



Dr. Jehan-Antoine Vayssade

19 juillet 2025

Instructions

- **Durée** : 1 heure 30 min
- **Note totale / maximale** : 20 points
- **Les réponses doivent être argumentées**
- **Prenez uniquement un stylo**

1 Perceptron Simple (4 points)

Expliquez le fonctionnement du perceptron simple et donnez son équation mathématique de sortie en supposant deux variables d'entrée x_0 et x_1 . Vous ferez apparaître les différentes variables ainsi que le résultat intermédiaire, souvent noté $z = \dots$. Puis écrivez la solution $f(z)$.

À partir des points suivants, déterminez si le perceptron simple peut le résoudre et expliquez pourquoi. Sinon, proposez une solution logique.

Point	a_0	a_1	Classe
A	1	1	1
B	2	2	1
C	1	3	0
D	3	1	0

2 Fonction d'activation (5 points)

On considère plusieurs fonctions d'activation couramment utilisées dans les réseaux de neurones : Heaviside, ReLU, Leaky ReLU, Sigmoid, Tanh et GELU. L'image ci-dessous représente ces fonctions (courbe bleue) ainsi que leurs dérivées (courbe orange pointillée).

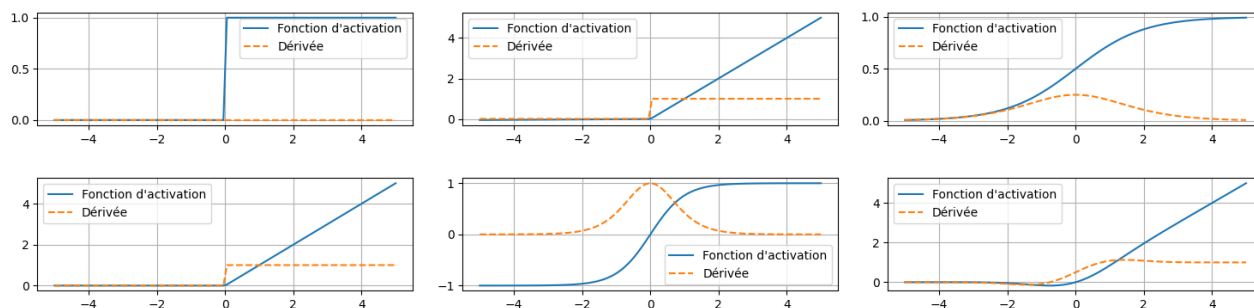


FIGURE 1 – Fonctions d'activation et leurs dérivées

- (1 pt) Pour chaque graphe, nommez la fonction d'activation représentée.
- (1 pt) Parmi les fonctions présentées, lesquelles sont différentiables partout ?
- (1 pt) Quelle(s) fonction(s) d'activation peut (peuvent) provoquer un phénomène de *vanishing gradient* ?
- (1 pt) Comparez la fonction ReLU et la fonction Leaky ReLU. Pourquoi la version « Leaky » a-t-elle été introduite ? Quels problèmes corrige-t-elle ?
- (bonus) À partir de vos observations sur les dérivées, qu'apporte GeLU ?
- (1 pt) Écrivez l'équation de la fonction ReLU et de sa dérivée.

3 Perceptron multicouche (5 points)

1. (2 pts) Expliquez pourquoi les fonctions d'activation sont nécessaires dans un réseau de neurones. Justifiez votre réponse à l'aide d'un graphique représentant plusieurs neurones, leurs courbes d'activation, ainsi que la sortie finale.
2. (2 pts) Étendez le concept à plusieurs couches et déduisez en fonction de la "hauteur" et de la "largeur" du réseau, quels sont le degré et la précision de la fonction reconstruite.
3. (1 pt) Expliquez le principe de la descente de gradient, notamment le rôle de la fonction de perte (loss).

4 Réseau convolutif (3 points)

1. (2 pts) Expliquez le principe de l'opération de convolution dans le contexte des réseaux de neurones. Exemple : Quelle est son utilité par rapport à une couche dense classique ? Quelle sont les intérêts par rapport à des filtres classiques tels que **Sobel** ou **Laplace** ?

Rapelle le filtre de Sobel est défini par :

$$G(X) = G_x(X) + G_y(X) = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} * X + \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} * X$$

2. (1 pt) Quelle est la fonction d'une couche de pooling (max pooling ou average pooling) ? En quoi aide-t-elle l'apprentissage ?
3. (1 pt) Proposez le résultat d'une couche de pooling et/ou de convolution sur l'image suivante :

$$I = \begin{bmatrix} 12 & 50 & 80 & 50 & 12 \\ 50 & 100 & 150 & 100 & 50 \\ 80 & 150 & 255 & 150 & 80 \\ 50 & 100 & 150 & 100 & 50 \\ 12 & 50 & 80 & 50 & 12 \end{bmatrix}$$

5 Ensembles d'entraînement (2 points)

Expliquez la différence entre les ensembles d'entraînement, de validation et de test dans le cadre de l'apprentissage supervisé. Quel est le rôle de chacun ? Pourquoi est-il important de bien les séparer ? Donnez un exemple de problème pouvant survenir si cette séparation n'est pas respectée.