Calcul du Gradient

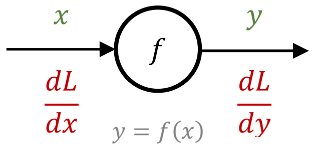
Objectif : Calculer l'influence des paramètres d'une fonction sur le résultat final

La première étape pour appliquer la méthode du gradient est tout d'abord de décomposer ce calcul en un assemblage de 'blocs' d'opérations élémentaires (somme, multiplication, fonction, …). Il devient alors possible de suivre les valeurs intermédiaires entre chaque 'sous-calcul' qui ont permis ce calcul.

Le gradient peut à présent être calculé grâce à la rétropropagation, c'est-à-dire : re-parcourir le calcul mais cette fois depuis le résultat final jusqu'aux valeurs de départ, en identifiant comment chaque 'sous-calcul' modifie l'influence sur le résultat final de chaque valeur intermédiaire.

Autrement dit, l'objectif est de calculer à chaque étape la dérivée $\frac{\partial L}{\partial x}$ où x est la valeur en entrée du bloc et L est le résultat final du calcul.

Bloc de 'sous-calcul' :



Le calcul de $\frac{\partial L}{\partial x}$ devient possible en utilisant le *'Théorème de dérivation des fonctions composées'* :

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial x}$$

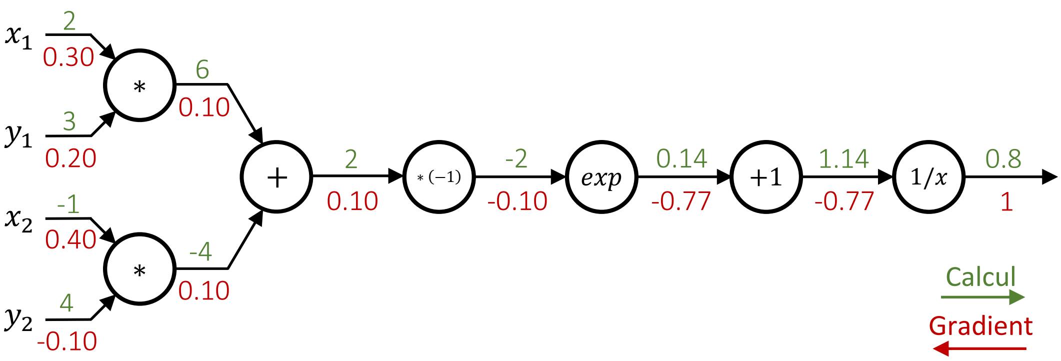
(ou *'Chain Rule'* en anglais)

En effet, $\frac{\partial L}{\partial y}$ correspond au gradient du bloc d'avant, il ne reste plus qu'à calculer $\frac{\partial y}{\partial x}$ qui correspond à la dérivée du bloc étudié : car $y = f(x)$ et donc $\frac{\partial y}{\partial x} = f'(x)$.

En calculant progressivement le gradient pour chaque 'sous-calcul', le gradient de chaque paramètre en entrée est ainsi obtenu.

La valeur du gradient calculée permet alors d'évaluer comment varierait le résultat du calcul si l'un de ses paramètres est légèrement modifié. Ainsi, le signe du gradient permet de déterminer si le résultat augmentera $(+)$ ou diminuera $(-)$ lors de l'augmentation de ce paramètre ; et plus la valeur du gradient du paramètre est grande, plus une variation de ce paramètre modifiera le résultat final.

Illustration : La fonction Sigmoïde σ avec deux paramètres : $\sigma(x, y) = \dfrac{1}{1 + e^{-(x_1 \times y_1 + x_2 \times y_2)}}$



On observe ainsi qu'augmenter la valeur de x_1, y_1 ou x_2 augmenterait le résultat du calcul, alors qu'augmenter la valeur de y_2 diminuerait le résultat.

Aussi : une petite variation de x_2 aura une plus grande influence sur le résultat final que pour tous les autres paramètres (pris indépendamment).