# Exercice 1

**Question 1 :** Soit $F$ une fonction

- ☐ $F$ continue et scalaire $\Rightarrow$ $F$ dérivable
- ☑ $F$ dérivable et scalaire $\Rightarrow$ $F$ continue
- ☑ $F$ différentiable $\Rightarrow$ $F$ continue
- ☐ $F$ continue $\Rightarrow$ $F$ différentiable

**Question 2 :** Soit $A$ une matrice

- ☐ $A$ nilpotente (il existe $n$ tel que $A^n = 0$) $\Rightarrow$ $A$ inversible
- ☐ $A$ nilpotente $\Rightarrow$ $A$ est carrée
- ☐ $A$ inversible $\Rightarrow$ $A$ nilpotente

**Question 3 :** La descente de gradient repose sur la formule

- ☐ $\theta \leftarrow -\theta - \mu J_{\mathcal{L}\circ F}$
- ☐ $\theta \leftarrow \theta + \mu J_{\mathcal{L}\circ F}$
- ☑ $\theta \leftarrow \theta - \mu J_{\mathcal{L}\circ F}$
- ☐ $\theta \leftarrow -\theta + \mu J_{\mathcal{L}\circ F}$

**Question 4 :** La chain rule donne

- ☐ $J_{f_n\circ\cdots\circ f_1} = J_{f_n\circ\cdots\circ f_1} \times \cdots \times J_{f_1}$
- ☐ $J_{f_n\circ\cdots\circ f_1} = J_{f_n} \times \cdots \times J_{f_1}$
- ☑ $J_{f_n\circ\cdots\circ f_1} = J_{f_n}(f_{n-1}\circ\cdots\circ f_1) \times \cdots \times J_{f_1}$
- ☐ $J_{f_n\circ\cdots\circ f_1} = J_{f_1} \times \cdots \times J_{f_n}(f_{n-1}\circ\cdots\circ f_1)$

**Question 5 :** Soit $A$ une matrice carrée

- ☑ si $A$ est inversible alors le système $AX = B$ admet une solution
- ☐ si $A$ est inversible alors le système $AX = B$ admet aucune solution
- ☑ si $A$ n'est pas inversible alors le système $AX = B$ admet une solution
- ☐ si $A$ n'est pas inversible alors le système $AX = B$ admet aucune solution

# Exercice 2

Soient $F : X \mapsto AX$ et $\mathcal{L}(Y, X) \mapsto -\sum_{i=1}^{n} y_i \ln(x_i) + (1-y_i)\ln(1-x_i)$.

**1. Nous posons $X$ un vecteur de taille 2 et $F(X)$ est un scalaire. Donner les dimensions de $A$**

$A \in \mathbb{R}^{1\times 2}$

**2. Donner la Jacobienne de $F$ par rapport à X et par rapport à $A$**

Par rapport à $X$, $J_F = A$ et par rapport à $A$, $J_F = X^T$.

**3. Supposons que $X$ et $Y$ sont des scalaires. Donner la Jacobienne de $\mathcal{L}$ par rapport à $X$.**

La fonction $\mathcal{L}$ devient

$$\mathcal{L}(y, x) = -(y\ln(x) + (1-y)\ln(1-x))$$

et donc

$$\mathcal{L}'(y, x) = -\frac{y}{x} - \frac{1-y}{1-x} = -\frac{y + x - 2xy}{x(1-x)}$$

**4. Supposons que $X = \begin{pmatrix}1 & 0 & 0\end{pmatrix}^T$ $A = \begin{pmatrix}0.5 & 1 & 1\end{pmatrix}$ et $Y = 1$, Appliquer la descente de gradient une fois avec $\mu = 0.25$ pour optimiser $A$ (on dérivera donc par rapport à $A$**

Nous devons calculer la Jacobienne $J_{\mathcal{L}\circ F}$.

$$\begin{cases} J_{\mathcal{L}\circ F}(X) & = -\frac{y+AX-2AXy}{AX(1-AX)}X^T = -\frac{1+0.5-1}{0.5(1-0.5)}X^T = -2X^T \\ A & \leftarrow A - \mu J_{\mathcal{L}\circ F}(X) \end{cases}$$

donc la nouvelle valeur pour $A$ est $\begin{pmatrix}1 & 1 & 1\end{pmatrix}$.

**5. Montrer qu'avec la nouvelle valeur pour $A$, la fonction $F$ donne la bonne prédiction.**

On veut $F(X) \approx Y$. Or avec la nouvelle valeur de $A$, $F(X) = 1 = Y$.

Supposons maintenant que nous disposons un autre Modèle $G : X \mapsto BX$ avec $B$ une matrice carrée. Nous savons pour $BX = \begin{pmatrix}2\\0\\1\end{pmatrix}$ et $B = \begin{pmatrix}1 & -1 & 0\\0 & 1 & 1\\0 & 1 & -1\end{pmatrix}$.

**6. Inverser $B$ et donner la valeur de $X$**

Appliquons les pivots de Gauss

$$\begin{pmatrix}1 & -1 & 0\\0 & 1 & 1\\0 & 1 & -1\end{pmatrix}\Bigg|\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{pmatrix}$$

$$\begin{pmatrix}1 & -1 & 0\\0 & 2 & 0\\0 & 1 & -1\end{pmatrix}\Bigg|\begin{pmatrix}1 & 0 & 0\\0 & 1 & 1\\0 & 0 & 1\end{pmatrix}$$

$$\begin{pmatrix}1 & -1 & 0\\0 & 1 & 0\\0 & 1 & -1\end{pmatrix}\Bigg|\begin{pmatrix}1 & 0 & 0\\0 & 1/2 & 1/2\\0 & 0 & 1\end{pmatrix}$$

$$\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 1 & -1\end{pmatrix}\Bigg|\begin{pmatrix}1 & 1/2 & 1/2\\0 & 1/2 & 1/2\\0 & 0 & 1\end{pmatrix}$$

$$\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & -1\end{pmatrix}\Bigg|\begin{pmatrix}1 & 1/2 & 1/2\\0 & 1/2 & 1/2\\0 & -1/2 & 1/2\end{pmatrix}$$

$$\begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{pmatrix}\Bigg|\begin{pmatrix}1 & 1/2 & 1/2\\0 & 1/2 & 1/2\\0 & 1/2 & -1/2\end{pmatrix}$$

Donc l'inverse de $B$ est

$$B^{-1} = \frac{1}{2}\begin{pmatrix}2 & 1 & 1\\0 & 1 & 1\\0 & 1 & -1\end{pmatrix}$$

Calculons donc $B^{-1}BX$

$$B^{-1}BX = \frac{1}{2}\begin{pmatrix}2 & 1 & 1\\0 & 1 & 1\\0 & 1 & -1\end{pmatrix}\begin{pmatrix}2\\0\\1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}5\\1\\-1\end{pmatrix}$$

## Exercice 3 : Densité des Matrices Inversibles

Le but de cet exercice est de prouver que l'ensemble des matrices inversibles est dense. Pour cela, on va prouver que pour chaque matrice non-inversible il existe une autre matrice inversible aussi proche que l'on veut de la première. Pour cela nous allons définir une distance sur l'espace des matrices carrée de taille $n \times n$.

$$d(M_1, M_2) = \max_{i,j}\{(|M_1 - M_2|)_{i,j}\}$$

Ce qui correspond, en français, à l'écart maximal entre $M_1$ et $M_2$. Pour rappel une distance est une application telle que $d(M_1, M_2) = d(M_2, M_1)$, $d(M_1, M_2) = 0$ si et seulement si $M_1 = M_2$ et $d(M_1, M_3) \leq d(M_1, M_2) + d(M_2, M_3)$.

**1. vérifier que $d$ est une distance**

Vérifions les trois propriétés. Dans l'ordre de l'énonce, soient $M_1$ et $M_2$ deux matrices quelconques de taille $n \times n$. Alors

$$d(M_1, M_2) = \max_{i,j}\{(|M_1 - M_2|)_{i,j}\} = \max_{i,j}\{(|M_2 - M_1|)_{i,j}\} = d(M_2, M_1)$$

On a donc vérifié la première propriété. Ensuite,

$$d(M_1, M_2) = 0 \Leftrightarrow \max_{i,j}\{(|M_1 - M_2|)_{i,j}\} = 0 \Leftrightarrow \forall i, j \quad (|M_1 - M_2|)_{i,j} = 0 \Leftrightarrow M_1 = M_2$$

Et enfin,

$$d(M_1, M_3) = \max_{i,j}\{(|M_1 - M_3|)_{i,j}\} \leq \max_{i,j}\{(|M_1 - M_2| + |M_2 - M_3|)_{i,j}\}$$
$$\leq \max_{i,j}\{(|M_1 - M_2|)_{i,j}\} + \max_{i,j}\{(|M_2 - M_3|)_{i,j}\} = d(M_1, M_2) + d(M_2, M_3)$$

$d$ est donc bien une distance.

**2. Soit la suite $M_k = M + \frac{1}{n}\mathrm{Id}$, montrer que la suite $M_k$ converge vers $M$**

Il suffit de montrer que

$$\lim_{n\to\infty} d(M, M_k) = 0$$

En effet,

$$\begin{aligned} d(M, M_k) &= \max_{i,j}\{(|M - M_k|)_{i,j}\} \\ &= \max_{i,j}\{(|M - M - \frac{1}{n}\mathrm{Id}|)_{i,j}\} \\ &= \max_{i,j}\{(|\frac{1}{n}\mathrm{Id}|)_{i,j}\} \\ &= \frac{1}{n} \rightarrow_{n\to\infty} 0 \end{aligned}$$

**3. Rappeler le critère d'inversibilité d'une matrice par rapport au déterminant.**

Une matrice est inversible si et seulement si

$$\det(A) \neq 0$$

La définition générale du déterminant d'une matrice $A$ est récursive : si $A$ est de taille 1 alors $\det(A) = A_{1,1}$ et sinon

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} A_{1,j} \det(A_{[1,j]})$$

où $A_{[1,j]}$ désigne la matrice $A$ privée de la première ligne et $j^{\text{th}}$ colonne. Cette matrice est donc de taille $(n-1)\times(n-1)$.

**4. Montrer que la fonction $x \mapsto \det(M - x\mathrm{Id})$ est un polynôme de degré $n$ en $x$**

Là on va devoir s'accrocher un petit coup... Nous allons remplacer $A$ dans la définition du déterminant par $M - x\mathrm{Id}$, on obtient alors

$$\det(M - x\mathrm{Id}) = \sum_{j=1}^{n} (-1)^{1+j} (M - x\mathrm{Id})_{1,j} \det((M - x\mathrm{Id})_{[1,j]})$$

Or le terme $(M - x\mathrm{Id})_{1,j}$ est un polynôme de degré au plus 1 en $x$. En effet, il s'agit d'une coordonnée de la matrice $M - x\mathrm{Id}$, donc c'est un terme qui contient au plus $x$ et ne contient pas de termes en $x^m$ pour $m \geq 2$. Ainsi le déterminant est un polynôme de degré :

$$\text{degré}(\det(\text{matrice de taille } n \times n)) = \text{degré}(\det(\text{matrice de taille } (n-1) \times (n-1))) + 1$$

Ainsi, on en déduite que le déterminant d'une matrice $M - x\mathrm{Id}$ est un polynôme de degré au plus $n$ en $x$.

**5. Montrer qu'une infinité de terme de la suite $M_k$ sont inversibles**

Puisque de le déterminant de $M_k$ est un polynôme de degré $n$, il admet au plus $n$ racines et donc ne s'annule qu'au plus $n$ fois. Donc il y a au plus $n$ termes de déterminant nul dans la suite $M_k$. En conséquence il y a une infinité de termes inversibles

**6. Conclure**

En recoupant les résultats précédents, il existe une suite dans la suite $M_k$ de matrices inversibles qui convergent vers $M$ ce qui conclut l'exercice.

## Exercice 4 : Théorème d'Approximation Universelle

Montrons le théorème d'approximation universelle mentionnée en cours. Considérons un réseau de neurones artificiels de la forme suivante $F : X \mapsto B\sigma(AX)$ avec $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times m}$.

**1. On pose $F : \mathbb{R}^i \to \mathbb{R}^o$. Donner les valeurs de $i$ and $o$ en fonction de $m$ et $n$**

$$n = i = o$$

On veut montrer que l'on peut approximer toute fonction $G$ continue sur $[-1;1]$ (donc $X \in [-1;1]$), on pose donc $n=1$ pour le reste de l'exercice.

**2. Montrer que toute suite bornée admet une sous-suite convergente**

C'est une démonstration du théorème de Bolzano-Weierstrass dans un cas simple. On pose la suite bornée $(x_n)_{n\in\mathbb{N}}$ bornée entre $a$ et $b$. Puisque la suite admet un nombre infini de terme. On note les intervalles $I_0 = [a,(a+b)/2]$ et $I_1 = [(a+b)/2]$. $(x_n)_{n\in\mathbb{N}}$ admet une infinité de termes dans au moins un des deux intervalles. On construit alors itérativement la sous-suite suivante : premier terme dans $I = [a;b]$ puis un second terme dans l'intervalle $I_0$ ou $I_1$ contenant une infinité d'éléments. En répétant le processus avec $I_{00}$, $I_{01}$, $I_{10}$ et $I_{11}$ (en écriture binaire), nous construisons bien une sous suite dont l'ensemble de termes voit leur distance bornée et tendre vers 0. C'est donc une suite de Cauchy qui est convergente (car dans $\mathbb{R}$).

**3. Montrer que $G$ est bornée**

Pour montrer que $G$ est bornée, il suffit de montrer que $|G|$ est majorée. Comme la fonction valeur absolue est continue sur $[-1;1]$ donc $|G|$ aussi. Supposons que $|G|$ ne soit pas majorée. Alors il existe une suite $(x_n)_{n\in\mathbb{N}}$ dans $[-1;1]^{(N)}$ tels que $\lim_{n\to\infty} |G(x_n)| = \infty$. Par la question précédente, il existe une sous suite convergente $(x_{\mu(n)})_{n\in\mathbb{N}}$ de $(x_n)_{n\in\mathbb{N}}$. On a donc

$$\infty = \lim_{n\to\infty} |G(x_n)| = \lim_{n\to\infty} |G(x_{\mu(n)})| = |G(\lim_{n\to\infty} x_{\mu(n)})| \neq \infty$$

Ce qui achève la preuve.

**4. Montrer que pour toute constante $\epsilon > 0$ il existe une liste de nombre $a_1,...a_N$ telle que pour tout $x, y \in [a_n; a_{n+1}]^2$ on a $|G(x) - G(y)| < \epsilon$**

On note $a_i$ les points de $[a;b]$ tels que $a_i < a_{i+1}$ et pour tout $x$ dans $[a_i; a_{i+1}[$ on ait $|G(a_i) - G(x)| < \epsilon$ et $|G(a_i) - G(a_{i+1})| > \epsilon$. Si la suite des $a_i$ est finie alors la preuve s'arrête. Supposons par l'absurde que la suite des $a_i$ est infinie. Alors il existe une limite $l$ à la suite des $a_i$ dans $[a;b[$. Or par continuité de $G$ la suite des $G(a_i)$ converge et donc la suite des $a_i$ est stationnaire.

**5. Montrer le théorème d'approximation universelle pour $\sigma =$ ReLU.**

Il suffit de prendre $A$ et $B$ de telle sorte à construire une fonction affine par morceaux (exercice vu en cours) sur chaque intervalle $[a_n; a_{n+1}]$.