Calcul du Gradient dans un Graphe et Application à la Régression Linéaire

Alexis Emanuelli

November 1, 2023

1: Fonction dans un graphe

Question 1

Pour la Question 1, nous avons utilisé la règle de la chaîne pour les dérivées.

La formule générale pour une dérivée partielle utilisant la règle de la chaîne est donnée par :

$$\frac{\partial L \circ h}{\partial x_i} = \sum_j \frac{\partial L}{\partial h_j} \frac{\partial h_j}{\partial x_i}$$

Donc :

$$\frac{\partial L \circ h}{\partial x_i}(\mathbf{x}) = \sum_j (\nabla L)_j \frac{\partial h_j}{\partial x_i}(\mathbf{x})$$

2.1 Calcul du gradient (scalaire)

Question 2

Dans un premier temps, étant donné que $y \in \mathbf{R}$, nous avons :

$$\frac{\partial L \circ \mathrm{mse}}{\partial y}(y, \hat{y}) = L'(\mathrm{mse}(y, \hat{y}(\mathbf{x}))) \cdot 2(y - \hat{y}(\mathbf{x}))$$

et :

$$\frac{\partial L \circ \mathrm{mse}}{\partial \hat{y}}(y, \hat{y}) = -L'(\mathrm{mse}(y, \hat{y}(\mathbf{x}))) \cdot 2(y - \hat{y}(\mathbf{x}))$$

Ici, $L = \mathrm{id}$, donc on simplifie comme suit :

$$\frac{\partial L \circ \mathrm{mse}}{\partial y}(y, \hat{y}) = 2(y - \hat{y}(\mathbf{x}))$$

et

$$\frac{\partial L \circ \mathrm{mse}}{\partial \hat{y}}(y, \hat{y}) = -2(y - \hat{y}(\mathbf{x}))$$

Les dérivées partielles de la fonction $L \circ f$ par rapport à x_i sont données par la relation suivante, selon 1) :

$$\frac{\partial (L \circ f)}{\partial x_i}(\mathbf{x}, \mathbf{w}, b) = \sum_j (\nabla L)_j \frac{\partial f_j}{\partial x_i}(\mathbf{x}, \mathbf{w}, b)$$

D'où, avec $L = \mathrm{mse}(y, \cdot)$ et étant donné que $\hat{y} = f$ est à valeurs dans $\mathbb{R}$:

$$\frac{\partial (L \circ f)}{\partial x_i}(\mathbf{x}, \mathbf{w}, b) = \frac{\partial (\mathrm{mse}(y, \hat{y}))}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial x_i}(\mathbf{x}, \mathbf{w}, b)$$

Comme

$$\frac{\partial \mathrm{mse}(y, \hat{y})}{\partial \hat{y}}(y, \hat{y}) = -2(y - \hat{y}(\mathbf{x}, \mathbf{w}, b))$$

et :

$$\frac{\partial \hat{y}}{\partial x_i}(\mathbf{x}, \mathbf{w}, b) = w_i$$

On obtient :

$$\frac{\partial (L \circ f)}{\partial x_i}(\mathbf{x}, \mathbf{w}, b) = -2(y - \hat{y}(\mathbf{x}, \mathbf{w}, b)) \cdot w_i$$

2.2 Calcul du gradient (matriciel)

Question 3

Considérons la fonction linéaire $Y = f(X, W, b) = XW + b$ où $X \in \mathbb{R}^{q\times n}$ sont les entrées, $W \in \mathbb{R}^{n\times p}$ est la matrice de poids, et $b \in \mathbb{R}^{1\times p}$ est le biais. La sortie Y est donc une matrice dans $\mathbb{R}^{q\times p}$.

La fonction de coût aux moindres carrés (MSE) généralisée à un lot d'exemples et plusieurs sorties est donnée par:

$$\text{mse}(\hat{Y}, Y) = \frac{1}{q}\|\hat{Y} - Y\|^2$$

1. Pour $\frac{\partial L \circ \text{mse}}{\partial Y_{ij}}$ et $\frac{\partial L \circ \text{mse}}{\partial \hat{Y}_{ij}}$:

D'abord, trouvons les dérivées partielles de mse par rapport à Y_{ij} et $\hat{Y}_{ij}$:

$$\frac{\partial \text{mse}}{\partial Y_{ij}} = \frac{2}{q}(Y_{ij} - \hat{Y}_{ij})$$
$$\frac{\partial \text{mse}}{\partial \hat{Y}_{ij}} = \frac{-2}{q}(Y_{ij} - \hat{Y}_{ij})$$

Ensuite, utilisons la règle de la chaîne pour trouver les dérivées de $L \circ \text{mse}$ par rapport à Y_{ij} et $\hat{Y}_{ij}$:

$$\begin{aligned}\frac{\partial L \circ \text{mse}}{\partial Y_{ij}} &= \frac{\partial L}{\partial \text{mse}} \frac{\partial \text{mse}}{\partial Y_{ij}} \\ &= \frac{\partial L}{\partial \text{mse}} \cdot \frac{2}{q}(Y_{ij} - \hat{Y}_{ij})\end{aligned}$$

$$\begin{aligned}\frac{\partial L \circ \text{mse}}{\partial \hat{Y}_{ij}} &= \frac{\partial L}{\partial \text{mse}} \frac{\partial \text{mse}}{\partial \hat{Y}_{ij}} \\ &= \frac{\partial L}{\partial \text{mse}} \cdot \frac{-2}{q}(Y_{ij} - \hat{Y}_{ij})\end{aligned}$$

2. Pour les dérivées de la fonction linéaire $f = \hat{Y}$ par rapport à ses entrées:

$$\frac{\partial \hat{Y}_{ij}}{\partial \mathbf{X}_{kl}} = \mathbf{W}_{lj}\delta_{ik}$$
$$\frac{\partial \hat{Y}_{ij}}{\partial \mathbf{W}_{kl}} = \mathbf{X}_{ik}\delta_{jl}$$
$$\frac{\partial \hat{Y}_{ij}}{\partial \mathbf{b}_{l}} = \delta_{jl}$$

où δ_{ik} est le delta de Kronecker.

Ainsi, en utilisant la règle de la chaîne pour la fonction de coût L et la fonction linéaire f, nous avons:

$$\frac{\partial L \circ f}{\partial X_{kl}} = \sum_j \frac{\partial L}{\partial \hat{Y}_{ij}} W_{lj}\delta_{ik}$$
$$\frac{\partial L \circ f}{\partial W_{kl}} = \sum_i X_{ik} \frac{\partial L}{\partial \hat{Y}_{ij}} \delta_{jl}$$
$$\frac{\partial L \circ f}{\partial b_{l}} = \sum_i \frac{\partial L}{\partial \hat{Y}_{ij}} \delta_{jl}$$

Question 4

$$\frac{\partial \mathrm{mse}(\mathbf{Y}, \hat{\mathbf{Y}})}{\partial \mathbf{Y}} = \frac{2}{q}(\mathbf{Y} - \hat{\mathbf{Y}})$$
$$\frac{\partial \mathrm{mse}(\mathbf{Y}, \hat{\mathbf{Y}})}{\partial \hat{\mathbf{Y}}} = \frac{-2}{q}(\mathbf{Y} - \hat{\mathbf{Y}})$$

Pour les dérivées de la fonction linéaire f par rapport à ses entrées:

$$\frac{\partial \hat{\mathbf{Y}}}{\partial \mathbf{X}} = \mathbf{W}$$
$$\frac{\partial \hat{\mathbf{Y}}}{\partial \mathbf{W}} = \mathbf{X}$$
$$\frac{\partial \hat{\mathbf{Y}}}{\partial \mathbf{b}} = I_{q \times p} \text{ (la matrice identité de taille } q \times p)$$

Maintenant, en utilisant la règle de la chaîne, exprimons ces dérivées en notation matricielle pour le coût L et la fonction f:

$$\begin{aligned}
\frac{\partial L \circ \text{mse}(\mathbf{Y}, \hat{\mathbf{Y}})}{\partial \mathbf{Y}} &= \frac{\partial L}{\partial \text{mse}} \cdot \frac{2}{q}(\mathbf{Y} - \hat{\mathbf{Y}}) \\
\frac{\partial L \circ \text{mse}(\mathbf{Y}, \hat{\mathbf{Y}})}{\partial \hat{\mathbf{Y}}} &= \frac{\partial L}{\partial \text{mse}} \cdot \frac{-2}{q}(\mathbf{Y} - \hat{\mathbf{Y}})
\end{aligned}$$

Et,

$$\begin{aligned}
\frac{\partial L \circ f(\mathbf{X}, \mathbf{W}, \mathbf{b})}{\partial \mathbf{X}} &= \frac{\partial L}{\partial \hat{\mathbf{Y}}} \cdot \mathbf{W}^T \\
\frac{\partial L \circ f(\mathbf{X}, \mathbf{W}, \mathbf{b})}{\partial \mathbf{W}} &= \mathbf{X}^T \cdot \frac{\partial L}{\partial \hat{\mathbf{Y}}} \\
\frac{\partial L \circ f(\mathbf{X}, \mathbf{W}, \mathbf{b})}{\partial \mathbf{b}} &= \frac{\partial L}{\partial \hat{\mathbf{Y}}}
\end{aligned}$$

2.3 Calcul du gradient de C

Question 5

On a $\mathbf{C} = \text{MSE}_{\mathbf{Y}} \circ f$

En utilisant les dérivées calculées précédemment, les gradients de $\mathbf{C}$ par rapport à $\mathbf{W}$, et $\mathbf{b}$ sont les suivants:

$$\frac{\partial \mathbf{C}}{\partial \mathbf{W}} = \frac{\partial (\text{MSE}_{\mathbf{Y}} \circ f)}{\partial \mathbf{W}} = \frac{\partial \text{MSE}_{\mathbf{Y}}}{\partial \hat{\mathbf{Y}}} \cdot \frac{\partial \hat{\mathbf{Y}}}{\partial \mathbf{W}} = -\frac{2}{q}(\mathbf{Y} - \hat{\mathbf{Y}}) \cdot \mathbf{X}^T$$

$$\frac{\partial \mathbf{C}}{\partial \mathbf{b}} = \frac{\partial (\text{MSE}_{\mathbf{Y}} \circ f)}{\partial \mathbf{b}} = \frac{\partial \text{MSE}_{\mathbf{Y}}}{\partial \hat{\mathbf{Y}}} \cdot \frac{\partial \hat{\mathbf{Y}}}{\partial \mathbf{b}} = -\frac{2}{q}(\mathbf{Y} - \hat{\mathbf{Y}})$$