

Réseaux de neurones

Partie 2b

Multi-layer perceptron (MLP)

Bruno Bouzy

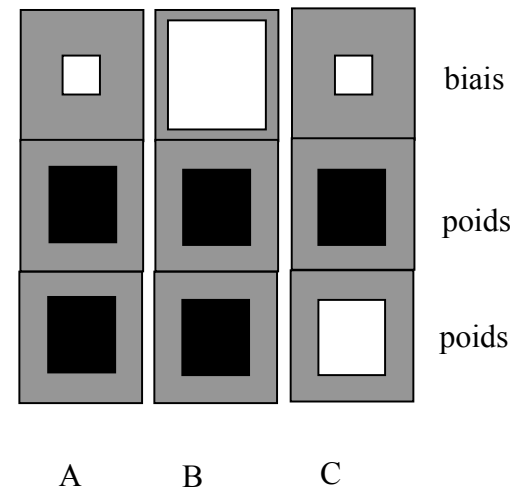
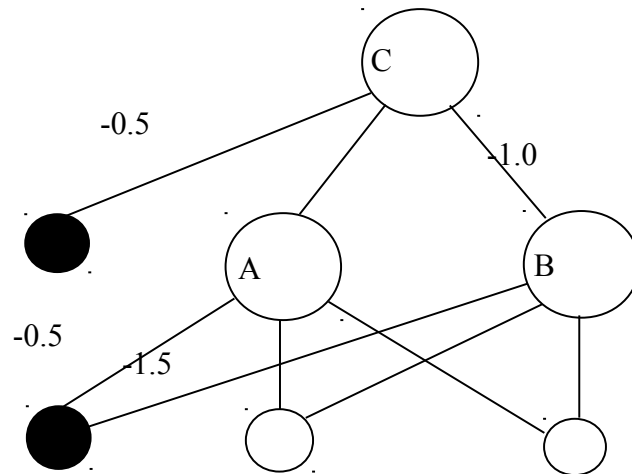
bruno.bouzy@parisdescartes.fr

Avril 2022

UE IA L3

Diagrammes de Hinton

- Soit un réseau à 3 neurones solution du XOR



- Un diagramme de Hinton visualise les poids des connexions du réseau
 - Carré noir : poids positif
 - Carré blanc : poids négatif
 - Taille du carré : valeur absolue du poids

Pourquoi ce réseau à 3 neurones ?

- Le tableau montre que ce réseau est solution du XOR

X	Y	A	f(A)	B	f(B)	C	f(C)
0	0	-0.5	0	-1.5	0	-0.5	0
0	1	+0.5	1	-0.5	0	+0.5	1
1	0	+0.5	1	-0.5	0	+0.5	1
1	1	+1.5	1	+0.5	1	-0.5	0

- D'où viennent ces valeurs ?

Hyperplans séparateurs et opérateurs logiques (1/2)

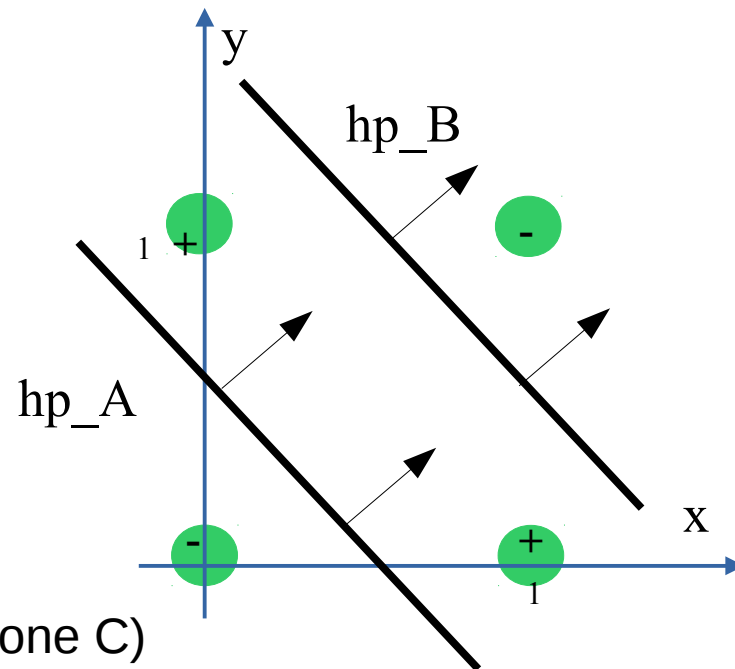
- 1ère couche : 2 hyperplans séparateurs (neurones A et B)

- hp_A :

- (0, 0) est négatif
 - (1, 0) (0, 1) (1, 1) sont positifs
 - $x + y - 0.5 > 0$

- hp_B :

- (1, 1) est positif
 - (1, 0) (0, 1) (0, 0) sont négatifs
 - $x + y - 1.5 > 0$



- 2ème couche : opérateurs logiques (neurone C)

- Sur le résultat de A et NON B
 - $A - B - 0.5 > 0$

Hyperplans séparateurs et opérateurs logiques (2/2)

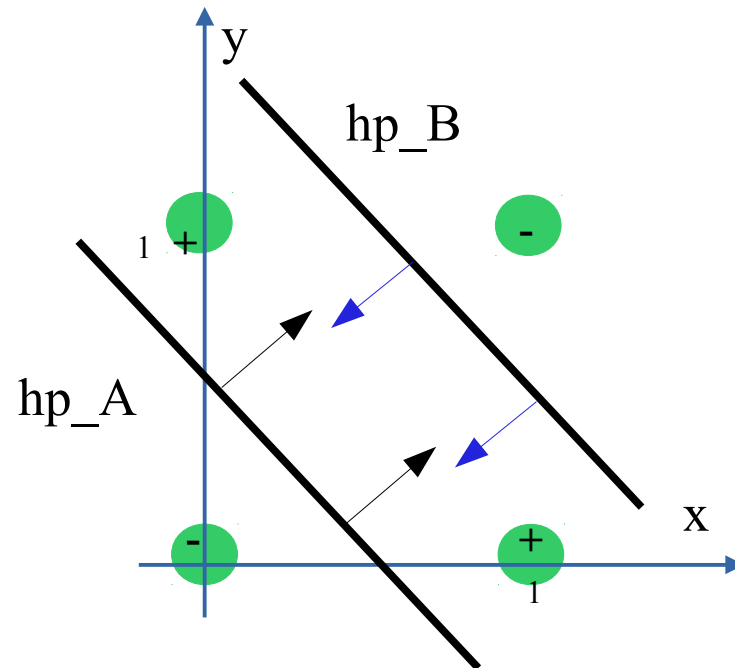
- Une autre solution
- 1ère couche : 2 hyperplans séparateurs (neurones A et B)

- hp_A :

- (0, 0) est négatif
 - (1, 0) (0, 1) (1, 1) sont positifs
 - $x + y - 0.5 > 0$

- hp_B :

- (1, 1) est négatif
 - (1, 0) (0, 1) (0, 0) sont positifs
 - $x + y - 1.5 < 0$



- 2ème couche : opérateurs logiques (neurone C)
 - Sur le résultat de A et B
 - $A + B - 1.5 > 0$
- Il y a d'autres solutions...

Vocabulaire

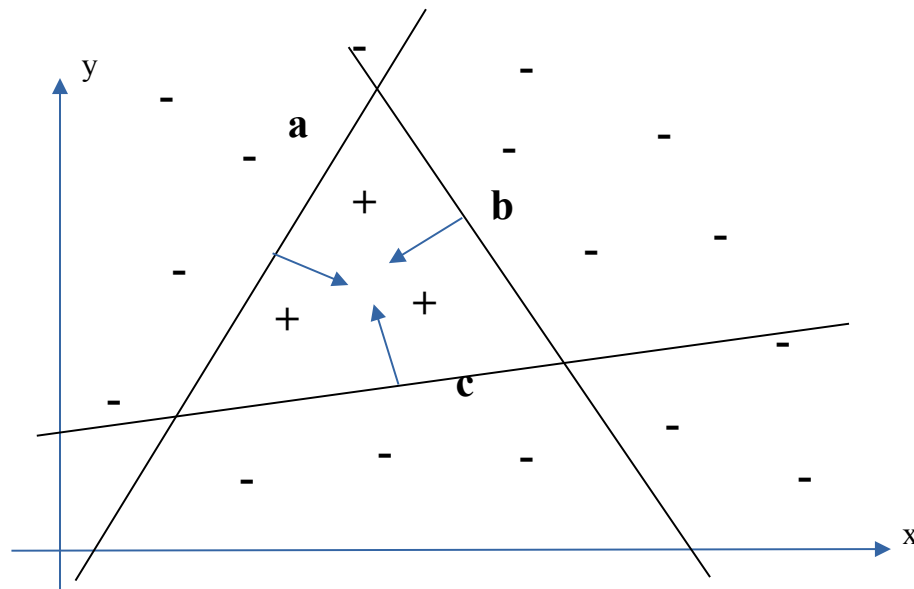
- Réseau Multi-Couches
- Multi-Layer Perceptron (MLP)
- Les entrées d'un neurone d'une couche N ne sont connectées qu'à des sorties de neurones de la couche N-1

Simuler un opérateur logique avec un neurone

- OU logique
 - $\text{Netz} = \text{Somme des entrées} - 0.5$
 - Si une entrée au moins vaut 1, la somme sera > 0.5
 - < -0.5 sinon
- ET logique
 - $\text{Netz} = \text{Somme des entrées} - N + 0.5$
 - N est le nombre d'entrées
 - Si toutes les entrées valent 1, $\text{somme} - N + 0.5 > 0.5$
 - < -0.5 sinon
- NON logique
 - Mettre le signe -1 pour cette entrée

Reconnaitre une zone convexe

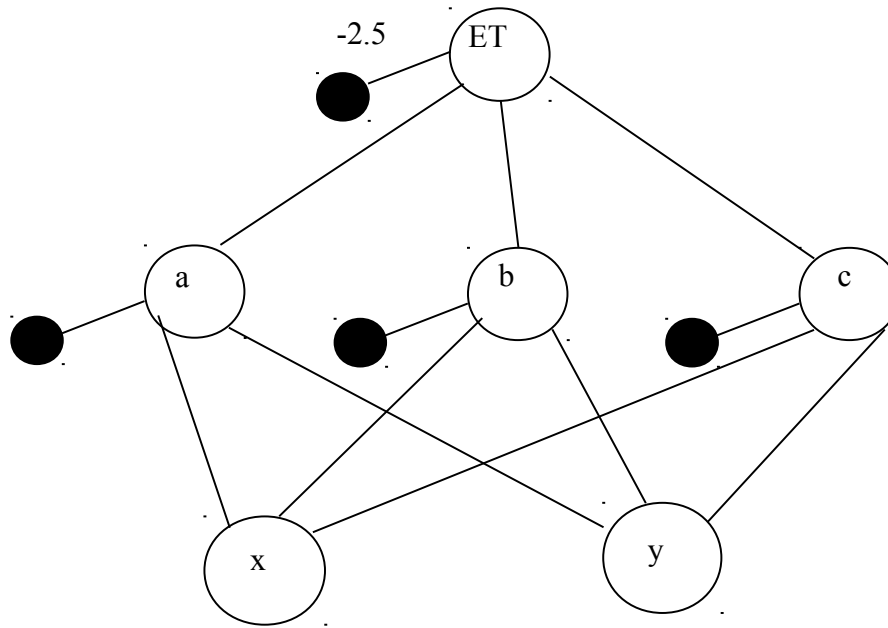
- Les exemples + sont à l'intérieur de la zone
- Les exemples – à l'extérieur



- On entoure la zone avec des hyperplans a, b, c
- On synthétise les résultats avec un ET logique

Avec un réseau à 2 couches

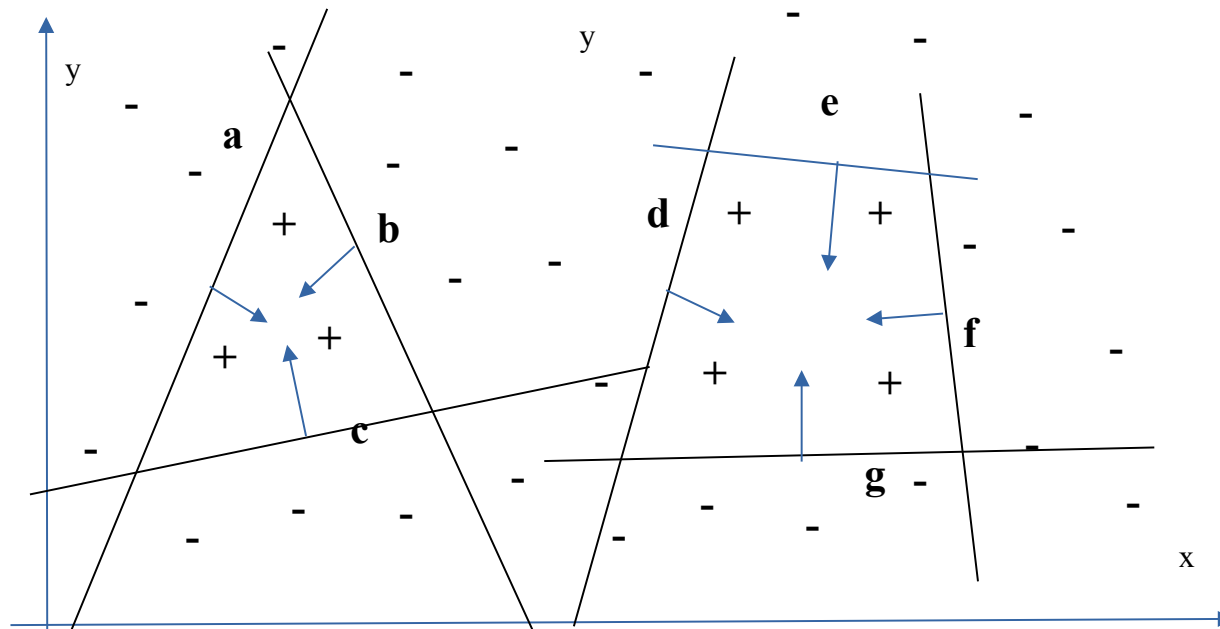
- 1ère couche : les hyperplans séparateurs



- 2ème couche : l'opérateur logique

Reconnaître une union de zones convexes

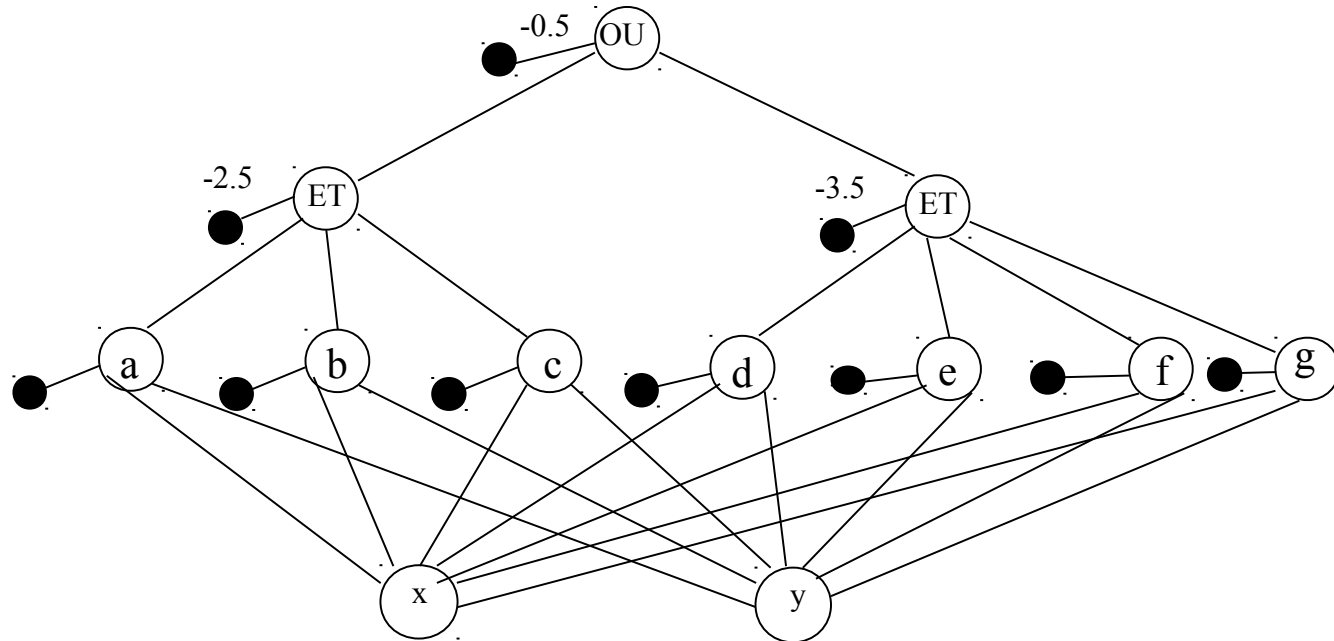
- Les exemples + sont à l'intérieur d'une union de zones convexes



- On entoure chaque zone avec des hyperplans
- On synthétise avec des ET et des OU logiques

Avec un réseau à 3 couches

- 1ère couche : hyperplans séparateurs



- 2ème couche : des ET logiques
- 3ème couche : un OU logique

Pratique et théorie (1/2)

- « théorie »
 - Preuve que l'on peut classifier avec telle ou telle architecture en suivant telle ou telle méthode
 - Par exemple, la mini-théorie sur les 2 ou 3 couches nécessaires pour classifier n'importe quoi avec des hyperplans des ET et de OU
 - Pourvu qu'il n'y ait pas de limitation sur le nombre de neurones en couche cachée, il est théoriquement possible de classer « à peu près n'importe quel ensemble d'exemples » avec un réseau à 1 couche cachée (c'est-à-dire 2 couches en tout).
 - Référence à retrouver...

Pratique et théorie (2/2)

- «pratique»
 - Faire converger un réseau avec telle ou telle architecture, avec Backprop (ou amélioration)
 - Le réseau n'a pas d'à priori (il ne connaît pas les ET et le OU, il les retrouve si besoin)
 - La première couche contient les hyperplans nécessaires à la distinction des caractéristiques du problème
 - Les couches supérieures abstraient des concepts sur les concepts des couches inférieures
 - Il est inhumain de comprendre comment un réseau fonctionne à partir de ses poids finaux
 - Un réseau de neurones reste une « boîte noire » qui marche ou pas.
 - Le concepteur essaie des architectures et des variantes de Backprop qui sont des descentes de gradients

Les actions et observations du concepteur

- Les observations sont celles d'une descente de gradient
 - Le réseau converge-t-il ?
 - Oui : recommencer avec un pas d'apprentissage plus grand pour aller plus vite
 - Non :
 - Le pas d'apprentissage est-il trop grand ?
 - Cela vaut-il le coup de relancer sur d'autres poids de départ ?
- Les actions
 - Changer la représentation des entrées
 - Augmenter le nombre de neurones en couches basses
 - Augmenter le nombre de couches
 - Utiliser une autre fonction d'activation ?
 - Sigmoid, ReLU, identité, autre ?
 - Utiliser des connexions résiduelles (sautant des couches)
 - Utiliser des réseaux convolutionnels ?
- Méthode expérimentale par essais erreurs

Conclusion

- Aujourd'hui, les réseaux de neurones constituent un outil très puissant de l'intelligence artificielle
- Ils ont été valorisés :
 - ces dernières décennies, par l'augmentation de la puissance des machines,
 - ces dernières années par l'utilisation de GPU en parallèle pour les réseaux convolutionnels
 - par l'amélioration des algorithmes de convergence
- La théorie repose sur des concepts
 - Anciens (1960 à 1990)
 - Perceptrons, MLP
 - Récents (depuis 2010)
 - Réseaux convolutionnels
 - Réseaux résiduels
 - ReLU
- Nombreux succès en IA :
 - Approximation de fonctions complexes (compliquées)
 - Traitement d'images
 - Traitement du langage naturel
 - Programmation des jeux
 - AlfaGo, alfaZero