Vous trouverez la solution à la suite de tous les énoncés.

Exercice 1

Question 1 : Soient $A \in \mathbb{R}^{4\times n}$, $B \in \mathbb{R}^{n\times 4}$ et $C \in \mathbb{R}^{n\times n}$ trois matrices,

- ☐ le produit ACB est défini quel que soit n
- ☐ le produit BAC est défini quel que soit n
- ☐ le produit BAC pour $n = 4$
- ☐ le produit ABC pour $n = 4$

Question 2 : Soit $f : x \mapsto A\sigma(x)$ avec $A \in \mathbb{R}^{6\times 3}$

- ☐ admet 3 dérivées partielles
- ☐ admet 6 dérivées partielles
- ☐ admet 3×6 dérivées partielles
- ☐ admet $3 \times 6 \times 3$ dérivées partielles
- ☐ admet $3 \times 6 \times 6$ dérivées partielles
- ☐ admet un nombre inconnu de dérivées partielles

Question 3 : Soit $f : A \mapsto A\sigma(x)$ avec $A \in \mathbb{R}^{6\times 3}$

- ☐ admet 3 dérivées partielles
- ☐ admet 6 dérivées partielles
- ☐ admet 3×6 dérivées partielles
- ☐ admet $3 \times 6 \times 3$ dérivées partielles
- ☐ admet $3 \times 6 \times 6$ dérivées partielles
- ☐ admet un nombre inconnu de dérivées partielles

Question 4 : Quelles propriétés sont vraies ?

- ☐ $f \circ g$ est croissante si et seulement si f est croissante et g est croissante
- ☐ $f : x \mapsto Ax$, avec $A \in \mathbb{R}^{2\times 2}$ admet un minimum global
- ☐ $f : x \mapsto e^{-x^2}$ admet un maximum global
- ☐ $f : x \mapsto ax^2$ admet un minimum global

Question 5 : Soit X une variable aléatoire sur $\mathbb{R}$

- ☐ $\mathbb{E}[X] = \int_{\mathbb{R}} f(x)dx$
- ☐ $\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx$
- ☐ $\mathbb{E}[X] = \sum_{i=0}^{\mathbb{N}} \mathbb{P}(X = i)$
- ☐ $\mathbb{E}[X] = \sum_{i=0}^{\mathbb{N}} i\mathbb{P}(X = i)$

Exercice 2 : prédire la fraude

Nous allons étudier les bases du modèle derrière la détection de fraude fiscale par imagerie satellitaire. Le modèle F traite des images de résolution $R = 768 \times 768 \times 3$ et prédit un "masque" binaire, c'est-à-dire une matrice dont les valeurs sont dans $[0; 1]$ de taille 768×768.

Question 1 : Donner l'expression de F de la forme $F :? \to ?$.

Question 2 : Donner le nombre de dérivées partielles de F.

Afin d'entrainer notre modèle, nous avons besoin d'une fonction de cout $\mathcal{L}$. Proposons la fonction de cout suivante $\mathcal{L} : (M_1, M_2) = \text{mIoU}(M_1, M_2)$ ou $M_1 \in \{0; 1\}^{768\times 768}$ et $M_2 \in \{0; 1\}^{768\times 768}$. La mIoU est l'intersection des masques divisée par leur union.

Question 3 : Pour calculer la mIoU, nous allons calculer l'intersection $I : (M_1, M_2) \mapsto I(M_1, M_2) \in \mathbb{R}$ entre les deux masques. Proposer une formule pour calculer l'intersection entre deux masques binaires.

indication : par exemple, pour calculer la taille de M_1, on pourrait utiliser la fonction $f : M \mapsto \sum_i \sum_j M_{i,j}$

Question 4 : Pour calculer la mIoU, nous allons calculer l'union $U : (M_1, M_2) \mapsto U(M_1, M_2) \in \mathbb{R}$ entre les deux masques. Proposer une formule pour calculer l'union entre deux masques binaires.

À partir de maintenant, la mIoU est uniquement une fonction de M_1, autrement dit M_2 est une constante.
Question 5 : La mIoU admet combien de dérivées partielles ?
Question 6 : Calculer les dérivées partielles de la mIoU.
Les masques utilisés dans la mIoU sont binaires, or les masques prédits par notre modèle ne sont pas binaires, mais dans $[0;1]$. Pour convertir nos masques, nous allons utiliser la fonction t
Question 7 : Avec $t : M \mapsto \lfloor M \rceil$ la fonction "arrondi au plus proche". Par exemple, $t(\begin{pmatrix} 0.1 & 0.54 & 0.98 \end{pmatrix}) = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$. Donner la jacobienne de t.
Question 8 : Peut-on optimiser F par descente de gradient avec les opérations suivantes $\mathcal{L}(t(F(X)), Y)$?

Exercice 2 (autre) : un peu de probabilité

Nous nous intéressons au problème suivant : prédire le salaire d'un salarié à partir de sa grille de performance accumulée sur 10 ans. Cette grille contient trois indicateurs, chacun représenté par un nombre réel.
Question 1 : Proposer une fonction f avec des paramètres pour résoudre le problème précédent.
indication : la question suivante donne un exemple de réponse pour cette question (à ne pas copier à l'identique).
Question 2 : On utilisera la fonction $f : \begin{array}{ccc} \mathbb{R}^3 & \rightarrow & \mathbb{R} \\ x = (x_1, x_2, x_3) & \mapsto & \text{relu}(Wx) \end{array}$, donner les dimensions de W et son nombre de dérivées partielles.
Question 3 : Calculer les dérivées partielles de f par rapport à W.
Question 4 : Nous proposons la fonction de cout $\mathcal{L}(x, y) = |x - y|$. Calculer sa dérivée par rapport à x.
Question 5 : Appliquer la descente de gradient sur f avec un pas d'apprentissage $\lambda = 1$, W_1 vaut un partout, x vaut un partout et y vaut 1, sur 2 étapes.
Question 6 : Notre modèle est alors évalué sur 1200 salariés. Pour savoir si notre modèle est correct, nous considérons une prédiction comme "bonne" si elle est à moins de 100$ du salaire réel. Nous obtenons une sensibilité de 0.9, une VPP de 0.9, une VPN de 0.95. Que vaut la spécificité ?
rappel : sensibilite $= \frac{TP}{TP+FN}$, *specificite* $= \frac{TN}{TN+FP}$, $VPP = \frac{TP}{TP+FP}$ *et* $VPN = \frac{TN}{TN+FN}$

Exercice 1

Question 1 : Soient $A \in \mathbb{R}^{4\times n}$, $B \in \mathbb{R}^{n\times 4}$ et $C \in \mathbb{R}^{n\times n}$ trois matrices,

- ☑ le produit ACB est défini quel que soit n
- ☐ le produit BAC est défini quel que soit n
- ☑ le produit BAC pour $n = 4$
- ☑ le produit ABC pour $n = 4$

Question 2 : Soit $f : x \mapsto A\sigma(x)$ avec $A \in \mathbb{R}^{6\times 3}$

- ☐ admet 3 dérivées partielles
- ☐ admet 6 dérivées partielles
- ☑ admet 3×6 dérivées partielles
- ☐ admet $3 \times 6 \times 3$ dérivées partielles
- ☐ admet $3 \times 6 \times 6$ dérivées partielles
- ☐ admet un nombre inconnu de dérivées partielles

Question 3 : Soit $f : A \mapsto A\sigma(x)$ avec $A \in \mathbb{R}^{6\times 3}$

- ☐ admet 3 dérivées partielles
- ☐ admet 6 dérivées partielles
- ☐ admet 3×6 dérivées partielles
- ☐ admet $3 \times 6 \times 3$ dérivées partielles
- ☑ admet $3 \times 6 \times 6$ dérivées partielles
- ☐ admet un nombre inconnu de dérivées partielles

Question 4 : Quelles propriétés sont vraies ?

- ☐ $f \circ g$ est croissante si et seulement si f est croissante et g est croissante
- ☐ $f : x \mapsto Ax$, avec $A \in \mathbb{R}^{2\times 2}$ admet un minimum global
- ☑ $f : x \mapsto e^{-x^2}$ admet un maximum global
- ☐ $f : x \mapsto ax^2$ admet un minimum global

Question 5 : Soit X une variable aléatoire sur $\mathbb{R}$

- ☐ $\mathbb{E}[X] = \int_{\mathbb{R}} f(x)dx$
- ☑ $\mathbb{E}[X] = \int_{\mathbb{R}} xf(x)dx$
- ☐ $\mathbb{E}[X] = \sum_{i=0}^{\mathbb{N}} \mathbb{P}(X = i)$
- ☐ $\mathbb{E}[X] = \sum_{i=0}^{\mathbb{N}} i\mathbb{P}(X = i)$

Exercice 2 : prédire la fraude

Nous allons étudier les bases du modèle derrière la détection de fraude fiscale par imagerie satellitaire. Le modèle F traite des images de résolution $R = 768 \times 768 \times 3$ et prédit un "masque" binaire, c'est-à-dire une matrice dont les valeurs sont dans $[0; 1]$ de taille 768×768.

Question 1 : Donner l'expression de F de la forme $F :? \to ?$.

$$F : \mathbb{R}^R \to [0; 1]^{768\times 768}$$
$$F : \mathbb{R}^R \to \mathbb{R}^{768\times 768}$$
$$F : \mathbb{R}^{768\times 768\times 3} \to \mathbb{R}^{768\times 768}$$

Question 2 : Donner le nombre de dérivées partielles de F.

F admet R entrées et 768×768 sorties et donc $R \times (768 \times 768)$ dérivées partielles. Par ailleurs, $1,043,677,052,928$ (de tête bien évidemment).

Afin d'entrainer notre modèle, nous avons besoin d'une fonction de cout $\mathcal{L}$. Proposons la fonction de cout suivante $\mathcal{L} : (M_1, M_2) = \text{mIoU}(M_1, M_2)$ ou $M_1 \in \{0; 1\}^{768\times 768}$ et $M_2 \in \{0; 1\}^{768\times 768}$. La mIoU est l'intersection des masques divisée par leur union.

Question 3 : Pour calculer la mIoU, nous allons calculer l'intersection $I : (M_1, M_2) \mapsto I(M_1, M_2) \in \mathbb{R}$ entre les deux masques. Proposer une formule pour calculer l'intersection entre deux masques binaires.

L'intersection entre deux ensembles est le sous-ensemble des éléments appartenant aux deux ensembles. Nous pouvons donc partir sur la définition suivante

$$I(M_1, M_2) = \sum_i \sum_j 1 \text{ si element i,j dans "} M_1 \text{" et "} M_2 \text{, et 0 sinon}$$

Cette "formule" peut se simplifier très facilement, en effet, dire que l'élément i, j est dans M_1 est équivalent à dire $(M_1)_{i,j} = 1$, au contraire l'élément i, j n'est pas dans M_1 est équivalent à dire $(M_1)_{i,j} = 0$

$$I(M_1, M_2) = \sum_i \sum_j (M_1)_{i,j} (M_2)_{i,j}$$

Question 4 : Pour calculer la mIoU, nous allons calculer l'union $U : (M_1, M_2) \mapsto I(M_1, M_2) \in \mathbb{R}$ entre les deux masques. Proposer une formule pour calculer l'union entre deux masques binaires. Pour calculer, l'union, nous allons utiliser la propriété suivante taille de $U =$ taille de $M_1 +$ taille de $M_2 -$ taille de I. Donc, on obtient

$$U(M_1, M_2) = \sum_i \sum_j (M_1)_{i,j} + (M_2)_{i,j} - (M_1)_{i,j} (M_2)_{i,j}$$

À partir de maintenant, la mIoU est uniquement une fonction de M_1, autrement dit M_2 est une constante.

Question 5 : La mIoU admet combien de dérivées partielles ?

La mIoU admet 768×768 entrées et retourne un nombre. Cette fonction admet donc $768 \times 768 \times 1$ dérivées partielles.

Question 6 : Calculer les dérivées partielles de la mIoU.

Les dérivées partielles de la mIoU sont de la forme

$$\begin{aligned}
\frac{\partial \text{mIoU}}{\partial (M_1)_{i,j}}(M_1, M_2) &= \frac{\partial \frac{I(M_1,M_2)}{U(M_1,M_2)}}{\partial (M_1)_{i,j}} \\
&= \frac{\partial \frac{\sum_i \sum_j (M_1)_{i,j}(M_2)_{i,j}}{\sum_i \sum_j (M_1)_{i,j} + (M_2)_{i,j} - (M_1)_{i,j}(M_2)_{i,j}}}{\partial (M_1)_{i,j}}
\end{aligned}$$

Nous avons une dérivée de la forme $\frac{f}{g}$, nous allons donc calculer les dérivées partielles de "f" et "g".

$$\begin{cases}
\frac{\partial \text{I}}{\partial (M_1)_{i,j}}(M_1, M_2) = (M_2)_{i,j} \\
\frac{\partial \text{U}}{\partial (M_1)_{i,j}}(M_1, M_2) = 1 - (M_2)_{i,j}
\end{cases}$$

On obtient alors

$$\frac{\partial \text{mIoU}}{\partial (M_1)_{i,j}}(M_1, M_2) = \frac{(M_2)_{i,j} U(M_1, M_2) - (1 - (M_2)_{i,j}) I(M_1, M_2)}{U(M_1, M_2)^2} = \frac{(M_2)_{i,j}}{U(M_1, M_2)} - \frac{1 - (M_2)_{i,j}}{U(M_1, M_2)} \text{mIoU}(M_1, M_2)$$

Les masques utilisés dans la mIoU sont binaires, or les masques prédits par notre modèle ne sont pas binaires, mais dans $[0; 1]$. Pour convertir nos masques, nous allons utiliser la fonction t

Question 7 : Avec $t : M \mapsto \lfloor M \rceil$ la fonction "arrondi au plus proche". Par exemple, $t(\begin{pmatrix} 0.1 & 0.54 & 0.98 \end{pmatrix}) = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$. Donner la jacobienne de t.

Cette fonction admet 768×768 entrées et 768×768 sorties. Ce qui nous donne beaucoup de dérivées partielles. Calculons une dérivée partielle quelconque

$$\frac{\partial t(M)_{i,j}}{\partial M_{k,l}} = \frac{\partial \lfloor M \rceil_{i,j}}{\partial M_{k,l}} = 0$$

puisque la dérivée de la fonction arrondie est toujours zéro quand elle est définie.

Question 8 : Peut-on optimiser F par descente de gradient avec les opérations suivantes $\mathcal{L}(t(F(X)), Y)$?

À cause de t les gradients seront tous nuls, on ne peut donc pas optimiser notre modèle par descente de gradient de cette manière.

Exercice 2 (autre) : un peu de probabilité

Nous nous intéressons au problème suivant : prédire le salaire d'un salarié à partir de sa grille de performance accumulée sur 10 ans. Cette grille contient trois indicateurs, chacun représenté par un nombre réel.

Question 1 : Proposer une fonction f avec des paramètres pour résoudre le problème précédent.

$$f : \begin{array}{ccc} \mathbb{R}^3 & \to & \mathbb{R} \\ x = (x_1, x_2, x_3) & \mapsto & \mathrm{relu} W_2 \sigma(W_1 x) \end{array}$$

$$f : \begin{array}{ccc} \mathbb{R}^3 & \to & \mathbb{R} \\ x = (x_1, x_2, x_3) & \mapsto & \mathrm{relu} W_2 \sigma(W_1 x + b_1) \end{array}$$

$$f : \begin{array}{ccc} \mathbb{R}^3 & \to & \mathbb{R} \\ x = (x_1, x_2, x_3) & \mapsto & \mathrm{relu}(W_2(x + \mathrm{relu}(W_1 x))) \end{array}$$

indication : la question suivante donne un exemple de réponse pour cette question (à ne pas copier à l'identique).

Question 2 : On utilisera la fonction $f : \begin{array}{ccc} \mathbb{R}^3 & \to & \mathbb{R} \\ x = (x_1, x_2, x_3) & \mapsto & \mathrm{relu}(Wx) \end{array}$, donner les dimensions de W et son nombre de dérivées partielles.

La fonction f admet trois entrées et une sortie. Puisque $x\mathbb{R}^3$ et que nous multiplions W et x, W admet nécessairement trois colonnes. De plus, puisque $Wx \in \mathbb{R}$, W admet nécessairement une seule ligne. Donc $W \in \mathbb{R}^{1\times 3}$.

De plus, f admet 3×1 dérivées partielles.

Question 3 : Calculer les dérivées partielles de f par rapport à W.

Les dérivées partielles de f sont données par

$$\begin{aligned} \frac{\partial f(W)}{\partial W_{1,i}} &= \frac{\partial \mathrm{relu}(Wx)}{\partial W_{1,i}} \\ &= \frac{\partial \mathrm{relu}(W_{1,1}x_1 + W_{1,2}x_2 + W_{1,3}x_3)}{\partial W_{1,i}} \end{aligned}$$

Nous avons une composée de fonction.

$$\begin{cases} \frac{\partial \mathrm{relu}(x)}{\partial x} = \mathbb{1}_{x>0} \\ \frac{\partial W_{1,1}x_1 + W_{1,2}x_2 + W_{1,3}x_3}{\partial W_{1,i}} = x_i \end{cases}$$

En utilisant la chain rule, nous obtenons

$$\frac{\partial f(W)}{\partial W_{1,i}} = \mathbb{1}_{Wx>0} \times x_i$$

Question 4 : Nous proposons la fonction de cout $\mathcal{L}(x, y) = |x - y|$. Calculer sa dérivée par rapport à x.

$$\frac{\partial |x-y|}{\partial x} = \begin{cases} 1 \text{ si } x > y \\ -1 \text{ sinon} \end{cases}$$

Question 5 : Appliquer la descente de gradient sur f avec un pas d'apprentissage $\lambda = 1$, W_1 vaut un partout, x vaut un partout et y vaut un, sur 2 étapes.

Nous allons calculer les termes W_2 et W_3 par descente de gradient :

$$W_n = W_{n-1} - \lambda J_{\mathcal{L}\circ f}(W_{n-1})$$

Dans notre cas, on obtient

$$W_n = W_{n-1} - 1 \times \begin{cases} 1 \text{ si } \mathrm{relu}(W_{n-1}x) > y \\ -1 \text{ sinon} \end{cases} \quad \mathbb{1}_{W_{n-1}x>0} \times \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix}$$

Nous obtenons alors

$$\begin{aligned}
W_1 &= \begin{pmatrix}1 & 1 & 1\end{pmatrix}\\
W_2 &= W_1 - \begin{cases}1 \text{ si } \mathrm{relu}(W_1 x) > y\\ -1 \text{ sinon}\end{cases} \quad \mathbb{1}_{W_1 x > 0} \times \begin{pmatrix}x_1 & x_2 & x_3\end{pmatrix}\\
&= W_1 - \begin{cases}1 \text{ si } 6 > 1\\ -1 \text{ sinon}\end{cases} \quad \mathbb{1}_{6>0} \times \begin{pmatrix}1 & 1 & 1\end{pmatrix}\\
&= W_1 - 1 \times 1 \times \begin{pmatrix}1 & 1 & 1\end{pmatrix}\\
&= \begin{pmatrix}0 & 0 & 0\end{pmatrix}\\
W_3 &= W_2 - \begin{cases}1 \text{ si } \mathrm{relu}(W_2 x) > y\\ -1 \text{ sinon}\end{cases} \quad \mathbb{1}_{W_2 x > 0} \times \begin{pmatrix}x_1 & x_2 & x_3\end{pmatrix}\\
&= W_2 - \begin{cases}1 \text{ si } 0 > 1\\ -1 \text{ sinon}\end{cases} \quad \mathbb{1}_{0>0} \times \begin{pmatrix}1 & 1 & 1\end{pmatrix}\\
&= W_2 - (-1) \times 0 \times \begin{pmatrix}1 & 1 & 1\end{pmatrix}\\
&= \begin{pmatrix}0 & 0 & 0\end{pmatrix}
\end{aligned}$$

Question 6 : Notre modèle est alors évalué sur 1200 salariés. Pour savoir si notre modèle est correct, nous considérons une prédiction comme "bonne" si elle est à moins de 100$ du salaire réel. Nous obtenons une sensibilité de 0.9, une VPP de 0.9, une VPN de 0.95. Que vaut la spécificité ?

L'énonce nous donne le système d'équation suivant

$$\begin{cases}\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN} = 1200\\ \text{sensibilite} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} = 0.9\\ \mathrm{VPP} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} = 0.9\\ \mathrm{VPN} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FN}} = 0.95\end{cases}$$

Voici une façon de résoudre ce système

$$\begin{cases}\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN} = 1200\\ 0.1\mathrm{TP} = 0.9\mathrm{FN}\\ 0.1\mathrm{TP} = 0.9\mathrm{FP}\\ 0.05\mathrm{TN} = 0.95\mathrm{FN}\end{cases}$$

$$\begin{cases}\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN} = 1200\\ \mathrm{TP} = 9\mathrm{FN}\\ \mathrm{TP} = 9\mathrm{FP}\\ \mathrm{TN} = 19\mathrm{FN}\end{cases}$$

$$\begin{cases}9\mathrm{FN} + \mathrm{FN} + 19\mathrm{FN} + \mathrm{FN} = 1200\\ \mathrm{TP} = 9\mathrm{FN}\\ \mathrm{TP} = 9\mathrm{FP}\\ \mathrm{TN} = 19\mathrm{FN}\end{cases}$$

$$\begin{cases}\mathrm{FN} = 40\\ \mathrm{TP} = 360\\ \mathrm{FP} = 40\\ \mathrm{TN} = 760\end{cases}$$

On en déduit

$$\text{specificite} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} = \frac{760}{800} = 0.95$$