Rétropropagation

Introduction à l'apprentissage automatique
Master Sciences Cognitives
Aix Marseille Université

Alexis Nasr

Motivations

- La phase d'apprentissage des réseaux de neurones nécessite généralement de calculer **le gradient de la fonction d'erreur**, sur les données d'apprentissage.
- La taille des réseaux peut être **importante**, ainsi que le nombre de données d'aprentissage.
- Il est donc nécessaire de trouver un moyen **rapide** pour calculer le gradient.
- C'est ce que permet de faire l'algorithme de **rétropropagation**.

Objectifs

- Comprendre la notion de graphe de calcul d'une fonction
- Calcul de la valeur de la fonction en un point à partir du graphe
- Calcul du gradient à partir du graphe
- Calcul efficace du gradient à partir du graphe

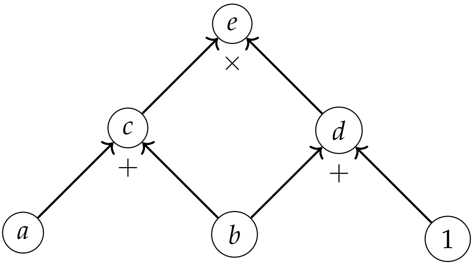
Plan

Graphe de calcul

- Soit la fonction : $f(a, b) = (a + b) \times (b + 1)$
- Elle comporte trois opérations : deux additions et une multiplication
- On introduit des **variables** pour représenter la valeur de la fonction (variable finale) et les résultats intermédiaires.
- On obtient trois équations, chacune composée d'une opération ou d'un appel de fonction :

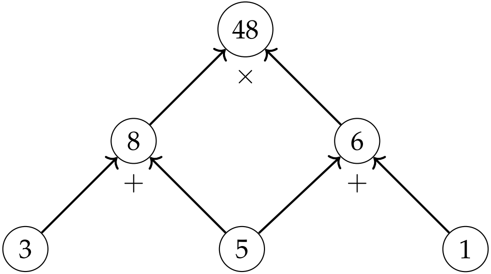
$$\begin{aligned} c &= a + b \\ d &= b + 1 \\ e &= c \times d \end{aligned}$$

Graphe de calcul



- On peut représenter les équations sous la forme d'un graphe.
- Chaque sommet du graphe correspond à une variable ou une constante.
- On représente au niveau de chaque sommet la fonction où l'opération qui lui correspond.
- On peut se servir du graphe pour réaliser le calcul, en donnant des valeurs à a et à b.

Graphe de calcul



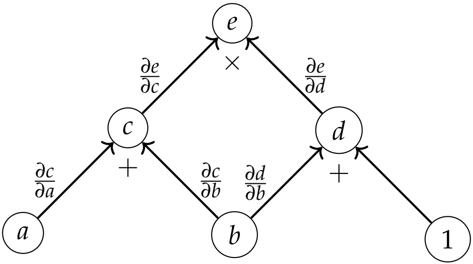
Bof !

Calcul du gradient

- On peut aussi se servir du graphe de calcul correspondant à une fonction pour calculer le gradient de la fonction.
- On associe à tout arc $x \to y$ la dérivée partielle $\frac{\partial y}{\partial x}$

$$x \xrightarrow{\frac{\partial y}{\partial x}} y$$

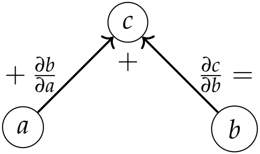
Calcul du gradient



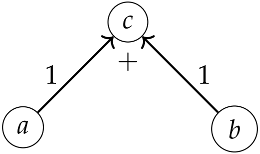
- On définit des règles spécifiques à chaque type de sommet
- qui dépendent de la fonction ou de l'opération associée au sommet.

Addition

$$\frac{\partial c}{\partial a} = \frac{\partial}{\partial a}(a+b) = \frac{\partial a}{\partial a} + \frac{\partial b}{\partial a}$$

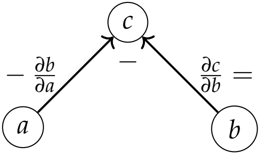


$$\frac{\partial c}{\partial b} = \frac{\partial}{\partial b}(a+b) = \frac{\partial a}{\partial b} + \frac{\partial b}{\partial b}$$

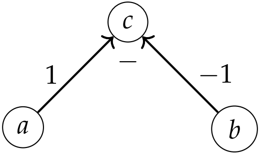


Soustraction

$$\frac{\partial c}{\partial a} = \frac{\partial}{\partial a}(a-b) = \frac{\partial a}{\partial a} - \frac{\partial b}{\partial a}$$

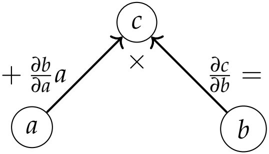


$$\frac{\partial c}{\partial b} = \frac{\partial}{\partial b}(a-b) = \frac{\partial a}{\partial b} - \frac{\partial b}{\partial b}$$

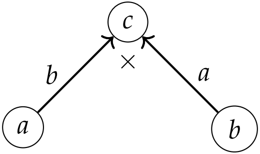


Multiplication

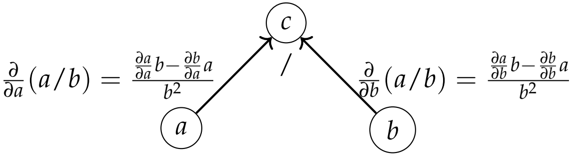
$$\frac{\partial c}{\partial a} = \frac{\partial}{\partial a}(a \times b) = \frac{\partial a}{\partial a} b + \frac{\partial b}{\partial a} a$$

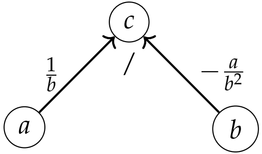


$$\frac{\partial c}{\partial b} = \frac{\partial}{\partial b}(a \times b) = \frac{\partial a}{\partial b} b + \frac{\partial b}{\partial b} a$$



Division


c
$/$
$\frac{\partial}{\partial a}(a/b) = \frac{\frac{\partial a}{\partial a}b - \frac{\partial b}{\partial a}a}{b^2}$
$\frac{\partial}{\partial b}(a/b) = \frac{\frac{\partial a}{\partial b}b - \frac{\partial b}{\partial b}a}{b^2}$
a
b


c
$/$
$\frac{1}{b}$
$-\frac{a}{b^2}$
a
b

Exponentielle

- $y = b^x$

$$x \xrightarrow{\ln(b) \times b^x} y$$

- $y = e^x$

$$x \xrightarrow{e^x} y$$

Logarithme

- $y = \log_b(x)$

$$x \xrightarrow{\frac{1}{\ln(b) \times x}} y$$

- $y = \ln(x)$

$$x \xrightarrow{\frac{1}{x}} y$$

Opposée et inverse

- $y = -x$

$$x \xrightarrow{-1} y$$

- $y = \frac{1}{x}$

$$x \xrightarrow{-\frac{1}{x^2}} y$$

Pas nécessaire mais pratique.

Composition

- $h(x) = f \circ g(x) = f(g(x))$
- introduction de variables intermédiaires

$$v = f(u)$$
$$u = g(x)$$

- Graphe correspondant

$$x \xrightarrow{\frac{\partial u}{\partial x}} u \xrightarrow{\frac{\partial v}{\partial u}} v$$

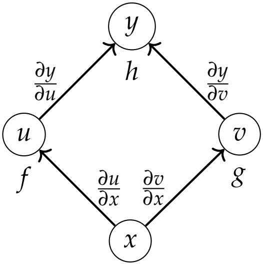
- Règle de dérivation des fonctions composées :

$$\frac{\partial v}{\partial x} = \frac{\partial v}{\partial u} \frac{\partial u}{\partial x}$$

- C'est le calcul qui est effectué si on multiplie les dérivées sur les arcs du graphe !

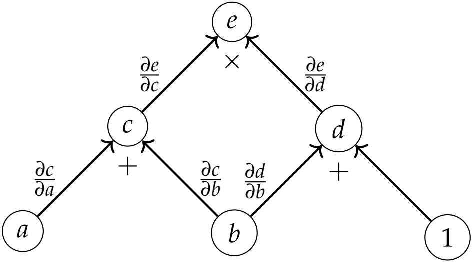
Composition

- $y = h(f(x), g(x))$



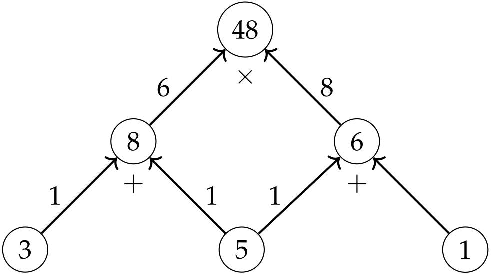
- $\frac{\partial y}{\partial x} = \frac{\partial y}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial y}{\partial v}\frac{\partial v}{\partial x}$
- On fait la somme des poids des chemins !

Calcul du gradient



1. On attribue une valeur aux variables libres.
2. On calcule la valeur des différentes variables du graphe en partant des variables libres.
3. On calcule les dérivées partielles.
4. Pour chaque variable libre, on calcule le poids de tous les chemins menant au dernier sommet

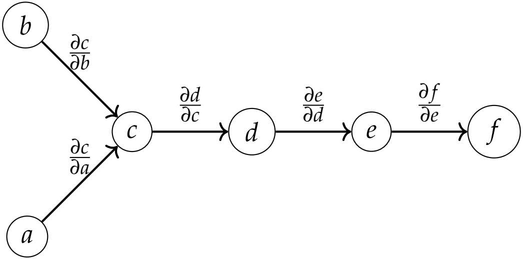
Exemple : $a = 3$ et $b = 5$



1. On attribue une valeur aux variables libres.
2. On calcule la valeur des différentes variables du graphe en partant des variables libres.
3. On calcule les dérivées partielles.
4. Pour chaque variable libre, on calcule le poids de tous les chemins menant au dernier sommet

- $\frac{\partial e}{\partial a} = 6 \times 1$
- $\frac{\partial e}{\partial b} = 6 \times 1 + 8 \times 1$
- $\nabla f(3,5) = [6,14]^T$

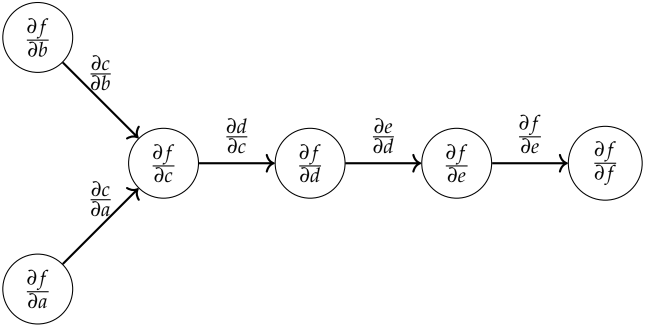
Problème d'efficacité



$$\frac{\partial f}{\partial a} = \frac{\partial c}{\partial a} \times \frac{\partial d}{\partial c} \times \frac{\partial e}{\partial d} \times \frac{\partial f}{\partial e}$$
$$\frac{\partial f}{\partial b} = \frac{\partial c}{\partial b} \times \frac{\partial d}{\partial c} \times \frac{\partial e}{\partial d} \times \frac{\partial f}{\partial e}$$

Le produit $\frac{\partial d}{\partial c} \times \frac{\partial e}{\partial d} \times \frac{\partial f}{\partial e}$ est répété deux fois

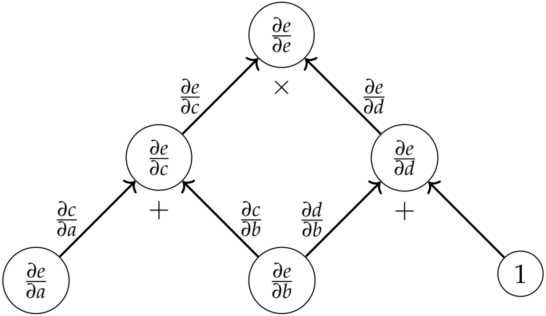
Problème d'efficacité



$$
\begin{array}{ll}
\frac{\partial f}{\partial e} = \frac{\partial f}{\partial e} \times \frac{\partial f}{\partial f} & \frac{\partial f}{\partial d} = \frac{\partial e}{\partial d} \times \frac{\partial f}{\partial e} \\
\frac{\partial f}{\partial c} = \frac{\partial d}{\partial c} \times \frac{\partial f}{\partial d} & \frac{\partial f}{\partial a} = \frac{\partial c}{\partial a} \times \frac{\partial f}{\partial a} \\
\frac{\partial f}{\partial b} = \frac{\partial c}{\partial b} \times \frac{\partial f}{\partial b} &
\end{array}
$$

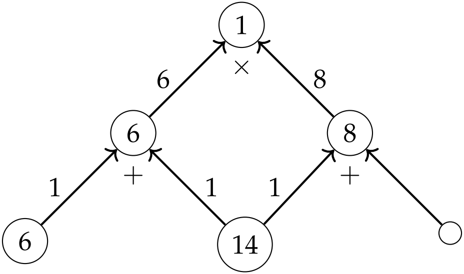
Chaque produit n'est fait qu'une fois et stocké dans un sommet du graphe.

Calcule efficace du gradient



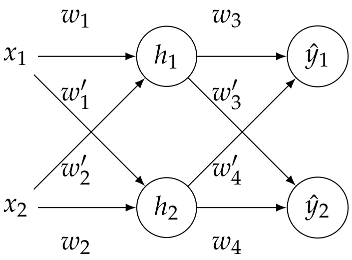
1. On attribue une valeur aux variables libres.
2. On calcule la valeur des différentes variables du graphe en partant des variables libres.
3. On calcule les dérivées partielles.
4. On calcule la dérivée partielle de la variable finale par rapport à chaque variable du graphe.

Calcule efficace du gradient



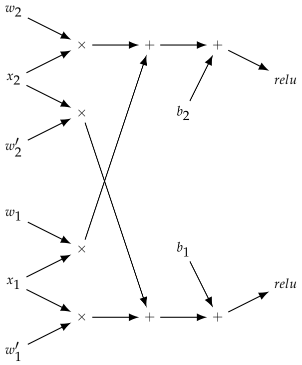
1. On attribue une valeur aux variables libres.
2. On calcule la valeur des différentes variables du graphe en partant des variables libres.
3. On calcule les dérivées partielles.
4. On calcule la dérivée partielle de la variable finale par rapport à chaque variable du graphe.

Un exemple plus réaliste



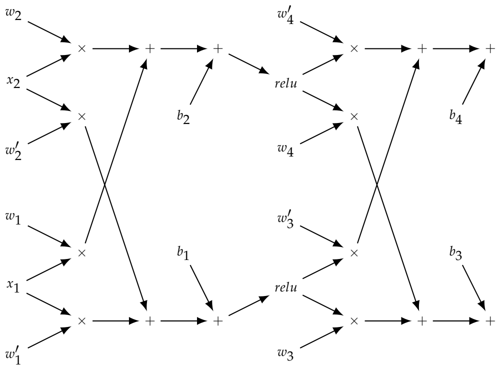
$$
\begin{aligned}
h_1 &= ReLU(w_1 x_1 + w'_2 x_2 + b_1)\\
h_2 &= ReLU(w'_1 x_1 + w_2 x_2 + b_2)\\
\hat{y}_1 &= w_3 \times h_1 + w_4 \times h_2 + b_3\\
\hat{y}_2 &= w'_3 \times h_1 + w'_4 \times h_2 + b_4
\end{aligned}
$$

Graphe de calcul - première couche



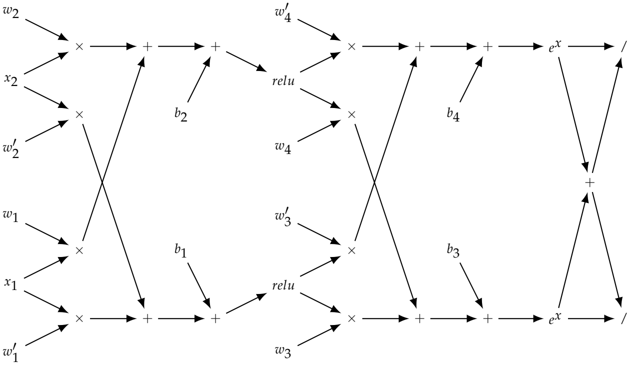
$$h_1 = ReLU(w_1 x_1 + w'_2 x_2 + b_1)$$
$$h_2 = ReLU(w'_1 x_1 + w_2 x_2 + b_2)$$

Graphe de calcul - deuxième couche



$$\hat{y}_1 = w_3 \times h_1 + w_4 \times h_2 + b_3$$
$$\hat{y}_2 = w'_3 \times h_1 + w'_4 \times h_2 + b_4$$

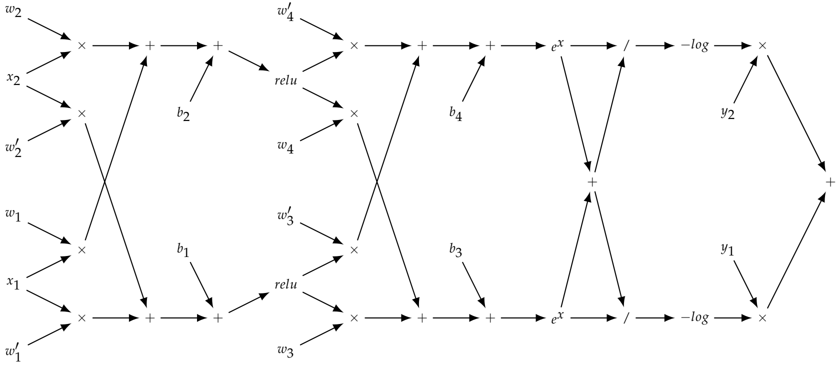
Graphe de calcul - Softmax



$$e_1 = \frac{e^{\hat{y}_1}}{e^{\hat{y}_1} + e^{\hat{y}_2}}$$

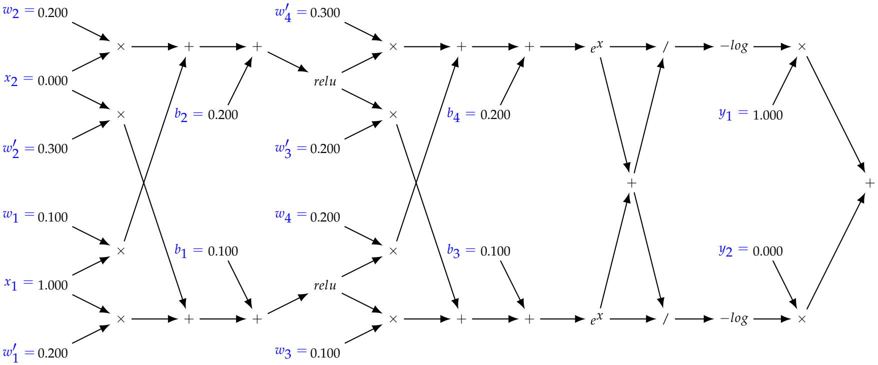
$$e_2 = \frac{e^{\hat{y}_2}}{e^{\hat{y}_1} + e^{\hat{y}_2}}$$

Graphe de calcul - entropie croisée

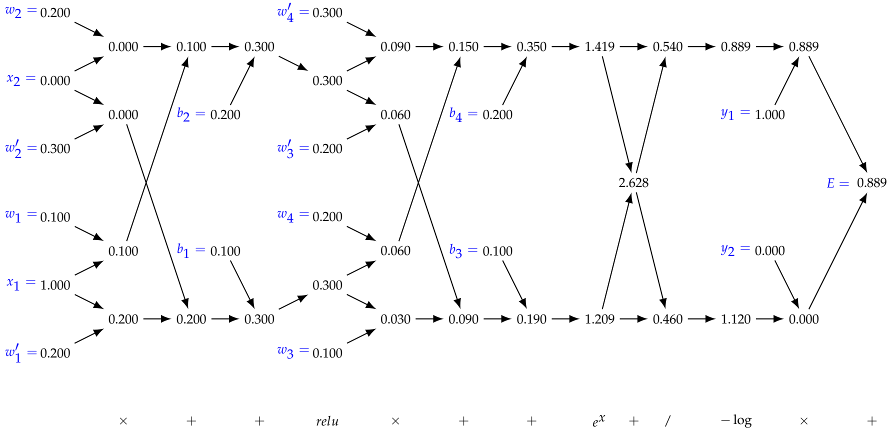


$$E = y_1 \times -\log_2(e_1) + y_2 \times -\log_2(e_2)$$

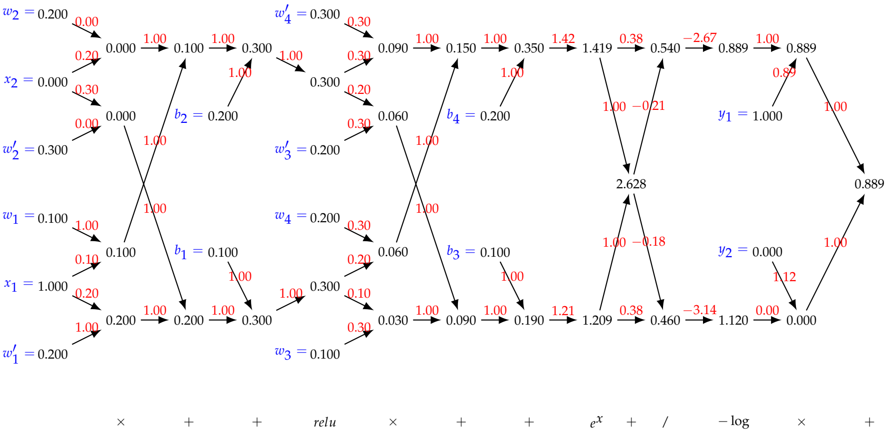
Calcul de la valeur de l’erreur



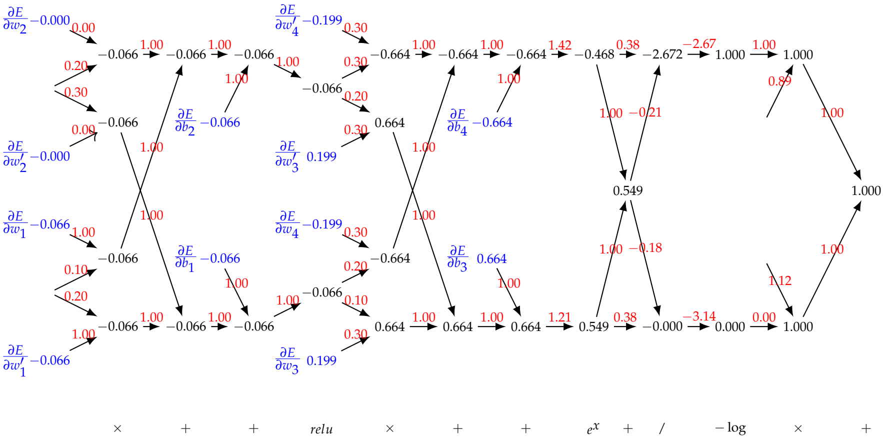
Calcul de la valeur de l'erreur



Calcul des dérivées partielles



Calcul du gradient



Sources

- Ian Goodfellow, Yoshua Bengio, Aaron Courville, *Deep Learning*, MIT Press, 2016.
- Le blog de Christopher Colah `http://colah.github.io`