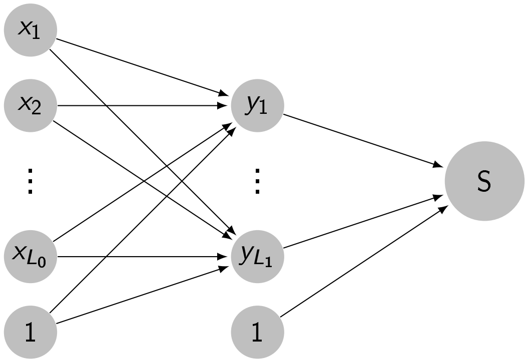
Ex :

| tanh | $\frac{1}{1+e^{-x}}$ |
| --- | --- |
|  |   |

# Exemple - Terre et Mer



# Couche intermédiaire



# $\Sigma^d(g)$

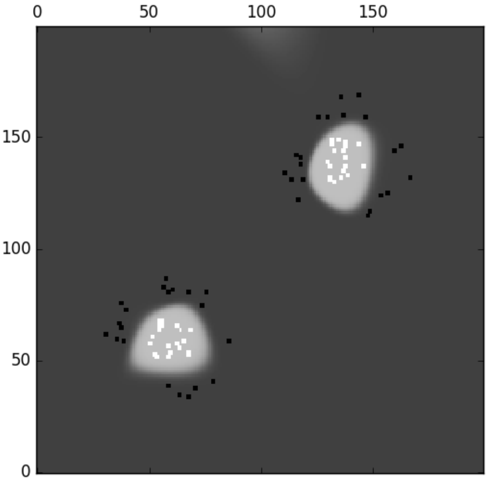
$g \in \mathcal{C}^0(\mathbb{R})$ et $d \in \mathbb{N}^*$.
$\mathbb{A}^d$ ensemble des fonctions affines de $\mathbb{R}^d$ dans $\mathbb{R}$.
On note

$$\Sigma^d(g) = \Big\{ f : \mathbb{R}^d \to \mathbb{R} \;\Big|\; \exists q \in \mathbb{N}^* \, \forall i \in [\![1, q]\!] \, \exists (\beta_i, A_i) \in \mathbb{R} \times \mathbb{A}^d, \\ \forall x \in \mathbb{R}^d \;\; f(x) = \sum_{i=1}^{q} \beta_i g(A_i(x)) \Big\}.$$

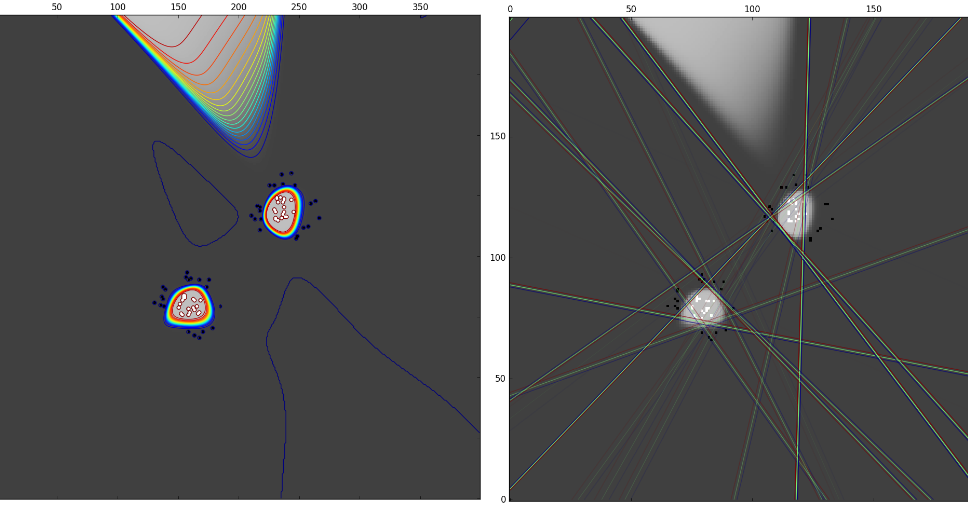
# Couches intermédiaires

- Rétropropagation
- Non-linéarité des fonctions d'activation

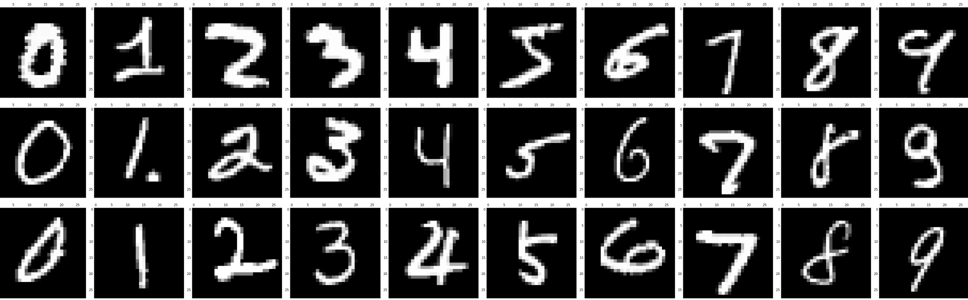
# Exemple - Îles



# Exemple - Au cœur du réseau



# Exemple - Reconnaissance d'images

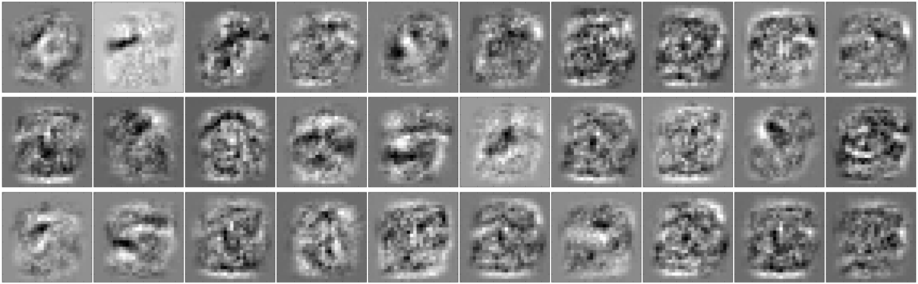


60000 exemples dans la base d'entraînement.
10000 exemples dans la base de test.

$$\mathbb{R}^{28\cdot 28} = \mathbb{R}^{784} \to \mathbb{R}^{10} \to [\![0, 9]\!]$$

# Exemple - Poids

- Environ $10^6$ itérations
- Erreur sur la base d'apprentissage : 1.13%
- Erreur sur la base de test : 2.28%
- Activations : tanh, sigmoïde.
- 200 neurones intermédiaires :



# Énoncé

### Théorème (Stone-Weierstrass)

*$K$ espace compact*
*$A$ sous-algèbre de $\mathcal{C}(K, \mathbb{R})$ telle que :*

- $\forall x, y \in K \ \exists f \in A, \ f(x) \neq f(y)$
- $\forall x \in K \ \exists f \in A, \ f(x) \neq 0$

*Alors, au sens de la convergence uniforme,*

$$\overline{A} = \mathcal{C}(K, \mathbb{R}).$$

# Fonctions saturantes

On appelle *fonction saturante* toute fonction $\psi : \mathbb{R} \to \mathbb{R}$ continue et croissante telle que $\lim_{x \to -\infty} \psi(x) = 0$ et $\lim_{x \to +\infty} \psi(x) = 1$.

# $\Sigma^d_{|K}(g)$

**Définition**

*$K \subset \mathbb{R}^d$, $g$ continue. $\Sigma^d_{|K}(g) = \left\{ f_{|K} \;\middle|\; f \in \Sigma^d_{|K}(g) \right\}$*

**Théorème**

*$K \subset \mathbb{R}^d$ compact, $\psi$ saturante.*
*Alors, au sens de la convergence uniforme,*

$$\overline{\Sigma^d_{|K}(\psi)} = \mathcal{C}(K, \mathbb{R}).$$

# $\Sigma\Pi^d(g)$

**Définition**

*$g \in \mathcal{C}^0(\mathbb{R})$ et $d \in \mathbb{N}^*$. On note*

$$\Sigma\Pi^d(g) = \left\{ f : \mathbb{R}^d \to \mathbb{R} \;\middle|\; \exists q \in \mathbb{N}^* \; \forall i \in [\![1, q]\!] \; \exists (\beta_i, l_i) \in \mathbb{R} \times \mathbb{N}^* \right.$$

$$\left. \forall k \in [\![1, l_i]\!] \; \exists A_{i,k} \in \mathbb{A}^d, \; \forall x \in \mathbb{R}^d \; f(x) = \sum_{i=1}^{q} \prod_{k=1}^{l_i} \beta_{i,k} g(A_{i,k}(x)) \right\}.$$

**Définition**

*$K \subset \mathbb{R}^d$, $g$ continue.* $\Sigma\Pi^d_{|K}(g) = \left\{ f_{|K} \;\middle|\; f \in \Sigma\Pi^d_{|K}(g) \right\}$

# $\Sigma\Pi^d(g)$

**Théorème**

*$g \in \mathcal{C}^0(\mathbb{R})$ non constante, $d \in \mathbb{N}^*$, $K \subset \mathbb{R}^d$ compact.*
*Alors, au sens de la convergence uniforme,*

$$\overline{\Sigma\Pi^d_{|K}(g)} = \mathcal{C}(K, \mathbb{R}).$$

# Réduction (1)

**Lemme**

$f, g \in \mathcal{C}^0(\mathbb{R})$, $d \in \mathbb{N}^*$.
*Si* $g \in \Sigma^1(f)$, *alors*

$$\Sigma^d(g) \subset \Sigma^d(f).$$

**Lemme**

$f, g \in \mathcal{C}^0(\mathbb{R})$, $d \in \mathbb{N}^*$.
*Si* $\forall M > 0$, $g_{|[-M,M]} \in \overline{\Sigma^1_{|[-M,M]}(f)}$,
*alors pour tout compact* $K \subset \mathbb{R}^d$,

$$\Sigma^d_{|K}(g) \subset \overline{\Sigma^d_{|K}(f)}$$

# Réduction (2)

## Proposition

*$\psi$, $\varphi$ des fonctions saturantes.*

$$\varphi \in \overline{\Sigma^1(\psi)}.$$

## Proposition

*La fonction suivante est saturante :*

$$\varphi : x \mapsto \begin{cases} 0 & \textit{si } x < -\pi/2 \\ \frac{1+\cos(x)}{2} & \textit{si } -\pi/2 \leq x < \pi/2 \\ 1 & \textit{si } \pi/2 \leq x. \end{cases}$$

*De plus,* $\forall M > 0$, $\cos_{|[-M,M]} \in \Sigma^1_{|[-M,M]}(\varphi)$.

# Réduction (3)

$K \subset \mathbb{R}^d$ compact. $\psi$ saturante.

## Proposition

$$\Sigma^d_{|K}(\cos) \subset \overline{\Sigma^d_{|K}(\psi)}.$$

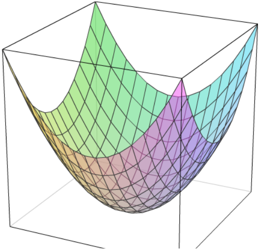
On en déduit que

$$\mathcal{C}(K,\mathbb{R}) = \overline{\Sigma\Pi^d_{|K}(\cos)} = \overline{\Sigma^d_{|K}(\cos)} \subset \overline{\Sigma^d_{|K}(\psi)} \subset \mathcal{C}(K,\mathbb{R}).$$

# Nouvelles fonctions d'agrégation

$$\overrightarrow{w} \cdot \overrightarrow{x} + b \longrightarrow K(\overrightarrow{w}, \overrightarrow{x}) + b$$

Ex :



$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \\ x^2 + y^2 \end{pmatrix}$$

On a alors accès aux

$$f : \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto ax + by + (a^2 + b^2)z$$

# Espaces de Hilbert (réels)

$(H, \langle\cdot,\cdot\rangle)$ préhilbertien réel, complet $\left[\langle\cdot,\cdot\rangle\right]$.
Exemples :

- $\mathbb{R}^d$
- $\ell^2_{\mathbb{R}} = \{u \in \mathbb{R}^{\mathbb{N}} | \sum_{n=0}^{+\infty} u_n^2 < +\infty\}$,
$$\langle u, v\rangle = \sum_{n=0}^{+\infty} u_n v_n$$
- $H^1_{\mathbb{R}} = \{u \in \mathbb{R}^{\mathbb{N}} | u_0^2 + \sum_{n=1}^{+\infty} n^2 u_n^2 < +\infty\}$,
$$\langle u, v\rangle = u_0 v_0 + \sum_{n=1}^{+\infty} n^2 u_n v_n$$

# Sous-ev fermés : Projection, orthogonal

$H$ espace de Hilbert, $F$ sev fermé de $H$.
Il existe $P_F : H \to F$ linéaire telle que

$$\forall x \in F,\ ||x - P_F(x)|| = d(x, F).$$

De plus, $x - P_F(x) \in F^\perp$.
Ainsi,

$$F \oplus F^\perp = H.$$

# Représentation de Riesz

$(H, \langle \cdot, \cdot \rangle)$ espace de Hilbert réel.

$$H' = \{f \in \mathcal{L}(H, \mathbb{R}) |\ f \text{ continue}\}.$$

$$\forall f \in H',\ \exists! \widetilde{f},\ \forall x \in H,\ f(x) = \langle \widetilde{f}, x \rangle.$$

En effet,

$$\begin{array}{cccc} \varphi & : & H & \rightarrow & H' \\ & & y & \mapsto & \langle y, \cdot \rangle \end{array}$$ est un isomorphisme isométrique.

# Noyau reproduisant

$X$ ensemble, $H \subset \mathbb{R}^X$ espace de Hilbert.
$H$ est un *espace de Hilbert à noyau reproduisant* ssi l'évaluation en chaque point $x \in X$ est continue

$$\begin{array}{rcccl} L_x & : & H & \to & \mathbb{R} \\ & & f & \mapsto & f(x). \end{array}$$

Dans ce cas,

$$\forall x \in X,\ \exists! K_x \in H,\ \forall f \in H,\ f(x) = \langle f, K_x \rangle.$$

On définit alors le *noyau reproduisant* de $H$

$$\begin{array}{rcccl} K & : & X \times X & \to & \mathbb{R} \\ & & (x, y) & \mapsto & K_x(y) = \langle K_x, K_y \rangle. \end{array}$$

# Théorème de Moore-Aronszajn

## Théorème

*$X$ ensemble, $K : X \times X \to \mathbb{R}$ noyau symétrique défini positif.*
*Alors, il existe un unique espace de Hilbert minimal $H \subset \mathbb{R}^X$ dont $K$ est un noyau reproduisant.*

Pour $x \in X$, $K_x = K(x, \cdot)$.
Préhilbertien $H_0$, complété en $H$ si nécessaire :

$$H_0 = Vect\{K_x | x \in X\}$$

$$\left\langle \sum_{i=1}^{n} a_i K_{x_i}, \sum_{j=1}^{m} b_j K_{y_j} \right\rangle_{H_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j K(x_i, y_j)$$

# Construction d'espace

$$H_0 = Vect\{K_x | x \in X\}$$

$$\left\langle \sum_{i=1}^{n} a_i K_{x_i}, \sum_{j=1}^{m} b_j K_{y_j} \right\rangle_{H_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j K(x_i, y_j)$$

Les suites de Cauchy de $H_0$ convergent simplement.

$$H = \{f \in \mathbb{R}^X | \exists (f_n) \in H_0^{\mathbb{N}} \text{ de Cauchy, } f_n \xrightarrow{\text{cvs}} f\}$$

$$\langle f, g \rangle_H = \lim_{n \to +\infty} \langle f_n, g_n \rangle_{H_0} \text{ où } f_n \xrightarrow{\text{cvs}} f, \ g_n \xrightarrow{\text{cvs}} g, \text{ de Cauchy } \in H_0^{\mathbb{N}}$$