

Réseaux de neurones

Partie 2a

Séparabilité et Backprop

Bruno Bouzy

bruno.bouzy@parisdescartes.fr

UE IA L3

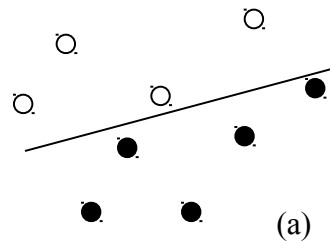
Avril 2022

Séparabilité

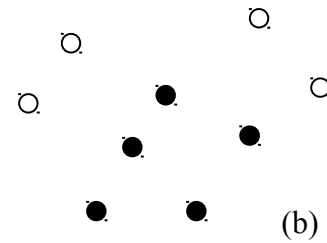
- Un neurone classifiant sépare l'espace de ses entrées en deux demi-espaces

- Séparation :

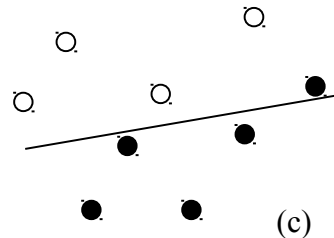
- Correcte (a)



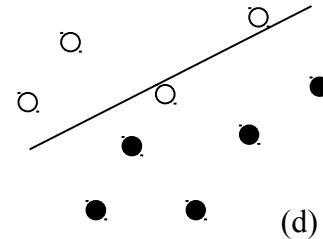
- Impossible (b)



- Incorrecte (c)



- Incorrecte (d)



- Théorème de convergence :

- Pour un ensemble d'exemples séparable, le perceptron trouve une solution en un nombre fini d'itérations.

Backprop aperçu de la théorie

- Un réseau organisé en couches

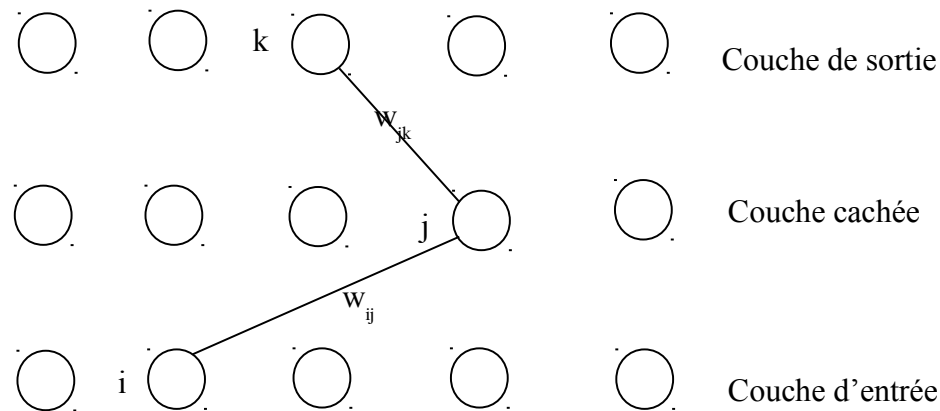


Figure 4 : Un réseau général à trois couches. Si w_{jk} change alors, seul k change.
Si w_{ij} change alors toutes les valeurs de la couche de sortie changent.

La fonction d'erreur

- Sigmoide

$$o_k = 1/(1+e^{-net_k}) \quad (7)$$

- Combinaison linéaire

$$net_k = \sum_j w_{jk} o_j \quad (8)$$

- Fonction d'erreur

$$E = 1/2 \sum_p (\sum_k (t_{pk} - o_{pk})^2) \quad (9a)$$

- Simplification

$$E = 1/2 \sum_k (t_k - o_k)^2 \quad (9b)$$

Mise à jour des poids d'un neurone de sortie

- Dérivation

$$\begin{aligned}\partial E / \partial w_{jk} &= \partial E / \partial o_k \quad \partial o_k / \partial \text{net}_k \quad \partial \text{net}_k / \partial w_{jk} \\ &= - (t_k - o_k) o_k (1 - o_k) o_j\end{aligned}\quad (10)$$

- Descente de gradient

$$\Delta w_{jk} = - \eta \partial E / \partial w_{jk} \quad (11)$$

- Formule

$$\Delta w_{jk} = \eta d_k o_j \quad (12)$$

Mise à jour des poids d'un neurone caché

- Dérivations

$$\partial E / \partial w_{ij} = \sum_k \partial E_k / \partial w_{ij}$$

$$\partial E / \partial w_{ij} = \sum_k \partial E_k / \partial o_k \partial o_k / \partial \text{net}_k \partial \text{net}_k / \partial o_j \partial o_j / \partial \text{net}_j \partial \text{net}_j / \partial w_{ij}$$

$$\partial E / \partial w_{ij} = \sum_k - (t_k - o_k) o_k (1 - o_k) w_{jk} o_j (1 - o_j) o_i$$

$$\partial E / \partial w_{ij} = \sum_k - d_k w_{jk} o_j (1 - o_j) o_i$$

- Suite...

$$\partial E / \partial w_{ij} = - o_j (1 - o_j) o_i \sum_k d_k w_{jk}$$

$$\partial E / \partial w_{ij} = - o_i d_j$$

- Formule

$$\Delta w_{ij} = \eta d_j o_i \quad (14)$$