Exercice 1 : prédire le loyer

On souhaite prédire le loyer y d'un appartement à partir de quelques informations $(x_1, ...)$ sur cet appartement. On dispose, pour chaque appartement, de sa surface au sol, de sa hauteur sous plafond et de sa distance au métro le plus proche.

1. Soit $f : \mathbb{R}^n \to \mathbb{R}^m$ une fonction qui pour chaque appartement prédit son loyer. Que valent n et m ?

Pour chaque appartement, on dispose de trois informations, chacune correspondant à un nombre, donc $n = 3$. De plus, on veut prédire le loyer qui correspond à un nombre, donc $m = 1$. Ainsi,

$$f : \begin{array}{ccc} \mathbb{R}^3 & \to & \mathbb{R} \\ (x_1, x_2, x_3) & \mapsto & f(x_1, x_2, x_3) \end{array}$$

2. À partir de maintenant f est un réseau de neurone tel que $f = f_3 \circ f_2 \circ f_1$ avec $f_1 : x \mapsto W_1 x$, $f_3 : x \mapsto W_2 x$ et $f_2 : x \mapsto \mathbb{1}_{x>0}$. On sait que $W_1 \in \mathbb{R}^{4\times 3}$. On note $f_1 : \mathbb{R}^n \to \mathbb{R}^{n_1}$, $f_2 : \mathbb{R}^{n_2} \to \mathbb{R}^{n_3}$ et $f_3 : \mathbb{R}^{n_4} \to \mathbb{R}^m$. Que valent n_1, n_2, n_3 et n_4 ?

Puisque la composée $f = f_3 \circ f_2 \circ f_1$ est définie, on sait que l'entrée de f_2 correspond à la sortie de f_1, donc $n_2 = n_1$. De même, on sait que la sortie de f_2 est l'entrée de f_3, donc $n_3 = n_4$. Il faut alors trouver n_2 et n_3.

On constate que f_2 prend un vecteur et retourne pour chaque coordonnée 1 si elle est positive et zéro sinon, donc la dimension de ce qui rentre dans f_2 ne change pas. Ainsi, on a $n_2 = n_3$. Donc

$$n_1 = n_2 = n_3 = n_4$$

Enfin, on sait que la sortie de f_1 dépend des dimensions de W_1. Plus précisément, le nombre de dimensions de la sortie de f_1 correspond au nombre de lignes de W_1, ici 4. On conclut alors

$$n_1 = n_2 = n_3 = n_4 = 4$$

3. Calculer la Jacobienne de f_1

La fonction $f_1 : \mathbb{R}^3 \to \mathbb{R}^4$ admet $3 \times 4 = 12$ dérivées partielles. La jacobienne est par définition

$$J_{f_1}(x) = \begin{pmatrix} \frac{\partial (f_1)_1}{\partial x_1}(x) & \frac{\partial (f_1)_1}{\partial x_2}(x) & \frac{\partial (f_1)_1}{\partial x_3}(x) \\ \frac{\partial (f_1)_2}{\partial x_1}(x) & \frac{\partial (f_2)_1}{\partial x_2}(x) & \frac{\partial (f_2)_1}{\partial x_3}(x) \\ \frac{\partial (f_3)_1}{\partial x_1}(x) & \frac{\partial (f_3)_1}{\partial x_2}(x) & \frac{\partial (f_3)_1}{\partial x_3}(x) \\ \frac{\partial (f_1)_4}{\partial x_1}(x) & \frac{\partial (f_4)_1}{\partial x_2}(x) & \frac{\partial (f_4)_1}{\partial x_3}(x) \end{pmatrix}$$

Or $f_1(x) = W_1 x$, on a vu en cours que dans ce cas,

$$J_{f_1}(x) = W_1$$

Sinon, on peut le re-démontrer. Prenons une dérivée partielle quelconque $\frac{\partial (f_1)_i}{\partial x_j}$. On remplace la i-eme sortie de f_1 et on obtient

$$\frac{\partial (f_1)_i}{\partial x_j}(x) = \frac{\partial \left(\sum_{k=1}^n W_{i,k} x_k\right)}{\partial x_j}$$

Or la somme $\sum_{k=1}^n W_{i,k} x_k$ ne dépend de x_j que pour $k = j$ et alors, on peut simplifier

$$\frac{\partial (f_1)_i}{\partial x_j}(x) = \frac{\partial \left(W_{i,j} x_j\right)}{\partial x_j}$$

Cette dérivée est très simple à calculer et donne

$$\frac{\partial (f_1)_i}{\partial x_j}(x) = W_{i,j}$$

Ainsi la coordonnée i, j de la jacobienne est égale à la coordonnée i, j de W_1, donc

$$J_{f_1}(x) = W_1$$

4. Calculer la Jacobienne de f_2

La fonction $f_2 : \mathbb{R}^4 \to \mathbb{R}^4$ admet $4 \times 4 = 16$ dérivées partielles. Calculons la dérivée partielle a la position i, j :

$$\frac{\partial (f_2)_i}{\partial x_j}(x) = \frac{\partial \left(\mathbb{1}_{x_i>0}\right)}{\partial x_j}$$

Le terme $\mathbb{1}_{x_i>0}$ ne dépend de x_j que pour $i = j$ et vaut donc 0 sinon. On en déduit que

$$J_{f_2}(x) = \begin{pmatrix} \frac{\partial (f_2)_1}{\partial x_1}(x) & 0 & 0 & 0 \\ 0 & \frac{\partial (f_2)_2}{\partial x_2}(x) & 0 & 0 \\ 0 & 0 & \frac{\partial (f_2)_3}{\partial x_3}(x) & 0 \\ 0 & 0 & 0 & \frac{\partial (f_2)_4}{\partial x_4}(x) \end{pmatrix}$$

De plus, $\mathbb{1}_{x_i>0}$ est constante pour $x > 0$ (et vaut 1), donc sa dérivée vaut 0 pour $x > 0$. De même, $\mathbb{1}_{x_i>0}$ est constante pour $x < 0$ (et vaut 0), donc sa dérivée vaut 0 pour $x < 0$. Enfin, elle n'est pas dérivable en 0, mais on peut prolonger par continuité sa dérivée en 0. On en déduit que la dérivée de $\mathbb{1}_{x_i>0}$ par rapport à x est 0. Donc,

$$J_{f_2}(x) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

5. Calculer la Jacobienne de f_3

On peut constater que f_3 est de la même forme que f_1 et déduire de la question 3 que

$$J_{f_3}(x) = W_2$$

6. Calculer la Jacobienne de f

On utilise la chain rule dont la formule générale est $J_{f_n}(f_{n-1}(\circ \ldots \circ f_1(x))) \times \ldots \times J_{f_1}(x)$. Dans ce cas, on a

$$J_f(x) = J_{f_3}(f_2(f_1(x))) \times J_{f_2}(f_1(x)) \times J_{f_1}(x)$$

En remplaçant avec les résultats précédents, on trouve

$$J_f(x) = W_2 \times 0 \times W_1 = 0$$

Attention, f admet 3 dérivées partielles et donc ici, le terme de droite est le vecteur 0 dans $\mathbb{R}^3$.

7. Montrer que la suite $(x_n)_{n\in\mathbb{N}}$ obtenue par descente de gradient est stationnaire. Autrement dit, la suite $x_n = x_{n-1} - a \times J_f(x_{n-1})$ est constante. Donc, on doit montrer que $x_n = x_{n-1}$.

En utilisant la question précédente, on a immédiatement

$$x_n = x_{n-1} - a \times J_f(x_{n-1}) = x_{n-1} - a \times 0 = x_{n-1}$$

Exercice 2 : Analyse du texte

On représente un texte comme une suite de vecteurs $v \in \mathbb{R}^n$ ou chaque vecteur représente un mot. Notre modèle f va prédire le mot suivant, étant donné un contexte de k mots.

1. En notant x le tenseur dont les lignes sont les mots du contexte, quelles sont les dimensions de x.
Par exemple, une mauvaise réponse serait : x est un vecteur avec 4 coordonnées.

Ici, x est une liste de vecteur, donc une matrice avec k lignes et n colonnes. Ainsi, $x \in \mathbb{R}^{k\times n}$.

2. Donner l'expression de f. La fonction f prend en entrée x et prédit un vecteur $v \in \mathbb{R}^n$. Donc,

$$\begin{array}{rccc} f : & \mathbb{R}^{k\times n} & \to & \mathbb{R}^n \\ & x & \mapsto & f(x) \end{array}$$

3. La fonction f admet combien de dérivées partielles ?

La fonction f admet $(k \times n) \times n = kn^2$ dérivées partielles.

Exercice 3 : Théorie de l'apprentissage

Soit f une fonction convexe. On dit que f est coercive si et seulement si

$$\lim_{x\to+\inf} f(x) = \lim_{x\to-\inf} f(x) = +\inf$$

Pour la suite de l'exercice, on considère f strictement convexe, coercive et dérivable, de dérivée continue.

1. Montrer que si x est un minimum local, alors $f'(x) = 0$.

Supposons par l'absurde que x est un minimum local avec $f'(x) \neq 0$. Par définition de la dérivée,

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

il existe h tel que

$$\begin{cases} \frac{f(x+h)-f(x)}{h} > 0 & \text{si } f'(x) > 0 \\ \frac{f(x+h)-f(x)}{h} < 0 & \text{si } f'(x) < 0 \end{cases}$$

Dans le second cas, on déduit tout de suite, pour $h > 0$ que $f(x+h) < f(x)$ donc x n'est pas un minimum local. Dans le second cas, on prend $h < 0$ et on obtient

$$\frac{f(x-h) - f(x)}{-h} > 0$$

qui devient

$$f(x-h) - f(x) < 0$$

enfin $f(x-h) < f(x)$ et donc dans les deux cas, x n'est pas un minimum local, ce qui est absurde.

2. Montrer que pour f (sous les hypothèses du début de l'exercice), il existe x tel que $f'(x) > 0$.

indication : utiliser $\int_0^b f'(x)dx = f(b) - f(0)$

Supposons, par l'absurde, qu'il n'existe pas de $x \in \mathbb{R}$ tel que $f'(x) > 0$. On sait que $f(0) \in \mathbb{R}$ et par coercite, on a

$$\lim_{b \to +\inf} \int_0^b f'(x)dx = +\inf$$

Or par notre hypothèse du raisonnement par l'absurde, $f' \leq 0$. Donc

$$\lim_{b \to +\inf} \int_0^b f'(x)dx \leq \lim_{b \to +\inf} \int_0^b 0 = 0$$

Ce qui est absurde.

3. Montrer que pour f (sous les hypothèses du début de l'exercice), il existe x tel que $f'(x) < 0$.

indication : utiliser le fait que $\int_a^b f = -\int_b^a f$

Supposons, par l'absurde, qu'il n'existe pas de $x \in \mathbb{R}$ tel que $f'(x) < 0$. On sait que $f(0) \in \mathbb{R}$ et par coercite, on a

$$\lim_{b \to -\inf} \int_b^0 f'(x)dx = \lim_{b \to -\inf} -\int_0^b f'(x)dx = +\inf$$

Or par notre hypothèse du raisonnement par l'absurde, $f' \geq 0$. Donc

$$\lim_{b \to +\inf} -\int_b^0 f'(x)dx \leq \lim_{b \to +\inf} \int_0^b 0 = 0$$

Ce qui est absurde.

4. Montrer que la stricte convexité de f garantie qu'elle n'est jamais croissante puis décroissante puis croissante.

Une fonction est strictement convexe si et seulement si, pour tout $x, y \in \mathbb{R}^2$ et tout $t \in [0; 1]$, on a

$$f(tx + (1-t)y) < tf(x) + (1-t)f(y)$$

Pour résoudre cet exercice, il est pratique de faire un dessin avec une fonction croissante, puis décroissante, puis croissante (par exemple un sinus).

Supposons, par l'absurde que f soit croissante puis décroissante puis croissante. On note x_1 le premier point d'inflexion (le point ou f passe de croissante à décroissante) et x_2 le second (le point ou f passe de décroissante à croissante).

Par définition de la croissance, il existe $x < x_1$ tel que $f(x) < f(x_1)$. De plus, par définition, de la décroissance $f(x_2) < f(x_1)$. Or, puisque $x < x_1 < x_2$, il existe $t \in [0; 1]$ tel que $tx + (1-t)x_2 = x_1$. On a alors

$$f(x_1) = f(tx + (1-t)x_2) > tf(x) + (1-t)f(x_2)$$

Ce qui contredit la définition de la convexité stricte.

5. Montrer que la stricte convexité de f garantie qu'elle n'est jamais décroissante puis croissante puis décroissante.

Procédons, de manière similaire a la question précédente. Supposons, par l'absurde, que f soit décroissante puis croissante puis décroissante. On note x_1 le premier point d'inflexion (le point ou f passe de décroissante à croissante) et x_2 le second (le point ou f passe de croissante à décroissante).

Cette fois, on pose x on point après x_2 tel que, par décroissance, on ait $f(x) < f(x_2)$. On a alors $x_1 < x_2 < x$ et les inégalités $f(x) < f(x_2)$ et $f(x_1) < f(x_2)$. Il existe $t \in [0;1]$ tel que $tx + (1-t)x_1 = x_2$ et

$$f(x_2) = f(tx + (1-t)x_1) > tf(x) + (1-t)f(x_1)$$

Ce qui contredit la définition de la convexité stricte.

6. Montrer que f est décroissante puis croissante.

On a montré que f ne peut ni être croissante puis décroissante puis croissante ni être décroissante puis croissante puis décroissante. Elle est donc soit constante, croissante, décroissante, croissante puis décroissante ou décroissante puis croissante.

Or par coercite, elle ne peut pas ère constante, ni croissante ($\lim_{x\to-\inf} f(x) = +\inf$), ni décroissante ($\lim_{x\to+\inf} f(x) = +\inf$), ni croissante puis décroissante.

Elle est donc décroissante puis croissante.

7. Montrer que f admet un minimum global.

Par la question précédente, on sait que f est de dérivée négative puis positive. Cette dérivée s'annule donc en au moins un point x, son minimum.

8. Donner un exemple pour f.

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