

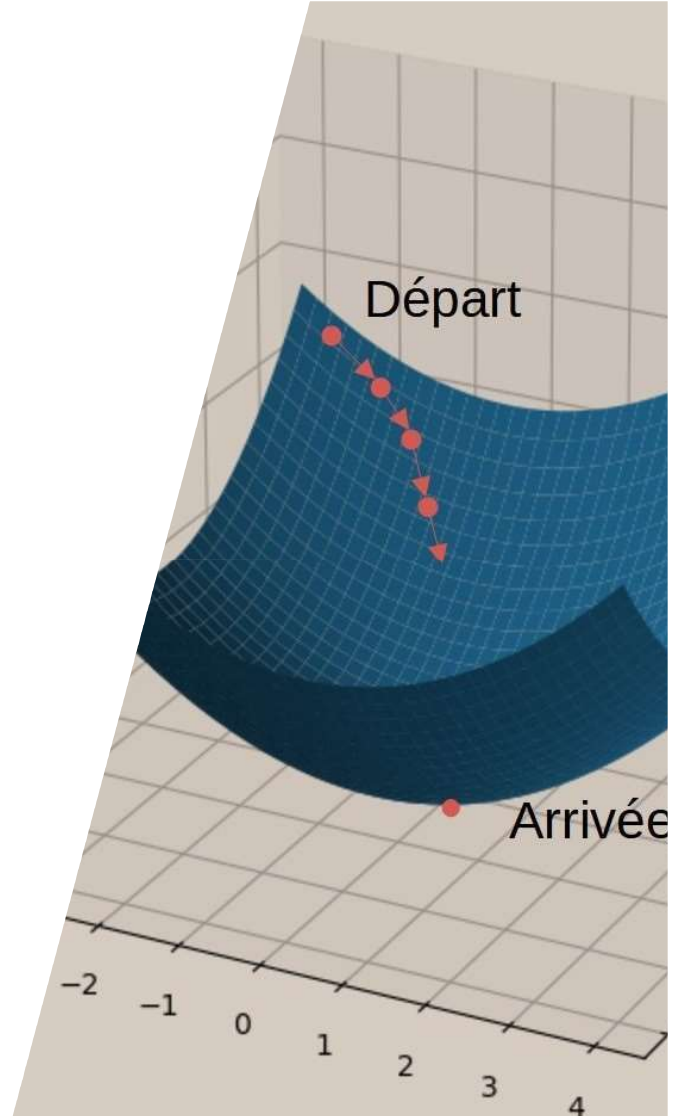
Chapitre 1

Algorithme du gradient

Ressource R4.O4

Tiphaine Jézéquel

2024-2025



Plan du chapitre

1. Optimisation des fonctions convexes
2. Idée de l'algorithme : analogie de la montagne
3. Algorithme du gradient pour une fonction à 1 variable
4. Des dérivées aux dérivées partielles
5. Algorithme du gradient sur une fonction à 2 variables
6. Gradient pour la Régression linéaire

Exercices



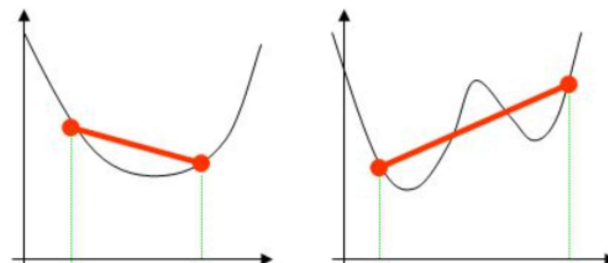
1. Optimisation des fonctions convexes

L'Optimisation est étudiée depuis longtemps en mathématiques, avec des ramifications déjà très développées, dont chacune est un domaine de recherche à part entière ! L'Optimisation convexe est l'un de ces domaines.

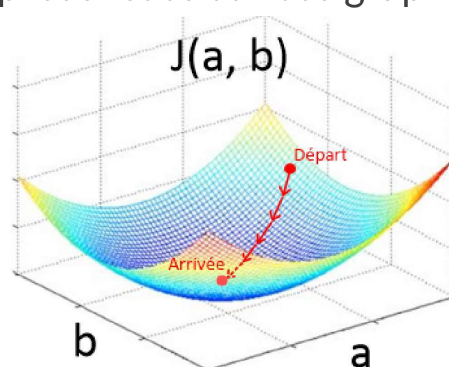
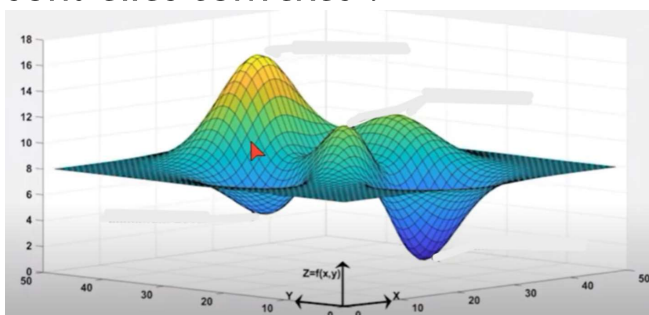
Définition

On dit qu'une **fonction est convexe** si, à chaque fois qu'on relie 2 points de son graphique par un segment, ce segment reste intégralement au-dessus du graphique.

fonction convexe fonction **non** convexe



Exercice. Les fonctions de 2 variables représentées sur ces graphiques sont-elles convexes ?

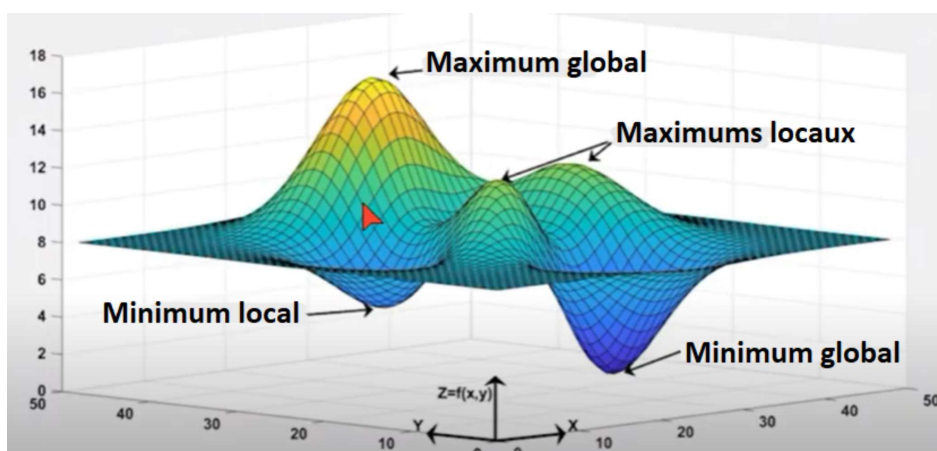


• Minimum d'une fonction convexe

Une fonction convexe a beaucoup de propriétés intéressantes, celle sur son/ses minimums va nous intéresser. On définit d'abord la notion général de minimum local/global.

Minimum local

On dit qu'un point est un **minimum local** s'il existe une zone autour de ce point sur laquelle le point est le minimum de la fonction.



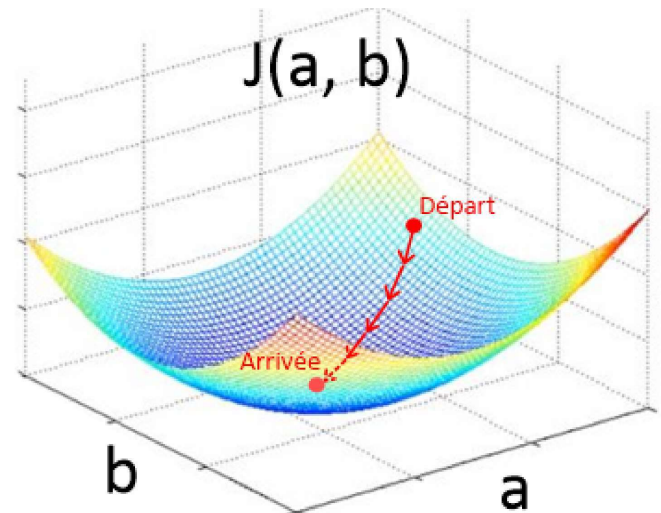
Minimum d'une fonction convexe

Pour une fonction convexe, tout minimum local est forcément un/le minimum global.

• Algorithme du gradient pour une fonction convexe

Algorithme du gradient pour une fonction convexe

Dans le cas où la fonction-objectif est convexe et dérivable, l'algorithme du gradient converge vers l'unique minimum de cette fonction.



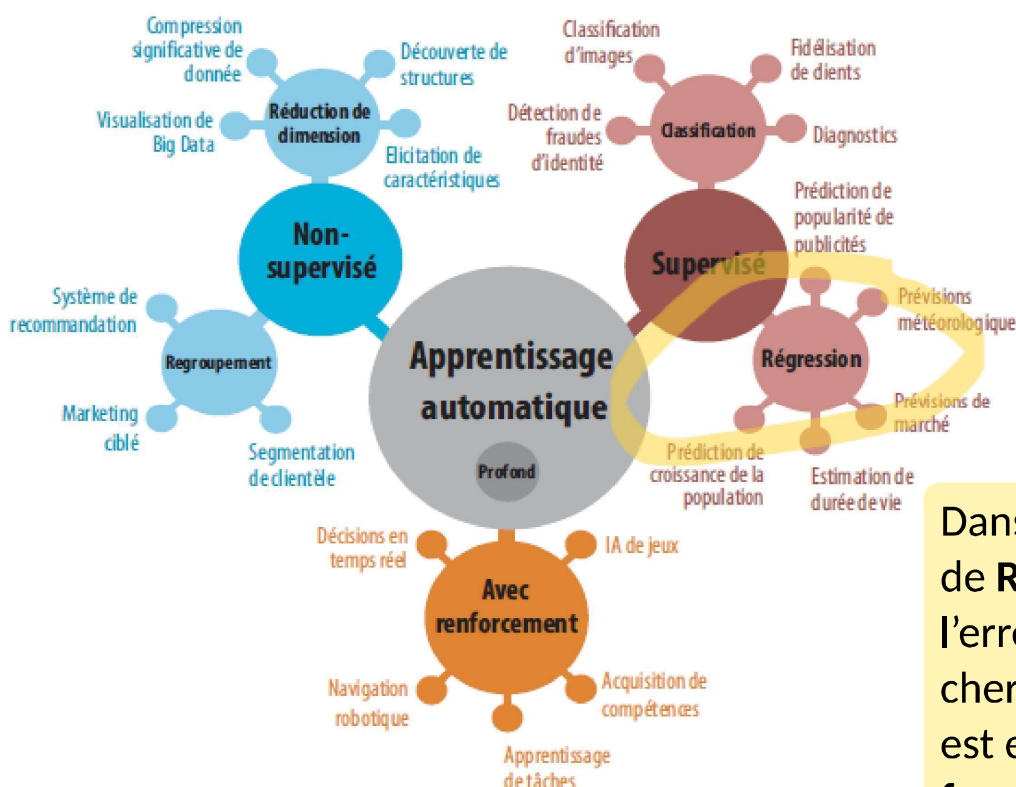
Remarque. Pour une fonction non convexe, l'algorithme du gradient convergera en général vers des minimums locaux.

Théorème précis (pour la culture)

Dans le cas où la fonction-objectif f est *fortement* convexe et dérivable de dérivée continue, l'algorithme du gradient à *pas optimal* converge vers le minimum de f .

• Convexité en Apprentissage automatique

Dans ce chapitre, on appliquera l'algorithme du gradient à un problème de Régression :

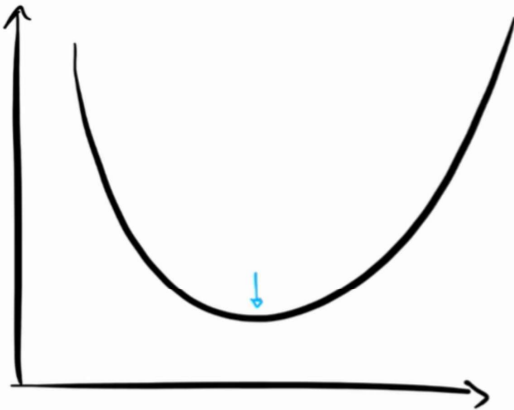


Dans les problèmes de **Régression**, l'erreur qu'on cherche à minimiser est en général une **fonction convexe** !



2. Idée de l'algorithme : analogie de la montagne

<http://machinelearnia.com/descente-de-gradient/>

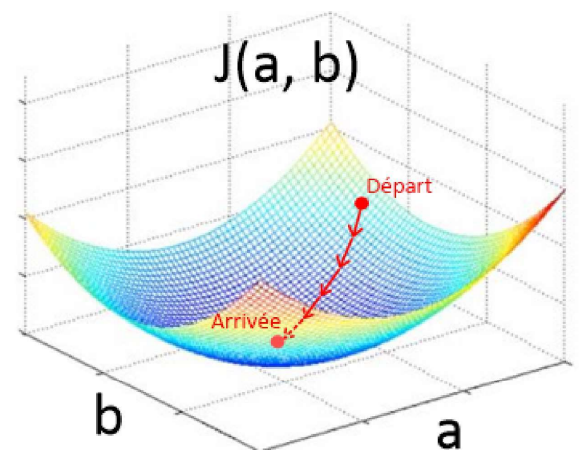
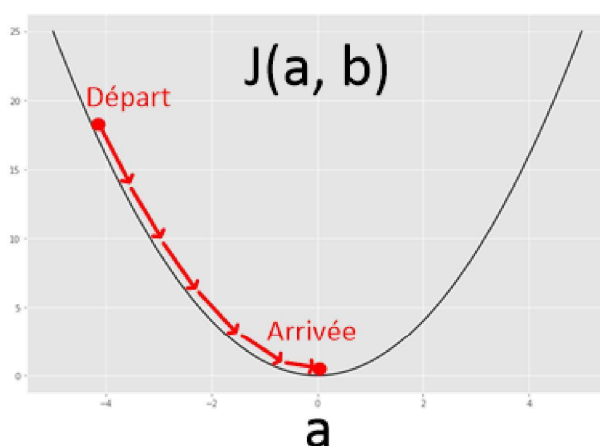
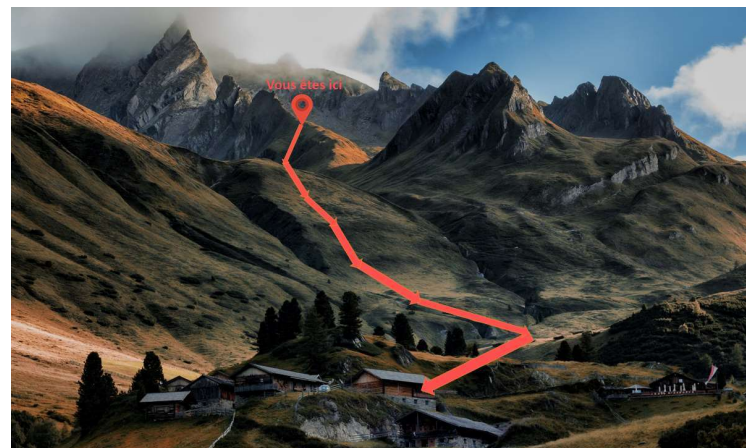


Gradient Descent : Algorithme d'optimisation qui converge vers le minimum d'une fonction convexe



Stratégie

On va se déplacer dans l'ensemble admissible en suivant la pente la + forte de la fonction-objectif ...comme si on était perdu en montagne et qu'on souhaitait atteindre un endroit le plus bas possible.





3. Algorithme du gradient pour une fonction à 1 variable

Algorithme du gradient pour une fonction à 1 variable

Si on cherche le minimum d'une fonction f à une variable $x \in \mathbb{R}$, l'algorithme du gradient consistera à calculer récursivement une suite de points x_i qui se rapprocheront progressivement du minimum de f , de la manière suivante :

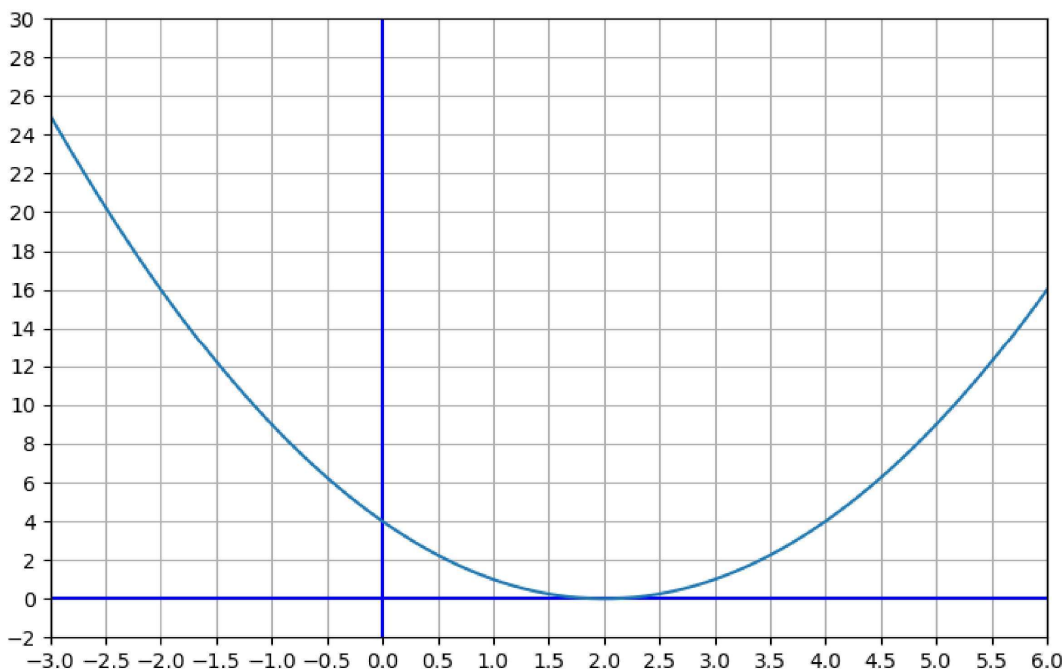
- Initialiser l'algorithme en 1 point $x_0 \in \mathbb{R}$,
- A partir d'un point x_i obtenu par l'algorithme, on calcule le point suivant x_{i+1} ainsi :

$$x_{i+1} = x_i - \alpha f'(x_i).$$

Remarque : le choix du paramètre α est une question importante dans la mise en oeuvre de l'algorithme... mais que nous n'aborderons pas dans ce cours !

Dans les exercices, la valeur du paramètre α vous sera donnée pour chaque implémentation.

- **Exemple.** On s'intéresse à la fonction $f(x) = x^2 - 4x + 4$ dont le graphique est représenté ici :



Dérivée de f :
 $f'(x) = \dots\dots\dots$

Algorithme du Gradient avec $x_0 = -2$ et $\alpha = 0.2$

- | | |
|------------------------------------|------------------------------------|
| • $x_1 = x_0 - 0.2 \times f'(x_0)$ | • $x_2 = x_1 - 0.2 \times f'(x_1)$ |
| avec $f'(x_0) = \dots$ | avec $f'(x_1) = \dots$ |
| donc $x_1 = \dots$ | donc $x_2 = \dots$ |



4. Des dérivées aux dérivées partielles

• Rappels sur les dérivées (fonctions à 1 variable)

Dérivées des fonctions usuelles, intervalle sur lequel la dérivée existe:

Fonction	Dérivée	Intervalle
$f(x) = a$	$f'(x) = 0$	$] -\infty, +\infty[$
$f(x) = x^n$	$f'(x) = nx^{n-1}$	$] -\infty, +\infty[$ si $n \in \mathbb{N}$
$f(x) = \frac{1}{x}$	$f'(x) = -\frac{1}{x^2}$	$] -\infty, 0[\cup] 0, +\infty[$
$f(x) = \sqrt{x}$	$f'(x) = \frac{1}{2\sqrt{x}}$	$] 0, +\infty[$
$f(x) = \ln(x)$	$f'(x) = \frac{1}{x}$	$] 0, +\infty[$
$f(x) = e^x$	$f'(x) = e^x$	$] -\infty, +\infty[$
$f(x) = \sin(x)$	$f'(x) = \cos(x)$	$] -\infty, +\infty[$
$f(x) = \cos(x)$	$f'(x) = -\sin(x)$	$] -\infty, +\infty[$

Dérivées des fonctions de la forme $f(u)$ où u est une fonction $u(x)$:

Fonction	Dérivée
u	u'
u^n	$n u' u^{n-1}$
$\frac{1}{u}$	$-\frac{u'}{u^2}$
\sqrt{u}	$\frac{u'}{2\sqrt{u}}$

Fonction	Dérivée
$\ln(u)$	$\frac{u'}{u}$
e^u	$u' e^u$
$\sin(u)$	$u' \cos(u)$
$\cos(u)$	$-u' \sin(u)$

• Exemples

1. $f(x) = 2 \ln(x)$, $f'(x) = \dots\dots\dots$

2. $f(x) = e^{3x} + 2$, $f'(x) = \dots\dots\dots$

3. $f(x) = 1 + 4x^3 + x^2$, $f'(x) = \dots\dots\dots$

4. $f(x) = ax^4 + b$, $f'(x) = \dots\dots\dots$

5. $f(x) = \cos(ax + b)$, $f'(x) = \dots\dots\dots$

6. $f(y) = \sin(ay + c)$, $f'(y) = \dots\dots\dots$

7. $f(a) = \ln(a) + c$, $f'(a) = \dots\dots\dots$

• Notion de dérivée partielle (fonctions à 2 variables)

Dérivée partielle

Soit $f(x, y)$ une fonction à 2 variables réelles x et y .

La *dérivée partielle de f par rapport à x* , notée $\frac{\partial f}{\partial x}(x, y)$, est la fonction à 2 variables obtenue en dérivant f par rapport à x en considérant y comme une constante.

On définit de même la dérivée partielle $\frac{\partial f}{\partial y}(x, y)$ de f par rapport à y .

• Exemples

1. $f(x, y) = 3 \cdot \frac{1}{x} + y,$

$$\frac{\partial f}{\partial x}(x, y) = \dots\dots\dots, \quad \frac{\partial f}{\partial y}(x, y) = \dots\dots\dots$$

2. $f(x, y) = \cos(2x) + y,$

$$\frac{\partial f}{\partial x}(x, y) = \dots\dots\dots, \quad \frac{\partial f}{\partial y}(x, y) = \dots\dots\dots$$

3. $f(x, y) = \cos(2x + y),$

$$\frac{\partial f}{\partial x}(x, y) = \dots\dots\dots, \quad \frac{\partial f}{\partial y}(x, y) = \dots\dots\dots$$

4. $f(x, y) = 2xy + 3,$

$$\frac{\partial f}{\partial x}(x, y) = \dots\dots\dots, \quad \frac{\partial f}{\partial y}(x, y) = \dots\dots\dots$$

5. $f(x, y) = \frac{2x}{y},$

$$\frac{\partial f}{\partial x}(x, y) = \dots\dots\dots, \quad \frac{\partial f}{\partial y}(x, y) = \dots\dots\dots$$



5. Algorithme du gradient sur une fonction à 2 variables

Algorithme du gradient pour une fonction à 2 variables

Si on cherche le minimum d'une fonction f à 2 variables $x, y \in \mathbb{R}$, l'algorithme du gradient consistera à calculer récursivement une suite de points (x_i, y_i) qui se rapprocheront progressivement du minimum de F , de la manière suivante :

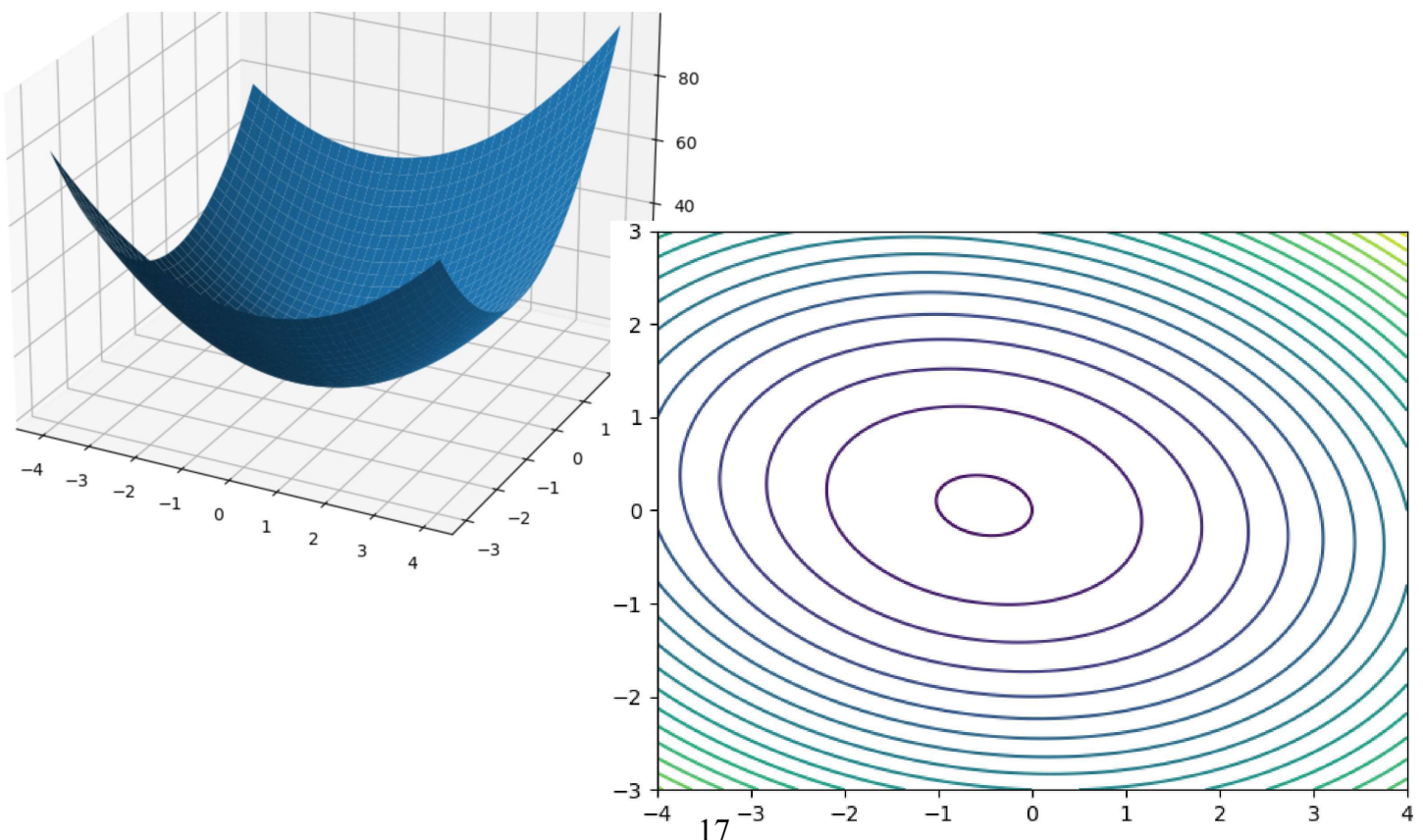
- Initialiser l'algorithme en 1 point $(x_0, y_0) \in \mathbb{R}^2$,
- A partir d'un point (x_i, y_i) obtenu par l'algorithme, on calcule le point suivant (x_{i+1}, y_{i+1}) ainsi :

$$x_{i+1} = x_i - \alpha \frac{\partial f}{\partial x}(x_i, y_i)$$

$$y_{i+1} = y_i - \alpha \frac{\partial f}{\partial y}(x_i, y_i)$$

• Exemple

On s'intéresse à la fonction $f(x, y) = 2x^2 + 2x + 5y^2 + xy$ dont voici une représentation graphique, puis ses lignes de niveau :



On calcule les dérivées partielles de f :

$$\frac{\partial f}{\partial x}(x, y) = \dots\dots\dots, \quad \frac{\partial f}{\partial y}(x, y) = \dots\dots\dots$$

Algorithme du Gradient avec $(x_0, y_0) = (-2, 2)$ et $\alpha = 0.2$

• $x_1 = x_0 - 0.2 \times \frac{\partial f}{\partial x}(x_0, y_0)$

• $y_1 = y_0 - 0.2 \times \frac{\partial f}{\partial y}(x_0, y_0)$

$$\frac{\partial f}{\partial x}(x_0, y_0) = \dots$$

$$\frac{\partial f}{\partial y}(x_0, y_0) = \dots$$

$$x_1 = \dots$$

$$y_1 = \dots$$

• $x_2 = x_1 - 0.2 \times \frac{\partial f}{\partial x}(x_1, y_1)$

• $y_2 = y_1 - 0.2 \times \frac{\partial f}{\partial y}(x_1, y_1)$

$$\frac{\partial f}{\partial x}(x_1, y_1) = \dots$$

$$\frac{\partial f}{\partial y}(x_1, y_1) = \dots$$

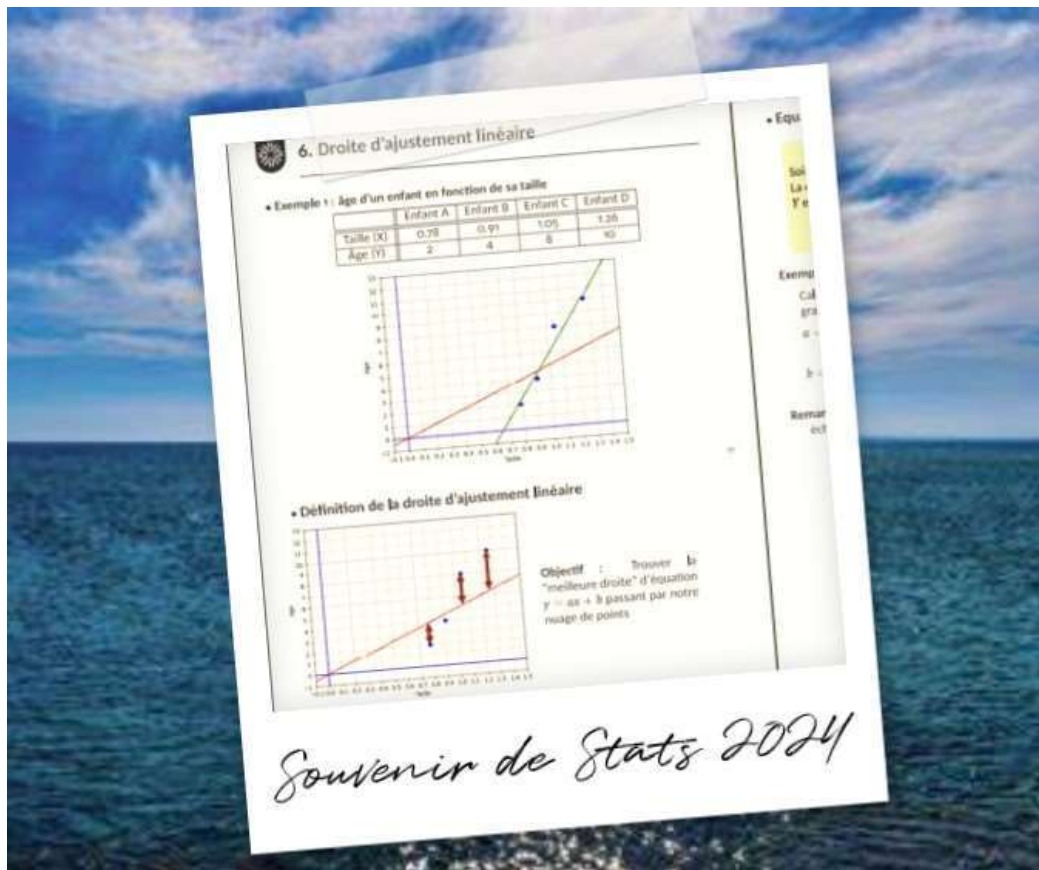
$$x_2 = \dots$$

$$y_2 = \dots$$

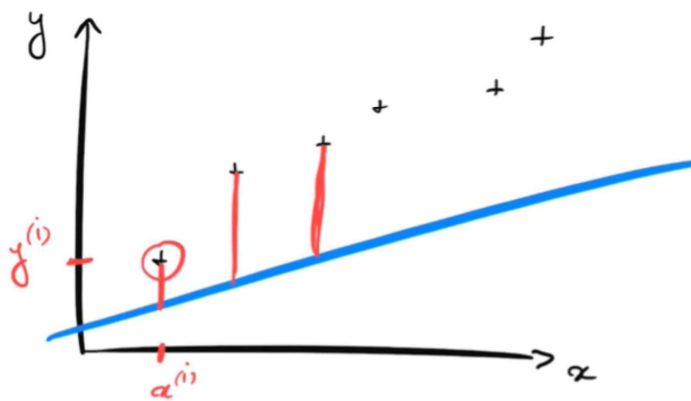


6. Gradient pour la Régression linéaire

• Souvenirs-souvenirs...



... exemple 5 dans le Cours de présentation



1. Dataset : (x, y) $m=6$, $n=1$ $x^{(i)}$

2. Modèle : $f(x) = ax + b$

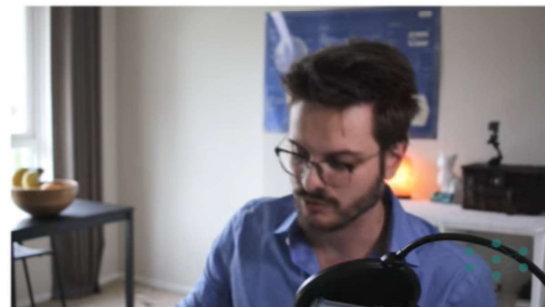
3. Fonction Coût :

1. Dataset
y : Target
x : features

2. Modèle
paramètres

3. Fonction Coût

4. Algorithme de
minimisation



<http://machinelearnia.com/regression-lineaire-simple/>

• Exemple de l'âge et la taille des enfants

Quand on est enfant, en regardant les personnes autour de nous, on apprend progressivement à reconnaître les adultes des enfants, et à évaluer l'âge des autres enfants, notamment en fonction de leur taille.

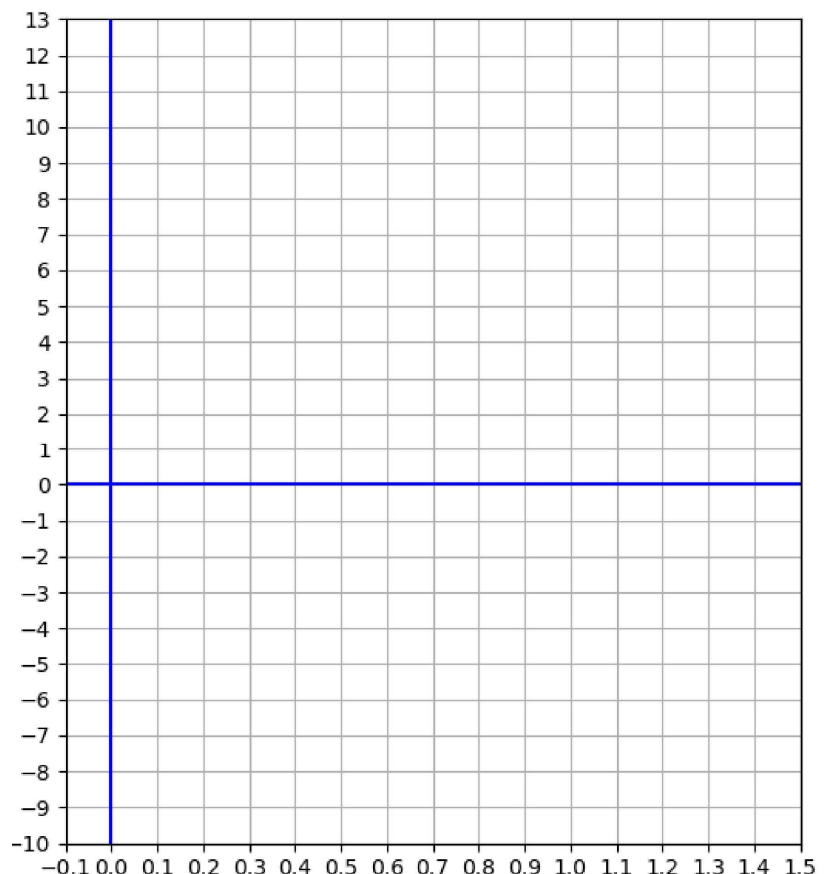
Essayons de faire apprendre à une machine à retrouver l'âge d'un.e enfant en fonction de sa taille, en lui apprenant à partir de quelques exemples.

	Enfant A	Enfant B	Enfant C	Enfant D	Enfant E
Taille	0.84	0.98	1.14	1.32	1.44
Âge	2	4	8	10	12

• Représentation des données

On représente graphiquement ces 5 données :

Elles nous semblent suivre approximativement une droite, on va chercher une droite qui soit le plus près possible des ces 5 points.



• Fonction à minimiser : F

Inconnues : on cherche une droite, donc son équation $y = ax + b$. Les inconnues qu'on veut déterminer sont donc

Objectif, Fonction-objectif : notre objectif est que les distances entre chaque point et la droite soient les + petites possibles :

- pour l'enfant A : $2 - (a \times 0.84 + b)$
- pour l'enfant B : $4 - (a \times 0.98 + b)$
- pour l'enfant C :
- pour l'enfant D :
- pour l'enfant E :

On décide de minimiser la fonction-objectif suivante :

$$\begin{aligned} F(a, b) &= (2 - (0.84a + b))^2 + (4 - (0.98a + b))^2 + (\dots\dots\dots)^2 \\ &\quad + (\dots\dots\dots)^2 + (\dots\dots\dots)^2 \\ &= \dots \\ &= 6.7816 a^2 + 5 b^2 + 11.44 ab - 90.4 a - 72 b + 328 \end{aligned}$$

• Algorithme du Gradient pour minimiser F

On calcule les dérivées partielles de F :

$$\frac{\partial F}{\partial a}(a, b) = \dots\dots\dots,$$

$$\frac{\partial F}{\partial b}(a, b) = \dots\dots\dots$$

Initialisons par exemple avec $a_0 = 20$, $b_0 = -10$, et $\alpha = 0.05$.

On peut alors calculer (a_1, b_1) :

$$\begin{aligned} a_1 &= 20 - 0.05 \times \frac{\partial F}{\partial a}(20, -10) \\ &= 20 - 0.05 \times \dots\dots\dots \\ &= 20 - 0.05 \times \dots\dots\dots \\ &= \end{aligned}$$

$$\begin{aligned} b_1 &= -10 - 0.05 \times \frac{\partial F}{\partial b}(20, -10) \\ &= -10 - 0.05 \times \dots\dots\dots \\ &= -10 - 0.05 \times \dots\dots\dots \\ &= \end{aligned}$$

$$a_2 = a_1 - 0.05 \times \frac{\partial F}{\partial a}(a_1, b_1)$$

$$= \dots - 0.05 \times \dots$$

$$= \dots - 0.05 \times \dots$$

$$=$$

$$b_2 = b_1 - 0.05 \times \frac{\partial F}{\partial b}(a_1, b_1)$$

$$= \dots - 0.05 \times \dots$$

$$= \dots - 0.05 \times \dots$$

$$=$$

• Lien avec les formules de Stats

Dans le cours de Stats on avait vu que les coefficients a et b optimaux étaient obtenus par ces formules :

• Equation de la droite d'ajustement linéaire

Proposition

Soient X et Y deux variables statistiques sur une même population.
La droite d'ajustement linéaire de Y en X (c'est-à-dire avec X en abscisse et Y en ordonnée) a pour équation $y = ax + b$ avec :

$$a = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \quad b = \bar{Y} - a\bar{X}.$$

Ce qui donnerait ici : $a = 16.8796$, $b = -12.1103$.