Exercice 1

Question 1 : Soit f une fonction de $\mathbb{R}$ dans $\mathbb{R}$

- ☐ si f est continue alors f est dérivable
- ☑ si f est dérivable alors f est continue
- ☐ si f est croissante puis décroissante, elle admet un minimum local
- ☑ si f est croissante puis décroissante, elle admet un maximum local

Question 2 : Soient A, B et C trois matrices, si je peux calculer

- ☐ $A+B+C$ alors je peux calculer ABC
- ☐ ABC alors je peux calculer $A+B+C$
- ☐ ABC alors je peux calculer CBA

Question 3 : Soit $f : \mathbb{R}^k \to \mathbb{R}^q$

- ☐ f admet 1 dérivées partielles
- ☐ f admet k dérivées partielles
- ☐ f admet q dérivées partielles
- ☑ f admet kq dérivées partielles

Question 4 : La fonction $f : (x, y, z) \mapsto Ax + y + Bz$ avec $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$, $z \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$ et $B \in \mathbb{R}^{n\times n}$

- ☐ f admet 3×1 dérivées partielles
- ☐ f admet $3 \times n$ dérivées partielles
- ☐ f admet $3n \times n^2$ dérivées partielles
- ☑ f admet $3n \times n$ dérivées partielles
- ☐ f admet $n^3 \times n$ dérivées partielles

Question 5 : La formule de la chain rule est

- ☐ $\nabla(f_n \circ \ldots \circ f_1)(x) = \nabla f_n(x) \times \ldots \nabla f_1(x)$
- ☐ $\nabla(f_n \circ \ldots \circ f_1)(x) = \nabla f_1(x) \times \ldots \nabla f_n(x)$
- ☑ $\nabla(f_n \circ \ldots \circ f_1)(x) = \nabla f_n(f_{n-1}(\circ \ldots \circ f_1(x))) \times \ldots \times \nabla f_1(x)$
- ☐ $\nabla(f_n \circ \ldots \circ f_1)(x) = \nabla f_n(f_{n-1}(x)) \times \ldots \times \nabla f_1(x)$
- ☐ $\nabla(f_n \circ \ldots \circ f_1)(x) = \nabla f_1(x) \times \ldots \times \nabla f_n(f_{n-1}(\circ \ldots \circ f_1(x)))$

Exercice 2

On utilise un modèle f pour prédire les précipitations sur Paris. Les précipitations sont représentées par un nombre réel (nombre de mm de pluie par unité de surface) par zone, sur 5 zones. Pour effectuer des prédictions, notre modèle utilise des tableurs fournis par l'OMM. Pour effectuer une prédiction, on utilise 4 valeurs issues de ces tableurs.

Question 1 : Donner les dimensions n et m de $f : \mathbb{R}^n \to \mathbb{R}^m$

Nous prenons 4 entrées ($n = 4$) et donnons un nombre par zone pour 5 zones ($m = 5$). Nous avons donc

$$f : \mathbb{R}^4 \to \mathbb{R}^5$$

Pour f, on propose de prendre la fonction $f : x \mapsto 500 \times \sigma(Wx)$ où W est une matrice et $\sigma : a \mapsto \frac{1}{1+e^{-a}}$

Question 2 : Combien de dérivées partielles f admet-elle ?

Par définition, f admet $4 \times 5 = 20$ dérivées partielles.

Question 3 : Donner la dérivée de la sigmoid σ

Comme vu en cours, la sigmoid admet pour dérivée : $\sigma' = \sigma(1 - \sigma)$. Si nous prenons une fonction de $\mathbb{R}^n$ alors la Jacobienne de la sigmoid est

$$J_\sigma(x) = \begin{pmatrix} \sigma(x_1)(1-\sigma(x_1)) & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \sigma(x_n)(1-\sigma(x_n)) \end{pmatrix}$$

Question 4 : Rappeler la formule de la chain rule

$$J_f(x) = J_{f_n}(f_{n-1}(\circ \ldots \circ f_1(x))) \times \ldots \times \nabla J_{f_1}(x)$$

Question 5 : Donner la jacobienne de f par rapport à x

Donner les composants de f

$$\begin{cases} f_1 : x \mapsto Wx f_2 : x \mapsto \sigma(x) f_3 : x \mapsto 500 \times x \end{cases}$$

On calcule chaque jacobienne

$$\begin{cases} J_{f_1} : x \mapsto W \\ J_{f_2} : x \mapsto J_\sigma(x) \\ J_{f_3} : x \mapsto \begin{pmatrix} 500 & 0 & 0 & 0 & 0 \\ 0 & 500 & 0 & 0 & 0 \\ 0 & 0 & 500 & 0 & 0 \\ 0 & 0 & 0 & 500 & 0 \\ 0 & 0 & 0 & 0 & 500 \end{pmatrix} \end{cases}$$

On applique alors la chain rule

$$J_f(x) = J_{f_n}(f_{n-1}(\circ \ldots \circ f_1(x))) \times \ldots \times \nabla J_{f_1}(x)$$

dans notre cas,

$$J_f(x) = \begin{pmatrix} 500\sigma((Wx)_1)(1-\sigma((Wx)_1)) & 0 & 0 & 0 & 0 \\ 0 & 500\sigma((Wx)_2)(1-\sigma((Wx)_2)) & 0 & 0 & 0 \\ 0 & 0 & 500\sigma((Wx)_3)(1-\sigma((Wx)_3)) & 0 & 0 \\ 0 & 0 & 0 & 500\sigma((Wx)_4)(1-\sigma((Wx)_4)) & 0 \\ 0 & 0 & 0 & 0 & 500\sigma((Wx)_5)(1-\sigma((Wx)_5)) \end{pmatrix} W$$

Question 6 : Donner la jacobienne de f par rapport à W

Le seul changement correspond au premier terme précédent

$$\begin{cases} J_{f_1} : x \mapsto \begin{pmatrix} \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix} & 0 & 0 & 0 & 0 \\ 0 & \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix} & 0 & 0 & 0 \\ 0 & 0 & \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix} & 0 & 0 \\ 0 & 0 & 0 & \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix} & 0 \\ 0 & 0 & 0 & 0 & \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix} \end{pmatrix} \\ J_{f_2} : x \mapsto J_\sigma(x) \\ J_{f_3} : x \mapsto \begin{pmatrix} 500 & 0 & 0 & 0 & 0 \\ 0 & 500 & 0 & 0 & 0 \\ 0 & 0 & 500 & 0 & 0 \\ 0 & 0 & 0 & 500 & 0 \\ 0 & 0 & 0 & 0 & 500 \end{pmatrix} \end{cases}$$

Question 7 : Peut-on appliquer la descente de gradient, ici ?

Il nous manque les valeurs pour x, W, y et une fonction de coût.

indication : l'objet de la question est de savoir s'il nous manque des informations dans l'énoncé pour la descente de gradient. Si tel est le cas, vous pouvez répondre en une phrase.

Question 8 : (bonus) compléter (si nécessaire) l'énoncé de la façon la plus simple et effectuer une étape de descente de gradient pour W et x valant 1 partout et y valant 250 partout.

Exercice 3 : Minima locaux

Nous travaillons sur un modèle de classification binaire de voiture : "neuve" / "pas neuve" (je n'ai pas mis "occasion" volontairement).

Question 1 : Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction admettant un unique minimum global. Donner un exemple d'une telle fonction.

$$f : x \mapsto x^2$$

Question 2 : Rappeler la définition d'un minimum global.

Une fonction f admet un minimum global en x si et seulement si pour tout $x' \neq x$, on a $f(x') > f(x)$

Question 3 : Si f admet un minimum global en x et que f est dérivable en tout point de $\mathbb{R}$, montrer que $f'(x) = 0$

Rappelons la définition de la dérivée de f en x :

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x)$$

Puisque f est dérivable, la limite existe et est unique. Par ailleurs, par définition d'un minimum global, pour $h \neq 0$, nous avons $f(x+h) - f(x) > 0$ et donc

$$\lim_{h \to 0} f(x+h) - f(x) > 0$$

Donc le signe de $f'(x)$ est le même que celui de h, or h tend vers 0 de la même manière en étant positif ou négatif, autrement dit

$$\lim_{h \to 0, h>0} \frac{f(x+h) - f(x)}{h} = f'(x) = \lim_{h \to 0, h<0} \frac{f(x+h) - f(x)}{h}$$

donc $0 \geq f'(x) \geq 0$ et donc $f'(x) = 0$.

Question 4 : Avec les hypothèses précédentes et en supposant que f' est dérivable en tout point de $\mathbb{R}$, montrer que $f''(x) \geq 0$.

La double dérivée est donnée par

$$\lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h} = f''(x)$$

par la question précédente,

$$\lim_{h \to 0} \frac{f'(x+h)}{h} = f''(x)$$

Supposons par l'absurde que $f''(x)$ soit négatif, alors $f'(x+h)$ est positif pour h négatif et donc f' est croissante pour des valeurs $x' < x$. Ainsi, f' est croissante et strictement positive puis vaut 0 en x, ce qui est absurde. Donc $f''(x)$ est positif.

Supposons que nous ayons atteint ce minimum global et soyons en train d'évaluer notre modèle. Nous avons $10,000$ sujets de test. Nous prédisons correctement 90% des voitures neuves, et trouvons les $4,000$ voitures non-neuves correctement avec $2,000$ faux positifs.

Question 5 : Donner le tableau de contingence

$$\begin{cases} FP = 2000 \\ TP = 3600 \\ TN = 4000 \\ FN = 400 \end{cases}$$

Question 6 : Peut-on obtenir une spécificité (TN / (TN + FP)) de 90% exactement en modifiant uniquement le nombre de faux positifs ?

Il faudrait trouver x tel que

$$\frac{4000}{4000+x} = \frac{9}{10} \Leftrightarrow 40,000 = 36,000 + 9x \Leftrightarrow x = \frac{4,000}{9} = \frac{2 \times 2 \times 2 \times 2 \times 2 \times 5 \times 5 \times 5}{3 \times 3}$$

x n'est pas un nombre entier, on ne peut donc pas donner de nouvelle valeur à FP de telle sorte à avoir exactement 90% de spécificité.

Exercice 4 : problème de symétrie

Dans les réseaux de neurones comme vu en cours, les poids W_1, W_2, ... peuvent présenter des redondances, autrement dit pour une matrice W_i deux lignes présentent exactement les même valeurs. Par exemple,

$$W_1 = \begin{pmatrix} 1 & 3 & -1 \\ 1 & 3 & -1 \\ 2 & 0 & 1 \end{pmatrix}$$

présente une redondance entre les deux premières lignes.
Question 1 : Montrer que toute symétrie est invariante par descente de gradient (autrement dit, si deux lignes sont identiques à un moment, elles resteront identiques pour toutes les étapes de descente de gradient).
Question 2 : Si une des composante de f est la fonction $x \mapsto Wx+b$, avec le vecteur biais b, montrer que les symétries dans b ne sont pas invariantes par descente de gradient.
Question 3 : Considérons le réseau $f : x \mapsto W_2 \mathrm{relu}(W_1 x + b)$ si W_1 présente une symétrie, peut-on simplifier W_1 sans changer f de telle sorte à supprimer cette symétrie ? et si oui comment ?