

Feuille d'exercices : Dérivées partielles et gradients

Fonctions d'activation, projections linéaires, compositions, exponentielles et logarithmes

Mathématiques pour le Deep Learning & Optimisation

Notations usuelles :

- Sigmoid standard : $\sigma(z) = \frac{1}{1+e^{-z}}$.
 - Tangente hyperbolique : $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$.
 - Rectified Linear Unit : $\text{ReLU}(z) = \max(0, z)$.
 - Logarithme naturel : $\ln(z)$ ou $\log(z)$.
 - Pour $x \in \mathbb{R}^n$, $\nabla_x f(x) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)^T$.
-

Partie 1 : Dérivées partielles scalaires et compositions simples (1 à 20)

Pour chaque exercice, calculez les dérivées partielles demandées.

1. Soit $f(x, y) = 3x^2y - 4y^3 + 5x$. Calculez $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$.
2. Soit $f(x, y) = e^{2x+3y}$. Calculez $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$.
3. Soit $f(x, y) = \ln(x^2 + y^2)$ pour $(x, y) \neq (0, 0)$. Calculez $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$.
4. Soit $\sigma(z) = \frac{1}{1+e^{-z}}$. Exprimez la dérivée $\sigma'(z)$ uniquement en fonction de $\sigma(z)$.
5. Soit $f(w, x, b) = \sigma(wx + b)$ où $w, x, b \in \mathbb{R}$. Calculez $\frac{\partial f}{\partial w}$, $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial b}$ en fonction de f .
6. Soit $g(z) = \tanh(z)$. Exprimez la dérivée $g'(z)$ uniquement en fonction de $\tanh(z)$.
7. Soit $f(w, x, b) = \tanh(wx + b)$. Calculez $\frac{\partial f}{\partial w}$ et $\frac{\partial f}{\partial b}$ en fonction de f .
8. Soit $h(z) = \text{ReLU}(z) = \max(0, z)$. Donnez la dérivée (au sens faible / sous-gradient) $h'(z)$ pour $z \neq 0$.
9. Soit $f(w, x, b) = \text{ReLU}(wx + b)$. Calculez $\frac{\partial f}{\partial w}$ pour $wx + b \neq 0$.
10. Soit $f(x, y) = x \ln(y) + ye^x$. Calculez $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$.
11. Soit $L(y, \hat{y}) = -(y \ln(\hat{y}) + (1 - y) \ln(1 - \hat{y}))$ la perte logistique binaire. Calculez $\frac{\partial L}{\partial \hat{y}}$.
12. Soit $L(w) = -[y \ln(\sigma(wx)) + (1 - y) \ln(1 - \sigma(wx))]$. En utilisant la règle de dérivation en chaîne, calculez $\frac{\partial L}{\partial w}$.
13. Soit $f(x, y) = \ln(1 + e^{x+y})$ (fonction Softplus à deux variables). Calculez $\frac{\partial f}{\partial x}$ et exprimez-la à l'aide de la fonction sigmoïde.
14. Soit $f(x, y) = \tanh(x^2y)$. Calculez $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$.
15. Soit $f(x, y) = \frac{x}{x+y}$ pour $x + y \neq 0$. Calculez $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$.
16. Soit $f(x_1, x_2) = \exp\left(-\frac{x_1^2 + x_2^2}{2\sigma^2}\right)$ (noyau gaussien). Calculez $\frac{\partial f}{\partial x_1}$.
17. Soit $f(x, y) = \text{ReLU}(x^2 - y)$. Déterminez $\frac{\partial f}{\partial x}$ pour les points où $x^2 - y > 0$.

18. Soit $f(x, y) = \sigma(x) \cdot \sigma(y)$. Calculez $\frac{\partial f}{\partial x}$ en factorisant par $f(x, y)$.
 19. Soit $f(x, y) = \ln(\sigma(x + 2y))$. Calculez $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$ en fonction de $\sigma(x + 2y)$.
 20. Soit $f(x, y) = \frac{e^x}{e^x + e^y}$. Exprimez $f(x, y)$ sous forme d'une sigmoïde et donnez $\frac{\partial f}{\partial x}$.
-

Partie 2 : Gradients vectoriels, projections linéaires et compositions (21 à 40)

Pour chaque exercice, déterminez le gradient ou la dérivée partielle par rapport aux vecteurs ou matrices indiqués.

21. Soit $f(x) = w^T x$ où $w, x \in \mathbb{R}^n$. Calculez le gradient $\nabla_x f(x)$.
22. Soit $f(x) = x^T A x$ où $A \in \mathbb{R}^{n \times n}$ est une matrice symétrique et $x \in \mathbb{R}^n$. Calculez $\nabla_x f(x)$.
23. Soit $f(x) = x^T A x$ où $A \in \mathbb{R}^{n \times n}$ n'est pas supposée symétrique. Calculez $\nabla_x f(x)$.
24. Soit $f(x) = \|x\|_2^2 = x^T x$ pour $x \in \mathbb{R}^n$. Calculez $\nabla_x f(x)$.
25. Soit $f(x) = \|Ax - b\|_2^2$ avec $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$ et $b \in \mathbb{R}^m$. Calculez $\nabla_x f(x)$.
26. Soit $f(w) = \sigma(w^T x + b)$ où $w, x \in \mathbb{R}^n$ et $b \in \mathbb{R}$. Calculez le gradient $\nabla_w f(w)$.
27. Soit $f(w) = \tanh(w^T x + b)$ où $w, x \in \mathbb{R}^n$ et $b \in \mathbb{R}$. Calculez le gradient $\nabla_w f(w)$ en fonction de $f(w)$.
28. Soit $f(x) = \ln(1 + e^{w^T x})$ avec $w, x \in \mathbb{R}^n$. Calculez $\nabla_x f(x)$ en utilisant la fonction sigmoïde.
29. Soit $L(w) = -[y \ln(\sigma(w^T x)) + (1 - y) \ln(1 - \sigma(w^T x))]$ avec $y \in \{0, 1\}$. Calculez $\nabla_w L(w)$.
30. Soit $f(W) = \text{Tr}(W^T X)$ où $W, X \in \mathbb{R}^{m \times n}$. Calculez la dérivée matricielle $\nabla_W f(W)$.
31. Soit $f(W) = \frac{1}{2} \|WX - Y\|_F^2$ où $W \in \mathbb{R}^{m \times d}$, $X \in \mathbb{R}^{d \times n}$, $Y \in \mathbb{R}^{m \times n}$ et $\|\cdot\|_F$ est la norme de Frobenius. Calculez $\nabla_W f(W)$.
32. Soit $f(x) = u^T \text{ReLU}(Wx + b)$ où $W \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, $u \in \mathbb{R}^m$, et la fonction ReLU s'applique composante par composante. Soit $z = Wx + b$. Donnez l'expression de $\nabla_x f(x)$ à l'aide d'une matrice diagonale d'activation.
33. Soit $f(x) = \ln(\sum_{i=1}^n e^{x_i})$ (la fonction LogSumExp). Calculez la k -ième composante $\frac{\partial f}{\partial x_k}$ et identifiez la fonction obtenue.
34. Soit $f(x) = \exp(u^T x)$ avec $u, x \in \mathbb{R}^n$. Calculez $\nabla_x f(x)$.
35. Soit $f(x) = \frac{1}{2} x^T A x - b^T x$ où A est une matrice symétrique inversible. Quelle est la valeur de x qui annule le gradient $\nabla_x f(x)$?
36. Soit $f(w, b) = \frac{1}{2m} \sum_{i=1}^m (w^T x^{(i)} + b - y^{(i)})^2$. Calculez $\nabla_w f(w, b)$ et $\frac{\partial f}{\partial b}$.
37. Soit $f(x) = \|x\|_2 = \sqrt{x^T x}$ pour $x \neq 0$. Calculez $\nabla_x f(x)$.
38. Soit $f(x) = \sigma(Ax)$ où $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, et σ s'applique terme à terme. Quelle est la matrice jacobienne $J_f(x) \in \mathbb{R}^{m \times n}$?
39. Soit la fonction résiduelle $f(x) = x + W_2 \text{ReLU}(W_1 x)$ où $x \in \mathbb{R}^n$, $W_1 \in \mathbb{R}^{h \times n}$ et $W_2 \in \mathbb{R}^{n \times h}$. Donnez l'expression de la matrice jacobienne $J_f(x)$.
40. Soit la fonction de perte avec régularisation $L(w) = E(w) + \frac{\lambda}{2} \|w\|_2^2$. Exprimez $\nabla_w L(w)$ en fonction de $\nabla_w E(w)$.

Solutions détaillées

Partie 1 : Solutions des exercices scalaires

1. $\frac{\partial f}{\partial x} = 6xy + 5$ et $\frac{\partial f}{\partial y} = 3x^2 - 12y^2$.
2. $\frac{\partial f}{\partial x} = 2e^{2x+3y}$ et $\frac{\partial f}{\partial y} = 3e^{2x+3y}$.
3. $\frac{\partial f}{\partial x} = \frac{2x}{x^2+y^2}$ et $\frac{\partial f}{\partial y} = \frac{2y}{x^2+y^2}$.
4. $\sigma'(z) = \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} \cdot \frac{e^{-z}}{1+e^{-z}} = \sigma(\mathbf{z})(1 - \sigma(\mathbf{z}))$.
5. Par règle de dérivation en chaîne :
$$\frac{\partial f}{\partial w} = xf(1-f), \quad \frac{\partial f}{\partial x} = wf(1-f), \quad \frac{\partial f}{\partial b} = f(1-f).$$
6. $g'(z) = 1 - \tanh^2(z) = \mathbf{1} - \mathbf{g}(\mathbf{z})^2$.
7. $\frac{\partial f}{\partial w} = x(1-f^2)$ et $\frac{\partial f}{\partial b} = 1-f^2$.
8. $h'(z) = \mathbf{1}$ si $z > 0$, et $\mathbf{0}$ si $z < 0$ (en $z = 0$, le sous-gradient est l'intervalle $[0, 1]$).
9. $\frac{\partial f}{\partial w} = x \cdot \mathbf{1}_{\{wx+b>0\}}$ (vaut x si $wx + b > 0$, et 0 sinon).
10. $\frac{\partial f}{\partial x} = \ln(y) + ye^x$ et $\frac{\partial f}{\partial y} = \frac{x}{y} + e^x$.
11. $\frac{\partial L}{\partial \hat{y}} = -\frac{y}{\hat{y}} + \frac{1-y}{1-\hat{y}} = \frac{\hat{\mathbf{y}}-\mathbf{y}}{\hat{\mathbf{y}}(1-\hat{\mathbf{y}})}$.
12. En posant $\hat{y} = \sigma(wx)$, on a $\frac{\partial L}{\partial w} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial w} = \frac{\hat{y}-y}{\hat{y}(1-\hat{y})} \cdot x\hat{y}(1-\hat{y}) = (\sigma(\mathbf{w}\mathbf{x}) - \mathbf{y})\mathbf{x}$.
13. $\frac{\partial f}{\partial x} = \frac{e^{x+y}}{1+e^{x+y}} = \frac{1}{1+e^{-(x+y)}} = \sigma(\mathbf{x} + \mathbf{y})$.
14. $\frac{\partial f}{\partial x} = 2xy(1 - \tanh^2(x^2y))$ et $\frac{\partial f}{\partial y} = x^2(1 - \tanh^2(x^2y))$.
15. $\frac{\partial f}{\partial x} = \frac{1(x+y)-x}{(x+y)^2} = \frac{\mathbf{y}}{(\mathbf{x}+\mathbf{y})^2}$ et $\frac{\partial f}{\partial y} = -\frac{\mathbf{x}}{(\mathbf{x}+\mathbf{y})^2}$.
16. $\frac{\partial f}{\partial x_1} = -\frac{x_1}{\sigma^2} \exp\left(-\frac{x_1^2+x_2^2}{2\sigma^2}\right) = -\frac{\mathbf{x}_1}{\sigma^2} \mathbf{f}(\mathbf{x}_1, \mathbf{x}_2)$.
17. Lorsque $x^2 - y > 0$, $\text{ReLU}(x^2 - y) = x^2 - y$, donc $\frac{\partial f}{\partial x} = 2\mathbf{x}$.
18. $\frac{\partial f}{\partial x} = \sigma'(x)\sigma(y) = \sigma(x)(1 - \sigma(x))\sigma(y) = \mathbf{f}(\mathbf{x}, \mathbf{y})(1 - \sigma(\mathbf{x}))$.
19. $\frac{\partial f}{\partial x} = \frac{\sigma'(x+2y)}{\sigma(x+2y)} = \frac{\sigma(x+2y)(1-\sigma(x+2y))}{\sigma(x+2y)} = 1 - \sigma(\mathbf{x} + 2\mathbf{y})$; de même $\frac{\partial f}{\partial y} = 2(1 - \sigma(\mathbf{x} + 2\mathbf{y}))$.
20. $f(x, y) = \frac{1}{1+e^{-(x-y)}} = \sigma(x - y)$. Ainsi, $\frac{\partial f}{\partial x} = \sigma(\mathbf{x} - \mathbf{y})(1 - \sigma(\mathbf{x} - \mathbf{y}))$.

Partie 2 : Solutions des exercices vectoriels et matriciels

21. $\nabla_x(w^T x) = \mathbf{w}$.
22. Si A est symétrique ($A = A^T$), $\nabla_x(x^T Ax) = 2\mathbf{A}\mathbf{x}$.
23. Dans le cas général, $\nabla_x(x^T Ax) = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$.
24. $\nabla_x(x^T x) = 2\mathbf{x}$.
25. En développant $f(x) = x^T A^T Ax - 2b^T Ax + b^T b$, on obtient $\nabla_x f(x) = 2A^T Ax - 2A^T b = 2\mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$.
26. Soit $z = w^T x + b$. $\nabla_w f = \sigma'(z)\nabla_w z = \sigma(\mathbf{w}^T \mathbf{x} + \mathbf{b})(1 - \sigma(\mathbf{w}^T \mathbf{x} + \mathbf{b})) \mathbf{x}$.
27. $\nabla_w f(w) = (1 - \tanh^2(w^T x + b)) x = (\mathbf{1} - \mathbf{f}(\mathbf{w})^2) \mathbf{x}$.
28. Par dérivation en chaîne : $\nabla_x f(x) = \frac{e^{w^T x}}{1+e^{w^T x}} w = \sigma(\mathbf{w}^T \mathbf{x}) \mathbf{w}$.

29. Résultat fondamental de la régression logistique : $\nabla_w L(w) = (\sigma(\mathbf{w}^T \mathbf{x}) - \mathbf{y}) \mathbf{x}$.
30. Comme $\text{Tr}(W^T X) = \sum_{i,j} W_{ij} X_{ij}$, on a directement $\nabla_W f(W) = \mathbf{X}$.
31. En utilisant la différentielle ou la trace : $\nabla_W f(W) = (\mathbf{W}\mathbf{X} - \mathbf{Y})\mathbf{X}^T$.
32. Soit $D = \text{diag}(\mathbf{1}_{\{z>0\}}) \in \mathbb{R}^{m \times m}$. Alors $\nabla_x f(x) = \mathbf{W}^T \mathbf{D} \mathbf{u}$.
33. $\frac{\partial f}{\partial x_k} = \frac{e^{x_k}}{\sum_{i=1}^n e^{x_i}} = \text{softmax}(\mathbf{x})_{\mathbf{k}}$. La dérivée de LogSumExp est la fonction Softmax.
34. $\nabla_x f(x) = \exp(u^T x) \mathbf{u}$.
35. $\nabla_x f(x) = Ax - b = 0 \iff \mathbf{x} = \mathbf{A}^{-1} \mathbf{b}$.
36. $\nabla_w f(w, b) = \frac{1}{m} \sum_{i=1}^m (\mathbf{w}^T \mathbf{x}^{(i)} + \mathbf{b} - \mathbf{y}^{(i)}) \mathbf{x}^{(i)}$ et $\frac{\partial f}{\partial b} = \frac{1}{m} \sum_{i=1}^m (\mathbf{w}^T \mathbf{x}^{(i)} + \mathbf{b} - \mathbf{y}^{(i)})$.
37. $\nabla_x \|x\|_2 = \frac{1}{2\sqrt{x^T x}} \cdot 2x = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}$ (le vecteur unitaire dans la direction de x).
38. $J_f(x) = \text{diag}(\sigma'(\mathbf{A}\mathbf{x})) \mathbf{A}$, où $\text{diag}(\sigma'(Ax))$ est une matrice diagonale $m \times m$ contenant $\sigma(a_i^T x)(1 - \sigma(a_i^T x))$ sur la diagonale.
39. $J_f(x) = \mathbf{I}_n + \mathbf{W}_2 \text{diag}(\mathbf{1}_{\{\mathbf{W}_1 \mathbf{x} > 0\}}) \mathbf{W}_1$.
40. $\nabla_w L(w) = \nabla_{\mathbf{w}} \mathbf{E}(\mathbf{w}) + \lambda \mathbf{w}$ (le terme λw correspond au *weight decay*).