Provenance sur les réseaux de neurones

Paul LANDRIER

1 Exemples

1.1 Réseaux de neurones

On s'intéresse d'abord à des réseaux de neurones denses avec une fonction d'activation f.

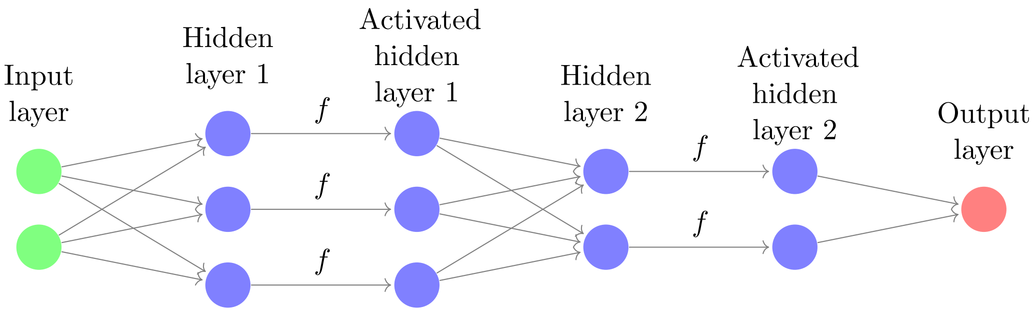


Figure 1: Réseau de neurone N

On appelle respectivement E, H_1, H_1', H_2, H_2' et o les vecteurs représentant les couches successives. On définit (A_1, B_1), (A_2, B_2) et (A_3, B_3) les couples matrices vecteurs tels que $H_1 = A_1E + B_1$, $H_2 = A_2H_1' + B_2$ et $o = A_3H_2' + B_3$.

En écrivant explicitement l'exécution générale, on obtient :

- $E = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$
- $H_1 = A_1E + B_1 = \begin{pmatrix} a^1_{1,1}x_1 + a^1_{1,2}x_2 + b^1_1 \\ a^1_{2,1}x_1 + a^1_{2,2}x_2 + b^1_2 \\ a^1_{2,1}x_1 + a^1_{2,2}x_2 + b^1_2 \end{pmatrix}$
- $H_1' = f(H_1) = \begin{pmatrix} f(a^1_{1,1}x_1 + a^1_{1,2}x_2 + b^1_1) \\ f(a^1_{2,1}x_1 + a^1_{2,2}x_2 + b^1_2) \\ f(a^1_{3,1}x_1 + a^1_{3,2}x_2 + b^1_3) \end{pmatrix}$
- $H_2 = A_2H_1' + B_2 =$ $\begin{pmatrix} a^2_{1,1}f(a^1_{1,1}x_1 + a^1_{1,2}x_2 + b^1_1) + a^2_{1,2}f(a^1_{2,1}x_1 + a^1_{2,2}x_2 + b^1_2) + a^2_{1,2}f(a^1_{3,1}x_1 + a^1_{3,2}x_2 + b^1_3) + b^2_1 \\ a^2_{2,1}f(a^1_{1,1}x_1 + a^1_{1,2}x_2 + b^1_1) + a^2_{2,2}f(a^1_{2,1}x_1 + a^1_{2,2}x_2 + b^1_2) + a^2_{2,2}f(a^1_{3,1}x_1 + a^1_{3,2}x_2 + b^1_3) + b^2_2 \end{pmatrix}$
- $H_2' = f(H_2) =$ $\begin{pmatrix} f(a^2_{1,1}f(a^1_{1,1}x_1 + a^1_{1,2}x_2 + b^1_1) + a^2_{1,2}f(a^1_{2,1}x_1 + a^1_{2,2}x_2 + b^1_2) + a^2_{1,2}f(a^1_{3,1}x_1 + a^1_{3,2}x_2 + b^1_3) + b^2_1) \\ f(a^2_{2,1}f(a^1_{1,1}x_1 + a^1_{1,2}x_2 + b^1_1) + a^2_{2,2}f(a^1_{2,1}x_1 + a^1_{2,2}x_2 + b^1_2) + a^2_{2,2}f(a^1_{3,1}x_1 + a^1_{3,2}x_2 + b^1_3) + b^2_2) \end{pmatrix}$
- $o = a^3_{11}f(a^2_{1,1}f(a^1_{1,1}x_1 + a^1_{1,2}x_2 + b^1_1) + a^2_{1,2}f(a^1_{2,1}x_1 + a^1_{2,2}x_2 + b^1_2) + a^2_{1,2}f(a^1_{3,1}x_1 + a^1_{3,2}x_2 + b^1_3) + b^2_1)$ $+ a^3_{22}f(a^2_{2,1}f(a^1_{1,1}x_1 + a^1_{1,2}x_2 + b^1_1) + a^2_{2,2}f(a^1_{2,1}x_1 + a^1_{2,2}x_2 + b^1_2) + a^2_{2,2}f(a^1_{3,1}x_1 + a^1_{3,2}x_2 + b^1_3) + b^2_2)$

1.2 Graphes de calculs

Les graphes de calculs sont une autre façon de représenter l'exécution d'un réseau de neurones, en mettant en évidence la structure des calculs.

Commençons par un exemple. Nous définissons la fonction $f(x, y, z) = xe^{(x^2-y^2)z}$. Son graphe de calcul est représenté en Figure 1.2.

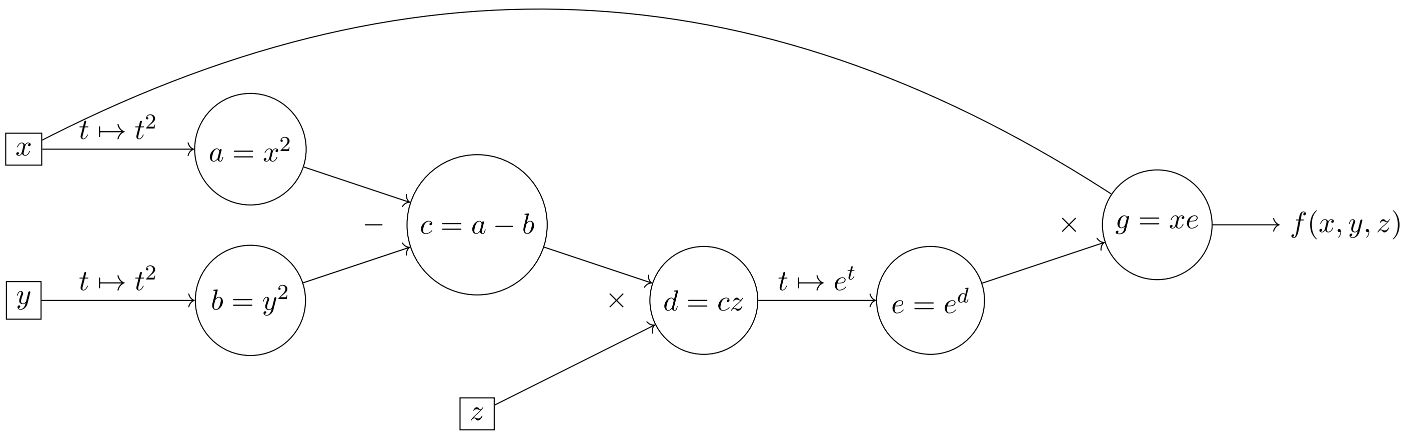
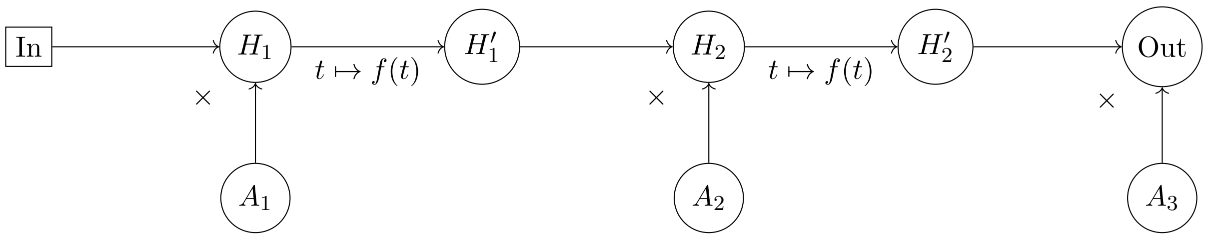


Figure 2: Graphe de calcul de la fonction f

Pour le réseau de neurones N, en considérant les opérations sur les vecteurs, cela donne :



1.3 Graphes de calculs - *Backward pass*

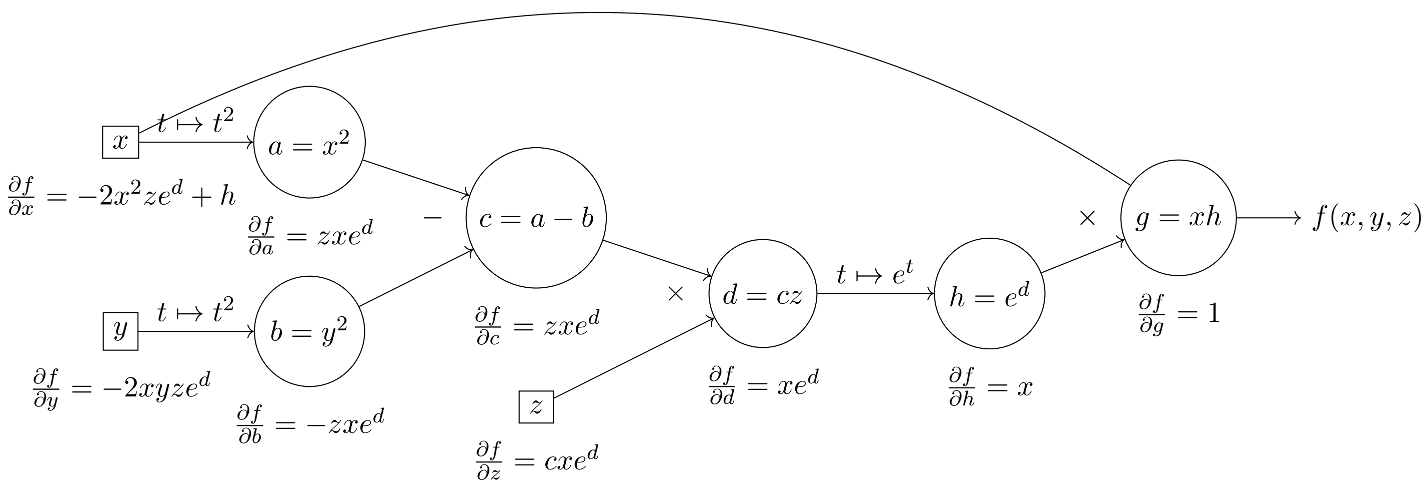


Figure 3: Exemple de *backward pass*.

Les graphes de calcul tels que représentés précédemment permettent de simuler une exécution du réseau de neurones ou de la fonction considérée. Cependant, leur intérêt ne se limite pas à cela puisqu'ils permettent aussi de calculer efficacement le gradient de la fonction considérée en remontant le graphe, c'est la *backward pass*.

L'idée repose sur la règle de la chaine :

$$\frac{\partial f}{\partial x} = \sum_{y \text{ child of } x} \frac{\partial y}{\partial x}\frac{\partial f}{\partial y}$$

Formellement, on voit d'abord f comme une fonction de g (la fonction identité) et on a donc $\frac{\partial f(g)}{\partial g} = 1$. Puis on voit g comme une fonction de x et e et on a $\frac{\partial f(g(x,e))}{\partial e} = x$, puis on voit e comme une fonction de d et ainsi de suite jusqu'à obtenir $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ et $\frac{\partial f}{\partial z}$ qui sont les trois coordonnées du gradient de f.

Le résultat de ces opérations sur le graphe de calcul est montré en figure 1.3.

2 Provenance

Nous souhaitons abstraire les méthodes précédentes à travers une strucuture algébrique adaptée, de façon analogue à la provenance par semi-anneaux dans les bases de données introduite dans [Green et al.].

2.1 Généralisation par les semi-anneaux

L'idée la plus directe consiste à remplacer

2.2 Généralisation

Du point de vue de cette provenance fine, on distingue trois grandes catégries d'objets. La première est constituée des éléments x_1, x_2 de l'input et des biais b, la deuxième est constituée des coefficients des matrices $a_{i,j}$ et la troisième est constituée de fonctions f. Nous appelons E le premier ensemble, A le deuxième et F le troisième.

Afin d'abstraire ce fonctionnement, il convient de trouver des structures adaptées pour ces trois catégories. Pour cela nous commençons par quelques observations :

- Les coefficients $a_{i,j}$ agissent sur les coefficients x_i, par une application $\left\{ \begin{array}{ccc} A \times E & \to & E \\ (a, x) & \mapsto & ax \end{array} \right.$
- L'ensemble E est muni d'une opération $+$.