Fonctions de plusieurs variables
Gradient

Introduction à l'apprentissage automatique
Master Sciences Cognitives
Aix Marseille Université

Alexis Nasr

Motivations

- Les fonctions vues jusque là étaient de la forme $f(x) = y$ avec $x \in \mathbb{R}$.
- Les fonctions qui nous intéressent, en particulier la fonction d'erreur, possèdent **plusieurs inconnues** : $f(x_1, \ldots, x_n)$.
- On aimerait trouver les valeurs des variables pour lesquelles la valeur de la fonction f est **minimale**.

$$\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

- Une méthode générale permettant de trouver ces valeurs, dans certaines conditions, est l'algorithme de **descente du gradient**, qui, comme son nom l'indique repose sur la notion de **gradient**.

Objectifs

- Comprendre la notion de **dérivée partielle** et de **dérivée directionnelle** d'une fonction.
- Comprendre la notion de **gradient** d'une fonction.
- Comprendre le principe de l'algorithme de **descente du gradient**.

Plan

Fonctions de plusieurs variables

Dérivée partielle

gradient

Descente du gradient

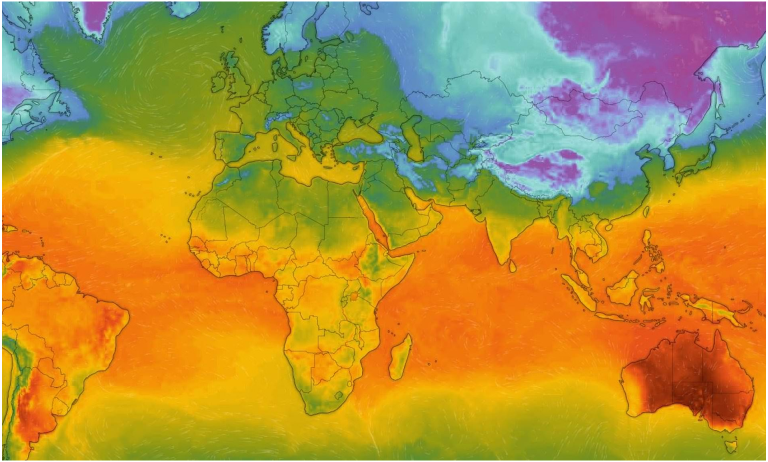
Fonctions de plusieurs variables

- Dans certains cas, une quantité peut dépendre de plusieurs autres.
- Exemple : on peut considérer que la température T en un point de la surface terrestre dépend de la longitude x et de la latitude y du point
- On peut alors voir T comme une fonction des deux variables x et y.

$$T = f(x, y)$$

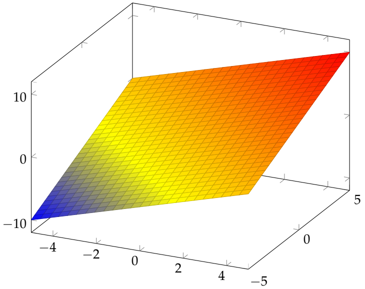
- Si f est une fonction de deux variables, alors de graphe de f est l'ensemble des points $(x, y, z) \in \mathbb{R}^3$ avec $z = f(x, y)$
- Le graphe de la fonction f se présente sous la forme d'une **surface**.

Température sur la surface terrestre le 19 décembre 2019



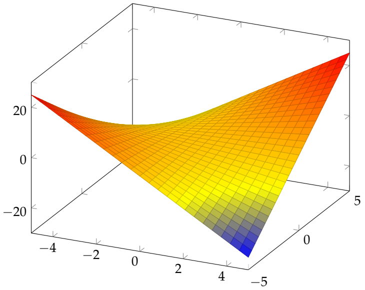
$$T = f(x, y)$$

$f(x, y) = ax + by$

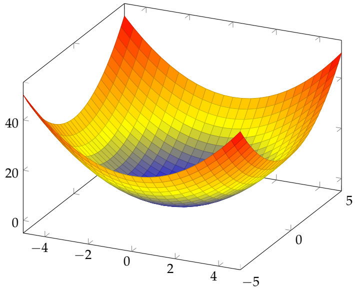


- f est linéaire en x et en y
- Le graphe de f est un plan.

$f(x,y) = xy$



$f(x, y) = x^2 + y^2$



Dérivées partielles en un point

- Soit f une fonction de deux variables x et y.
- On se demande comment varie f si on ne modifie qu'une seule variable : x ou y.
- La réponse est donnée par la **dérivée partielle** de f par rapport à cette variable.
- Si on fait varier x en gardant une valeur constante de y (par exemple $y = b$) alors on peut voir f comme une fonction d'une seule variable x.
- On peut écrire $g(x) = f(x, b)$.
- Si g est dérivable en a (pour $x = a$), alors on appelle cette dérivée, la dérivée partielle de f par rapport à x en (a, b) que l'on note $f_x(a, b)$

$$f_x(a, b) = g'(a)$$

Dérivées partielles en un point

- par définition de la dérivée, on a :

$$g'(a) = \lim_{h \to 0} \frac{g(a+h) - g(a)}{h}$$

- ou bien :

$$f_x(a,b) = \lim_{h \to 0} \frac{f(a+h,b) - f(a,b)}{h}$$

Dérivée partielle

- Si on fait varier le point (a,b), f_x et f_y deviennent des fonctions de deux variables.
- Si f est une fonction de deux variables, alors ses dérivées partielles sont les fonctions f_x et f_y définies de la façon suivante :

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

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

- Les règles de calcul des dérivées partielles sont celles des fonctions d'une seule variable !
- Notation de Leibnitz :
  - $f_x(x,y)$ est notée $\frac{\partial}{\partial x} f(x,y)$
  - $f_y(x,y)$ est notée $\frac{\partial}{\partial y} f(x,y)$

Exemples

- $f(x,y) = ax + by$

$$\frac{\partial}{\partial x} f(x,y) = a$$
$$\frac{\partial}{\partial y} f(x,y) = b$$

- $f(x,y) = xy$

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

Exemple

- $T = f(x, y)$ est la température en un point de la surface terrestre de longitude x et de latitude y.
- $\frac{\partial}{\partial x} f(x, y)$ est la variation de température lorsque x augmente : lorsqu'on se dirige vers l'**est**.
- $\frac{\partial}{\partial y} f(x, y)$ est la variation de température lorsque y augmente : lorsqu'on se dirige vers le **nord**.

Dérivée directionnelle

- On souhaite déterminer la variation de $f(x, y)$ lorsqu'on se déplace dans la direction du vecteur [1] $\mathbf{u} = [a, b]^T$.
- Cette variation est appelée **dérivée directionnelle** de la fonction f, dans la direction du vecteur **u**.
- La dérivée directionnelle est notée $\nabla_{\mathbf{u}} f(x, y)$, elle est définie de la façon suivante :

$$\nabla_{\mathbf{u}} f(x, y) = \frac{\partial}{\partial x} f(x, y) a + \frac{\partial}{\partial y} f(x, y) b$$

1. Pour des raisons techniques, que l'on comprendra plus tard, on considère que **u** est un vecteur unitaire (sa norme est égale à 1)

Exemple

- $T = f(x,y)$ est la température en un point de la surface terrestre de longitude x et de latitude y.
- $\frac{\partial}{\partial x} f(x,y)$ est la variation de température lorsque x augmente : lorsqu'on se dirige vers l'est.
- $\frac{\partial}{\partial y} f(x,y)$ est la variation de température lorsque y augmente : lorsqu'on se dirige vers le nord.
- Si on se déplace vers le nord est, $\mathbf{u} = \left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]$, la variation de température est donnée par la dérivée directionnelle :

$$\nabla_{\mathbf{u}} f(x,y) = \frac{\partial}{\partial x} f(x,y) \frac{\sqrt{2}}{2} + \frac{\partial}{\partial y} f(x,y) \frac{\sqrt{2}}{2}$$

- Si on se déplace vers le sud est, $\mathbf{u} = \left[\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right]$, la variation de température est donnée par la dérivée directionnelle :

$$\nabla_{\mathbf{u}} f(x,y) = \frac{\partial}{\partial x} f(x,y) \frac{\sqrt{2}}{2} - \frac{\partial}{\partial y} f(x,y) \frac{\sqrt{2}}{2}$$

Gradient

- La dérivée directionnelle peut être vue comme le produit scalaire de deux vecteurs :

$$\begin{aligned}\nabla_{\mathbf{u}} f(x,y) &= \frac{\partial}{\partial x} f(x,y) a + \frac{\partial}{\partial y} f(x,y) b \\ &= \left[\frac{\partial}{\partial x} f(x,y), \frac{\partial}{\partial y} f(x,y)\right]^T \cdot [a,b]^T \\ &= \left[\frac{\partial}{\partial x} f(x,y), \frac{\partial}{\partial y} f(x,y)\right]^T \cdot \mathbf{u}\end{aligned}$$

- Le premier vecteur est appelé **gradient** de la fonction $f(x,y)$, il est noté $\nabla f(x,y)$:

$$\nabla f(x,y) = \left[\frac{\partial}{\partial x} f(x,y), \frac{\partial}{\partial y} f(x,y)\right]^T$$

- La dérivée directionnelle s'écrit donc :

$$\nabla_{\mathbf{u}} f(x,y) = \nabla f(x,y) \cdot \mathbf{u}$$

Gradient

- Le gradient se généralise naturellement aux fonctions possédant un nombre quelconque n de variables.
- La dimension du gradient (le nombre de composantes) est égal au nombre de variables de la fonction.

$$\nabla f(\mathbf{x}) = \left[\frac{\partial}{\partial x_1} f(\mathbf{x}), \ldots, \frac{\partial}{\partial x_n} f(\mathbf{x})\right]^T$$

Maximisation de la dérivée directionnelle

- Etant donné une fonction de plusieurs variables $f(\mathbf{x})$
- Les dérivées directionnelles en un point P donnent la variation de f dans toutes les directions possibles $\mathbf{u}$.
- Quelle est la direction $\mathbf{u}^*$ pour laquelle la variation de f est la plus grande possible ?

$$\mathbf{u}^* = \underset{\mathbf{u} \in \mathbb{R}^n}{\arg\max} \nabla_{\mathbf{u}} f(\mathbf{x})$$

- Quelle est la valeur de cette variation ?

Maximisation de la dérivée directionnelle

$$\begin{aligned}
\nabla_{\mathbf{u}} f(\mathbf{x}) &= \nabla f(\mathbf{x}) \cdot \mathbf{u} \\
&= |\nabla f(\mathbf{x})| \times |\mathbf{u}| \times \cos(\theta) \\
&= |\nabla f(\mathbf{x})| \times \cos(\theta)
\end{aligned}$$

- θ est l'angle entre ∇f et $\mathbf{u}$
- La valeur maximale de $\cos(\theta)$ est 1
- Elle est atteinte pour $\theta = 0$, en d'autres termes lorsque ∇f et $\mathbf{u}$ ont la même direction.
- Dans ce cas, $\nabla_{\mathbf{u}} f(\mathbf{x}) = |\nabla f(\mathbf{x})|$

Points critiques et extremums d'une fonction à plusieurs variables

- Un point critique d'une fonction $f(\mathbf{x})$ est un point $\mathbf{a}$ tel que :

$$\nabla f(\mathbf{a}) = 0$$

- Les extremums locaux de f sont des points critiques de f.
- Mais tout point critique ne constitue pas forcément un extremum.
- La recherche des extremums de f suppose donc de résoudre l'équation :

$$\nabla f(\mathbf{x}) = 0$$

- Dans certains cas, on peut trouver une solution **exacte** de l'équation.
- Mais la plupart du temps (dans notre cas) on a recours à une méthode itérative qui donnera une solution **approchée** de l'équation.

Descente du gradient : idée générale

- Le gradient de la fonction f, calculé au point $\mathbf{x}$: $\nabla f(\mathbf{x})$, indique comment faire varier le vecteur $\mathbf{x}$ pour aboutir à l'augmentation maximale de f.
- Conséquences :

$$\mathbf{x}' = \mathbf{x} + \eta \nabla f(\mathbf{x}) \ \ \Rightarrow f(\mathbf{x}') \geq f(\mathbf{x})$$

$$\mathbf{x}' = \mathbf{x} - \eta \nabla f(\mathbf{x}) \ \ \Rightarrow f(\mathbf{x}') \leq f(\mathbf{x})$$

- η est appelé **pas d'apprentissage**, il permet de contrôler la variation de $\mathbf{x}$.

Exemple

- fonction

$$f(x,y) = (x+y) \times (y+1) = xy + x + y^2 + y$$

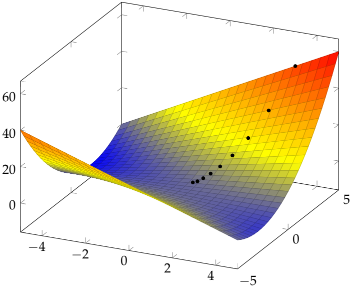
- dérivées partielles

$$\frac{\partial}{\partial x} f(x,y) = y + 1$$
$$\frac{\partial}{\partial y} f(x,y) = x + 2y + 1$$

- gradient : $\nabla f(x,y) = [y+1, x+2y+1]^T$
- gradient au point $(3,5)$:$\nabla f(3,5) = [6,14]^T$
- mise à jour ($\eta = 0.1$)

$$x = 3 - 0.1 \times 6$$
$$y = 5 - 0.1 \times 14$$

Exemple : $f(x,y) = (x+y) \times (y+1), \eta = 0.1$



i	x	y	$f(x,y)$
0	3	5	48
1	2.4	3.66	28.24
2	1.934	2.635	16.605
3	1.571	1.851	9.75
4	1.285	1.252	5.714
5	1.060	0.796	3.332
6	0.881	0.448	1.925
7	0.736	0.185	1.091
8	0.617	-0.014	1.091

Sources

- Stewart, James. Essential calculus : Early transcendentals. Cengage Learning, 8ème édition, 2012
- Tom Mitchell, *Machine Learning*, McGraw Hill, 1997.