Nom: Iordache Paul-Tiberiu, Soares Robin

Groupe: 21.4 A

Sujet: descente du gradient à pas optimal - steepest descent

Fiche de cours

I- INTRODUCTION

Le gradient d'une fonction avec plusieurs variables au point A est un vecteur qui traduit la variabilité de cette fonction au voisinage de A.

Dans un système de coordonnées cartésiennes, le gradient d'une fonction $f(x1,x2\ldots,xn)$ est le vecteur de composantes □f/□x(i) (avec $i=1,2,...,n$) c'est-à-dire les dérivées partielles de f par rapport aux coordonnées. Notation:

$$\nabla f=(\partial f/\partial x(1),\ \partial f/\partial x(2)...,\ \partial f/\partial x(n))$$

L'algorithme de la descente de gradient est un algorithme d'optimisation itératif (qui utilise principalement des boucles for et while) qui consiste à trouver le minimum d'une fonction convexe. C'est un algorithme très utilisé dans le machine learning et le deep learning.

Maintenant que nous comprenons un peu mieux notre sujet, nous allons pouvoir nous pencher sur le fonctionnement, la compréhension et l'utilisation de cet algorithme.

II- ALGORITHME DE LA DESCENTE DE GRADIENT

Analytiquement, pour minimiser ou maximiser une fonction, on va utiliser la dérivée de notre fonction égale à 0. Nous allons pouvoir en déduire la valeur de x qui minimise ou maximise notre fonction.

Prenons comme exemple la fonction suivante: $f(x)= x^2 +x +1$, on peut calculer le minimum de ($f(x) = x^2 +x +1$) car c'est une fonction simple.

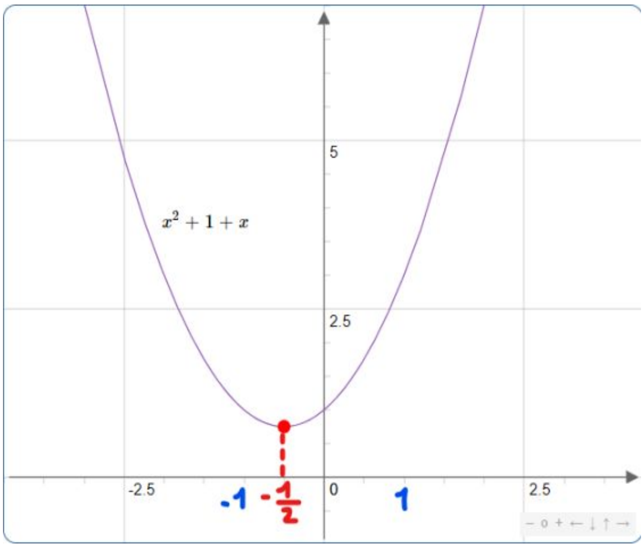
Pour trouver le point critique, il suffit de calculer sa dérivée en fonction de x et de trouver la valeur de x quand f'(x)=0:

$f'(x) = 2x+1 = 0 \Rightarrow x= -½$

Après, on doit calculer la deuxième dérivée de f par rapport à x:

si $f''(x)>0$ on a bien un point qui est minimal.
$f''(x)<0$ on a bien un point qui est maximal.

$f''(x)= 2 > 0$, donc on a bien notre minima dans le point $x = -1/2$



Ici, nous avons pris une fonction simple, mais calculer analytiquement le minima de fonctions plus complexes n'est pas chose aisée. Heureusement, nous pouvons utiliser l'algorithme de la descente de gradient (descente du gradient à pas optimal - steepest descent) qui va nous faciliter grandement la tâche.

La formule de l'algorithme de la descente du gradient tient en une ligne:

$$u_{n+1} = u_n - pas * \nabla f$$

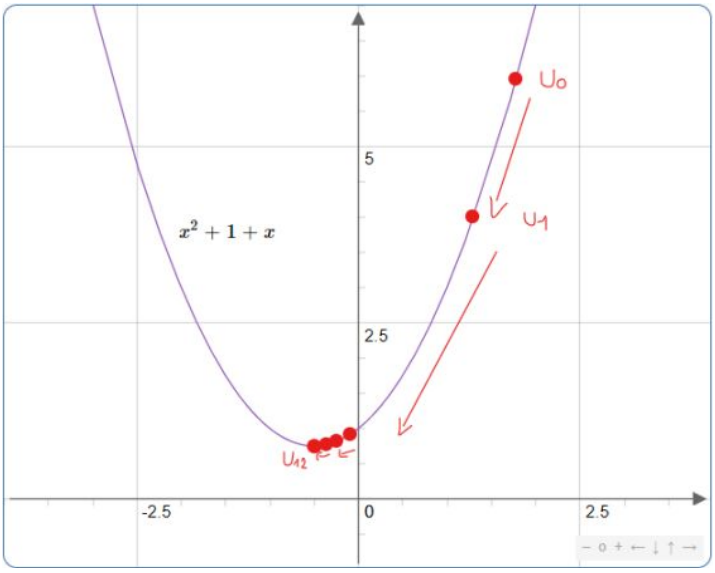
Avec "pas" étant la valeur du pas optimal et ∇f le gradient.

$\nabla f = \partial f/\partial x = f'(x)$ (parce qu'on a une seule variable)
Utilisons notre fonction $f(x)= x^2 +x +1$, pour notre exemple.

Pour utiliser cette formule, il faut commencer par prendre notre premier élément $\mathbf{u_0}$ aléatoirement dans notre fonction.

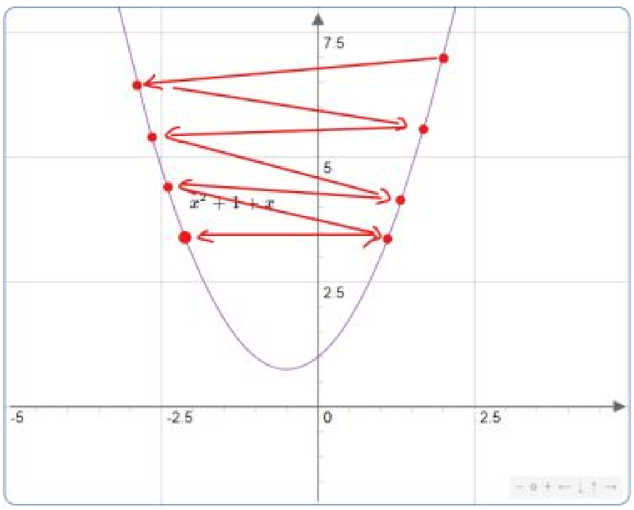
Nous prendrons $\mathbf{u_0}$ = 1.8 et **pas** = 0.3
Il faut ensuite appliquer la formule ci-dessus jusqu'à ce qu'il y ait convergence :

1	u(x)	f'(x)	y=f(x)
2	1.8		6.04
3	0.42	4.6	1.5964
4	-0.132	1.84	0.88542
5	-0.3528	0.736	0.77167
6	-0.4411	0.2944	0.75347
7	-0.4764	0.11776	0.75055
8	-0.4906	0.0471	0.75009
9	-0.4962	0.01884	0.75001
10	-0.4985	0.00754	0.75
11	-0.4994	0.00301	0.75
12	-0.4998	0.00121	0.75
13	-0.4999	0.00048	0.75
14	-0.5	0.00019	0.75



Attention au choix de son **pas** ! C'est peut être l'un des paramètres les plus importants ! En effet, si votre pas est trop grand, vous ne trouverez jamais votre minima. De même si il est trop petit, vous ferez trop de calculs…

Exemple d'un pas trop grand :



III- CONCLUSION

Pour conclure, l'algorithme du gradient est un algorithme utile dans de nombreux domaines, par exemple, le machine learning. Il est très utile car il présente la possibilité de paralléliser les algorithmes. De plus, il utilise les implémentations classiques des méthodes de data mining.
Il présente également de nombreux avantages par rapport aux autres algorithmes, ce qui est démontré par rapport à l'algorithme analytique. Par contre, il présente également quelques inconvénients, comme tous les algorithmes. Par exemple, le grand nombre de paramètres qui influent sur le comportement de l’algorithme peuvent poser problèmes.