

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

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

- 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 $w_1 = 2.5$, $w_2 = -6.5$ et $w_3 = -0.25$.

Si on refait le calcul de l'erreur, on obtient

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

4. Dérivées partielles

Pour $f(x, y) = \sqrt{x^2 + y^2}$:

Dérivée partielle par rapport à x :

$$\begin{aligned}\frac{\partial f}{\partial x} &= \frac{\partial}{\partial x} [(x^2 + y^2)^{1/2}] \\ &= \frac{1}{2} (x^2 + y^2)^{-1/2} \cdot 2x \\ &= \boxed{\frac{x}{\sqrt{x^2 + y^2}}}\end{aligned}$$

Dérivée partielle par rapport à y :

$$\begin{aligned}\frac{\partial f}{\partial y} &= \frac{\partial}{\partial y} [(x^2 + y^2)^{1/2}] \\ &= \frac{1}{2} (x^2 + y^2)^{-1/2} \cdot 2y \\ &= \boxed{\frac{y}{\sqrt{x^2 + y^2}}}\end{aligned}$$

5. Règle d'apprentissage L1

$$\mathbf{w} \leftarrow \mathbf{w} + \eta \sum_{e \in E} \text{sign}(y_e - \mathbf{w}^T \mathbf{x}_e) \mathbf{x}_e$$

6. Itération de descente de gradient

Données : $\mathbf{w} = (0, 1, 0.5)^T$, $\eta = 0.1$

Descente de gradient L2

Pour chaque exemple, calculons l'erreur $(y_e - \mathbf{w}^T \mathbf{x}_e)$:

- Exemple 1 : $y_1 - \mathbf{w}^T \mathbf{x}_1 = 1 - 2 = -1$
- Exemple 2 : $y_2 - \mathbf{w}^T \mathbf{x}_2 = 6 - 0.5 = 5.5$
- Exemple 3 : $y_3 - \mathbf{w}^T \mathbf{x}_3 = 1 - 0.5 = 0.5$

Calcul de la mise à jour :

$$\begin{aligned}\Delta \mathbf{w} &= \eta \sum_{e \in E} (y_e - \mathbf{w}^T \mathbf{x}_e) \mathbf{x}_e \\ &= 0.1 \left[(-1) \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} + 5.5 \begin{pmatrix} 1 \\ -2 \\ 5 \end{pmatrix} + 0.5 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right] \\ &= 0.1 \left[\begin{pmatrix} -1 \\ -1 \\ -2 \end{pmatrix} + \begin{pmatrix} 5.5 \\ -11 \\ 27.5 \end{pmatrix} + \begin{pmatrix} 0.5 \\ 0 \\ 0.5 \end{pmatrix} \right] \\ &= 0.1 \begin{pmatrix} 5 \\ -12 \\ 26 \end{pmatrix} \\ &= \begin{pmatrix} 0.5 \\ -1.2 \\ 2.6 \end{pmatrix}\end{aligned}$$

Nouveaux poids :

$$\mathbf{w}_{\text{new}} = \begin{pmatrix} 0 \\ 1 \\ 0.5 \end{pmatrix} + \begin{pmatrix} 0.5 \\ -1.2 \\ 2.6 \end{pmatrix} = \boxed{\begin{pmatrix} 0.5 \\ -0.2 \\ 3.1 \end{pmatrix}}$$

Vérification de l'erreur :

Nouvelles prédictions :

- ▶ $\mathbf{w}_{\text{new}}^T \mathbf{x}_1 = 0.5 - 0.2 + 6.2 = 6.5$, erreur : $(1 - 6.5)^2 = 30.25$
- ▶ $\mathbf{w}_{\text{new}}^T \mathbf{x}_2 = 0.5 + 0.4 + 15.5 = 16.4$, erreur : $(6 - 16.4)^2 = 108.16$
- ▶ $\mathbf{w}_{\text{new}}^T \mathbf{x}_3 = 0.5 + 0 + 3.1 = 3.6$, erreur : $(1 - 3.6)^2 = 6.76$

$$Error(\mathbf{w}_{\text{new}}) = \frac{1}{2}(30.25 + 108.16 + 6.76) = \boxed{72.585}$$