

Introduction au Deep LEarning

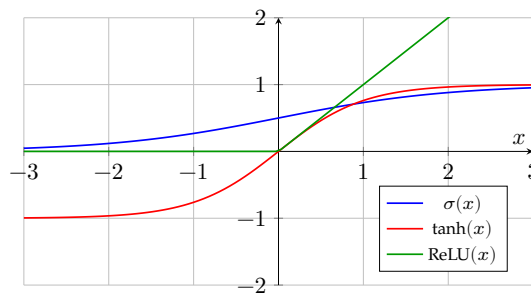
Architectures de réseaux de neurones

Olivier Schwander <olivier.schwander@sorbonne-universite.fr>

2025-2026

Exercice 1 - Fonctions d'activation

Les fonctions d'activation introduisent de la non-linéarité dans les réseaux de neurones. Étudions trois fonctions classiques :



Définitions :

- **Sigmoid** : $\sigma(x) = \frac{1}{1 + e^{-x}}$ (sortie dans $[0, 1]$)
- **Tanh** : $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ (sortie dans $[-1, 1]$)
- **ReLU** : $\text{ReLU}(x) = \max(0, x)$ (sortie dans $[0, +\infty[$)

Question 1 - Importance des activations

- a. Dans l'architecture précédente, que devient la passe avant si l'on supprime la fonction d'activation de la couche cachée ?
- b. Quel est l'intérêt de l'activation finale ?
- c. Conclure sur l'intérêt des activations.

Question 2 - Comportement

- a. Pour chaque fonction d'activation, donner :
 - Image (ensemble des valeurs possibles)
 - Dérivée
 - Comportement à l'infini
- b. Pourquoi la saturation pose-t-elle problème dans les réseaux profonds lors de la rétropropagation ?

Question 3 - Vanishing gradient

On considère un réseau à K couches cachées avec activation sigmoid.

- Si toutes les activations d'une couche sont proches de 0 (saturation à gauche), que vaut la dérivée $\sigma'(z) = \sigma(z)(1 - \sigma(z))$?
- Que devient le gradient après K couches si chaque dérivée locale est de l'ordre de 10^{-3} ? Conclure.
- La fonction ReLU a-t-elle le même problème ? Donner les valeurs possibles de sa dérivée.

Exercice 2 - Images et MLP

Question 4 - Explosion de nombre de paramètres

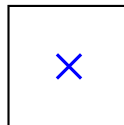
On considère une image noir&blanc de taille 64×64 pixels pour une tâche de classification binaire.

On construit un MLP avec l'architecture suivante :

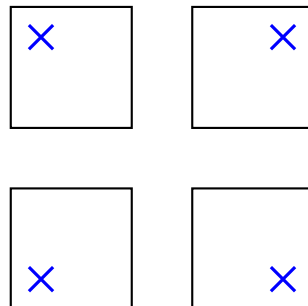
- Couche d'entrée** : Elle reçoit l'image aplatie (vecteur).
 - Couche cachée 1** : 1000 neurones, activation ReLU.
 - Couche de sortie** : 10 neurones (softmax).
- Quelle est la dimension du vecteur d'entrée ?
 - Calculer le nombre de paramètres de la première couche cachée.
 - Calculer le nombre total de paramètres du réseau.
 - Et sur une image de taille normale (par exemple 1920×1080). Conclure sur la faisabilité d'un MLP pour des images de grande taille.

Question 5 - Invariances spatiales

Supposons que notre réseau a été entraîné à reconnaître un "X" sur des images noir et blanc 64×64 . Durant l'entraînement, tous les exemples de "X" étaient parfaitement centrés dans l'image :



On présente maintenant au réseau des images de test contenant un "X", mais décalées par rapport au centre :



- Dans les couches complètement connectées de ce réseau, quelles connexions auront tendance à avoir des poids élevés ? (On peut imaginer un perceptron plutôt qu'un MLP pour simplifier.)
- Commenter la généralisation du modèle aux images avec le X décalé.
- Proposer une amélioration du dataset d'entraînement pour reconnaître le X peu importe sa position.
- Commenter la quantité de données nécessaire. Peut-on espérer généraliser ?

Question 6 - Besoin de données

Toujours avec des images 64×64 et une tâche de classification binaire, considérons que les objets d'intérêt sont petits par rapport à la taille de l'image (disons des X de taille 5×5).

- a. Combien y a-t-il de positions possibles pour placer ce X de 5×5 dans une image de 64×64 ?
- b. Si l'on considère que pour généraliser correctement un MLP a besoin d'au moins quelques exemples pour chaque configuration possible, combien d'images faudrait-il pour couvrir toutes les positions du cercle?
- c. Si l'on ajoute un degré de liberté supplémentaire (comme les rotations ou le changement d'échelle, comment évolue le nombre d'images nécessaires?
- d. Conclure sur les capacités des MLP à gérer des images complexes.

Question 7 - Vers les convolutions

Considérons maintenant un MLP plus simple, entraîné sur des images *10 times 10* contenant des X bien centrés de taille 5×5 .

- a. Commenter le nombre de paramètres de ce nouveau MLP pour réussir à bien généraliser.
- b. Comment utiliser ce modèle pour réaliser la classification sur les images 64×64 ?
- c. Commenter la généralisation, le nombre de paramètres et la quantité de données d'entraînement nécessaire avec cette nouvelle stratégie.