



Communauté Economique et Monétaire  
d'Afrique Centrale



**BP:** 294, Yaoundé (République de Cameroun)

**Email:** Isseacemac@yahoo.fr

**Tél:** (+237) 22-22-01-34

**Web:** www.issea-cemac.org

**Fax:** (+237) 22-22-95-21

**Année :** Janvier 2024-2025

## PROJET DE MACHINE LEARNING ET OPTIMISATION

Rédigé par:

- ANATO Diane
- NSIMOUESSA Dieuveil Nathan

---

**LA METHODE DE NEWTON CONVERGE TOUJOURS PLUS  
RAPIDEMENT QUE LA DESCENTE DE GRADIENT?**

---

*Sous la supervision de: Mme Marie Thérèse MBIA NDI, Ingénieur Statisticien Economiste*

## Introduction

Dans le domaine de l'optimisation, la recherche de solutions rapides et précises est primordiale. La méthode de Newton et la descente de gradient sont deux outils couramment utilisés pour atteindre cet objectif. Bien que toutes deux visent à minimiser une fonction, leurs approches diffèrent significativement. Nous nous proposons d'analyser en détail ces différences et de déterminer si la méthode de Newton conserve toujours son avantage en termes de vitesse de convergence.

## 1 Généralité sur la méthode de Newton et la méthode de descente de gradient

### 1.1 La méthode de Newton

La méthode de Newton, initialement conçue pour trouver les zéros d'une fonction, peut être adaptée pour résoudre des problèmes d'optimisation. L'idée de base est de chercher les points où la dérivée d'une fonction s'annule, car ces points correspondent généralement à des extrema (minimum ou maximum) de la fonction. Considérons une fonction  $f(x)$  que l'on souhaite minimiser. On calcule le gradient de  $f(x)$ , noté  $\nabla f(x)$ , qui représente la direction de plus forte pente de la fonction en un point donné. En pratique, on part d'un point de départ  $x_0$  et on calcule la nouvelle estimation  $x_1$  en utilisant la formule :

$$x_{k+1} = x_k - H^{-1}(x_k) \cdot \nabla f(x_k) \quad (1)$$

où :  $H^{-1}(x_k)$  est l'inverse de la matrice Hessienne évaluée en  $x_k$   $k \in \{0 \dots n\}$

Nous rappelons que la méthode de Newton est utilisée sur les fonctions deux fois différentiables et où la matrice hessienne est inversible.

### 1.2 Méthode de descente de gradient

La méthode de descente de gradient est une technique d'optimisation utilisée pour minimiser une fonction en ajustant progressivement ses paramètres dans la direction du gradient opposé. Elle est appliquée sur les fonctions convexes et différentiables. L'algorithme de cette méthode est la suivante :

$$x_{k+1} = x_k - \alpha \cdot \nabla f(x_k) \quad (2)$$

où :  $\alpha$  est le taux d'apprentissage et  $\nabla f(x_k)$  est le gradient de la fonction  $f$  évalué en  $x_k$ .

### 1.3 Comparaison de Convergence

La méthode de Newton converge rapidement si la fonction est deux fois continûment différentiable et si on commence suffisamment près d'un minimum local. Si ce n'est pas le cas, la méthode peut échouer, diverger ou être inefficace. La descente de gradient, quant à elle, est plus robuste et peut converger dans des situations où la méthode de Newton échoue, mais elle est souvent plus lente.

Par ailleurs, la méthode de Newton nécessite le calcul et l'inversion de la matrice hessienne (matrice des dérivées secondes), ce qui est coûteux pour les problèmes de grande dimension. Dans ce cas, la descente de gradient peut être plus pratique et efficace.

## 2 Présentation des résultats

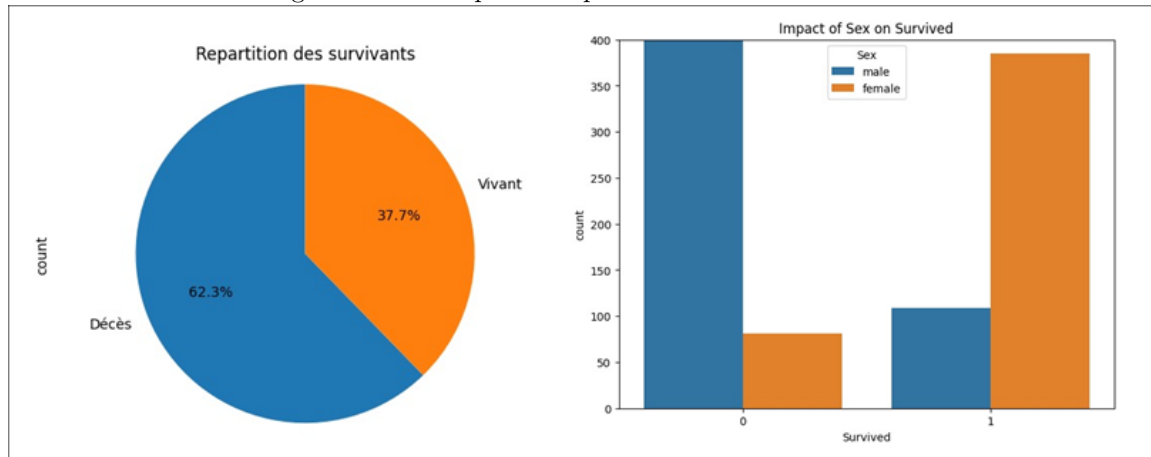
Afin de résoudre le problème soumis, nous avons utilisées 3 bases réelles (Titanic, Boston, Digit ) tous provenant respectivement de Kaggle et Scikit-learn afin de voir la méthode qui converge le plus rapidement selon les cas .

### 2.1 Le dataset de titanic

Le dataset Titanic est un jeu de données emblématique en science des données, utilisé pour l'analyse exploratoire et les modèles de classification. Il regroupe des informations sur les passagers du Titanic comprenant 1309 lignes et 12 colonnes, notamment leur âge, leur sexe, leur classe de voyage, le prix de leur billet et leur port d'embarquement. Il sera question de minimiser la fonction coût qui est l'opposée de la fonction de vraisemblance de prédiction du statut d'un individu (décès ou survivant).

Après traitement de la base de donnée, on remarque qu'il y a plus de décès que de survivant soit 62,3%. Parmi les décès, il y a plus d'hommes que de femmes. Cependant, on observe moins d'hommes survivants que de femmes survivantes.

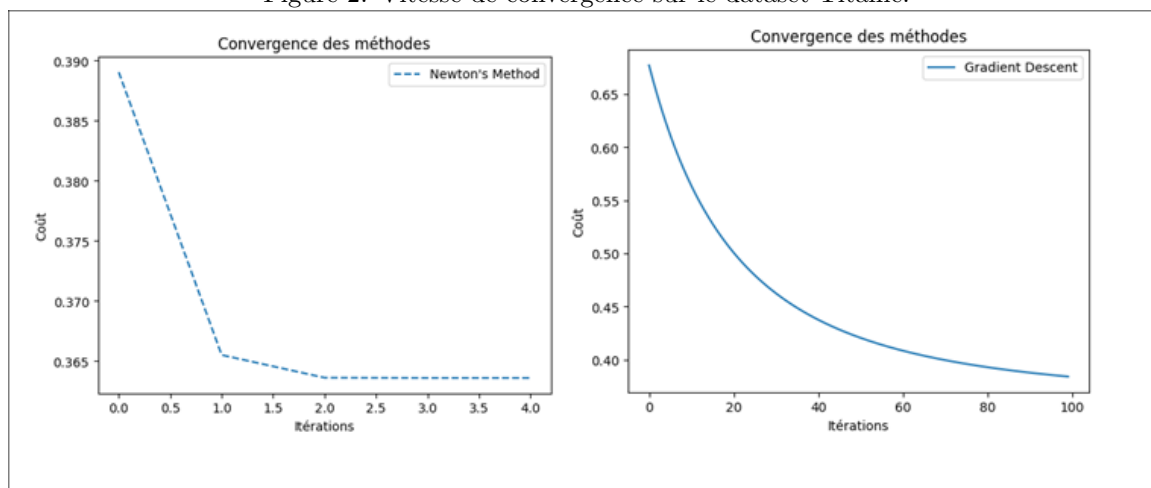
Figure 1: Statistique descriptive sur le dataset Titanic.



Source : kaggle

L'implémentation de la méthode de la descente gradient et celle de Newton avec une fonction objective logistique (Figure 2) révèle qu'à la **deuxième itération**, la méthode Newton converge ce qui n'est pas le cas de la méthode de la descente de gradient où la convergence s'observe après **100 itérations avec un taux d'apprentissage  $\gamma = 0,1$**

Figure 2: Vitesse de convergence sur le dataset Titanic.



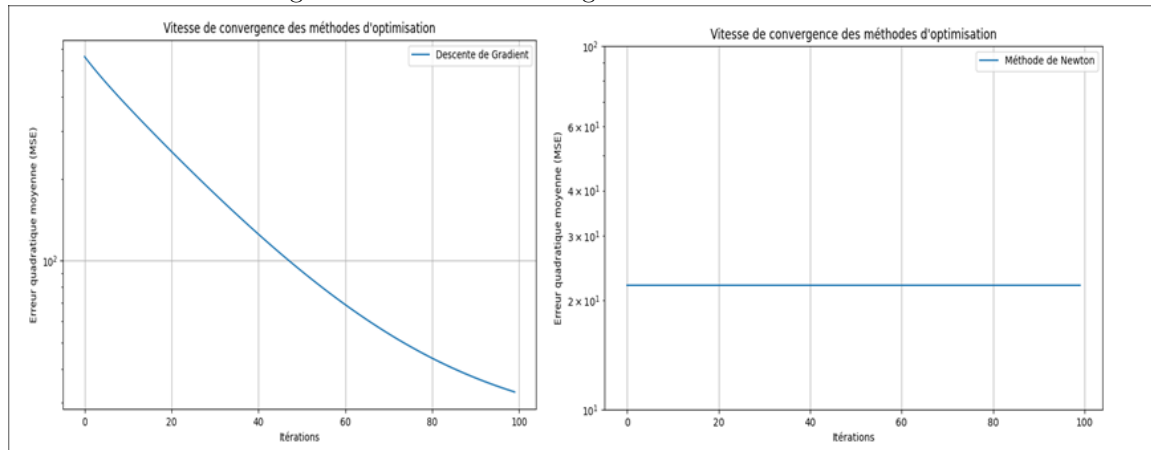
Source : kaggle

## 2.2 Le dataset de Boston

La base de données Boston de scikit-learn est un jeu de données classique contenant 506 quartiers de la ville de Boston, utilisé pour la régression linéaire, notamment pour prédire le prix des maisons en fonction de plusieurs caractéristiques. En appliquant la méthode de Newton et descente de gradient avec une fonction objectif quadratique on obtient le résultat ci dessous (Figure 3). Il contient des informations sur, avec pour objectif de prédire le prix médian des maisons en fonction de différentes

On note une **baisse lente de la vitesse convergence** (erreur quadratique moyenne) pour la méthode de gradient et une **vitesse constante** pour la méthode de Newton.

Figure 3: Vitesse de convergence sur le dataset Boston.

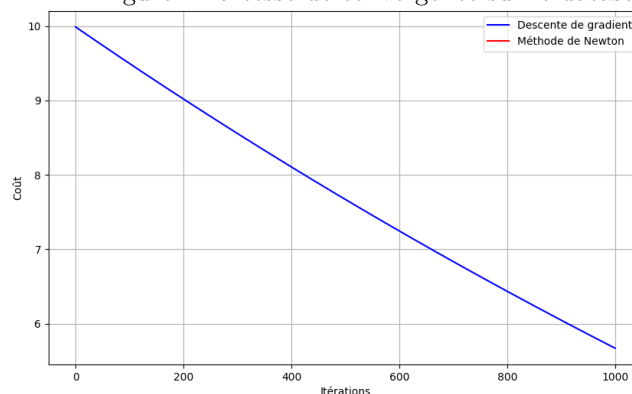


Source : kaggle

## 2.3 Le dataset de digit

La base de données Digits de scikit-learn est un jeu de données utilisé pour la classification d'images de chiffres manuscrits (0 à 9) contenant 1797 images de chiffres, chaque image est de taille 8x8 pixels, ce qui donne un total de 64 caractéristiques par image. En appliquant le méthode de Newton et descente de gradient sur cette base avec la fonction sigmoïde, on observe que **la matrice Hessienne n'est pas inversible** et par conséquent la **méthode de Newton ne converge pas**.

Figure 4: Vitesse de convergence sur le dataset Boston.



Source : kaggle

On observe seulement la convergence de Newtonne car la sklearn

## 3 Conclusion

Il était question dans notre travail de répondre à la problématique suivant laquelle si la méthode de Newton converge toujours plus rapidement que la méthode de gradient. Pour ce faire, il était question de mettre en exergue trois bases (Titanic, Boston, Digit) en utilisant comme fonction objective une régression logistique et une fonction quadratique. Il ressort des trois résultats que **la méthode de Newton converge plus rapidement avec la base de données du Titanic, et ne converge pas avec la base de données digit**. Ces résultats corroborent avec la théorie d'optimisation de ces deux méthodes du fait que la méthode de Newton est efficace et converge plus vite dans les datasets de petites tailles. Ainsi, nous pouvons conclure que la méthode de Newton **ne converge pas toujours plus rapidement** que la descente de gradient.