# Descente de gradient

1. La première partie consiste à appliquer la définition du réseau de neurones
   - On réalise le premier calcul en prenant comme vecteur d'entrée $x_0 = [123]^T$. On obtient ainsi

$$\begin{aligned} h(x_0) &= w^T \cdot x_0 \\ &= [0\ 1\ 0.5]^T \cdot [1\ 2\ 3] \\ &= 0 + 2 + 1.5 \\ &= 3.5 \end{aligned}$$

   - On refait le même calcul, mais cette fois avec la matrice $X$ lue en ligne et en indiquant l'erreur que l'on voudrait obtenir, le vecteur $y$ donné. On a ainsi

$$\begin{aligned} Errow(w) &= \frac{1}{2}((1 - [0\ 1\ 0.5]^T \cdot [1\ 1\ 2]) + (6 - [0\ 1\ 0.5]^T \cdot [1\ \ -2\ 5]) + (1 - [0\ 1\ 0.5]^T \cdot [1\ 0\ 1])) \\ &= \frac{1}{2}((1-2) + (6-0.5) + (1-0.5)) \\ &= \frac{1}{2}(-1 + 5.5 + 0.5) \\ &= 2.5 \end{aligned}$$

   - Il existe un ensemble de poids qui permettrait d'obtenir une meilleur erreur. En effet, nous avons un problème avec 3 poids (inconnues) et 3 équations ($X$ et $y$). On peut donc écrire le système à résoudre

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & -2 & 5 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 1 \end{bmatrix}$$

   La solution de ce système est

$$\begin{bmatrix} 1/3 \\ -1/2 \\ 5/6 \end{bmatrix}$$

2. On applique la définition de la dérivée partielle :

$$\begin{aligned} \frac{\partial f(x,y)}{\partial x} &= \frac{x}{\sqrt{x^2+y^2}} \\ \frac{\partial f(x,y)}{\partial y} &= \frac{y}{\sqrt{x^2+y^2}} \end{aligned}$$

3. Il suffit d'appliquer la mettre à jour $w$ avec la formule donnée. On obtient les nouveaux poids $w_1 = 2.5$, $w_= -6.5$ et $w_3 = -0.25$.
   Si on refait le calcul de l'erreur, on obtient

$$\begin{aligned} Errow(w) &= \frac{1}{2}((1 - [2.5\ \ -6.5\ \ -0.25]^T \cdot [1\ 1\ 2]) + (6 - [2.5\ \ -6.5\ \ -0.25]^T \cdot [1\ \ -2\ 5]) \\ &+ (1 - [2.5\ \ -6.5\ \ -0.25]^T \cdot [1\ 0\ 1])) \\ &= \frac{1}{2}((1+4.5) + (6-14.25) + (1-2.25)) \\ &= \frac{1}{2}(5.5 - 8.25 - 1.25) \\ &= -2 \end{aligned}$$