E-ISSN: 2583-9667
Indexed Journal
Peer Reviewed Journal
https://multiresearchjournal.theviews.in

Received: 03-04-2023
Accepted: 05-06-2023

INTERNATIONAL JOURNAL OF ADVANCE RESEARCH IN MULTIDISCIPLINARY

Volume 1; Issue 1; 2023; Page No. 926-929

To describe the newton's method in numerical optimization

[1]Purushottam Ji and [2]Dr. Ajay Singh

[1]Research Scholar, Department of Mathematics, Shri Krishna University, Chhatarpur, Madhya Pradesh, India
[2]Associate Professor, Department of Mathematics, Shri Krishna University, Chhatarpur, Madhya Pradesh, India

DOI: https://doi.org/10.5281/zenodo.15077220

Corresponding Author: Purushottam Ji

Abstract

In order to incorporate information criteria, this study expands the mathematical formulation using an existing method from the literature. By striking a balance between prediction accuracy and model complexity, these criteria aim to tackle overfitting, a prevalent issue in supervised learning tasks. Many of the desirable features of unconstrained Newton techniques are also present in projected Newton methods for optimisation over convex sets. Numerical solutions to equations can be effectively achieved using the Newton-Raphson method, sometimes known as the Newton Method. Its foundation is the elementary principle of linear approximation, The goal of this research is to fill a need in the existing literature on machine learning and data mining by creating new supervised learning algorithms that can handle classification and regression problems.

Keywords: Newtons, method, optimization, regression and machine

Introduction

For many machine learning issues, this approach works well, particularly in cases where f represents a loss and r represents a regularizer. But until we have some theory under our belts, let's put (5) instances on hold. We suggest addressing (5) using second-order techniques of the Newton type, which are known to perform well on problems without constraints. We may also think of Newton-type algorithms for restricted problems, which, much like their unconstrained counterparts, iteratively minimise a quadratic approximation to the goal, but with restrictions applied. This concept, known as a projected Newton technique, originates from Levitin and Polyak (1966, §7).

Many of the desirable features of unconstrained Newton techniques are also present in projected Newton methods for optimisation over convex sets. As an example, Bertsimas (1999) [13] cites the following results: global convergence shown under a form of the Armijo condition; high local convergence rates around local minima meeting strong convexity; and iterations guaranteed to increase the objective function given a small enough step size. Similarly, we may think of projected quasi-Newton techniques, which involve approximating the Hessian matrix by interpolating differences in parameter and gradient values.

What follows is an outline of the rest of this chapter. Beginning with the case when $r(x)=0$, we limit our attention to smooth optimisation. In Section 6, we address fundamental implementation concerns including Hessian approximation and line-search, and we provide projected Newton-type algorithms for this context. Following that, in Section 7, we go over two-metric projection techniques, and in Section 8, we cover inexact or truncated projected-Newton methods. At last, we go over the case of the nonsmoothed setting, $r(\boldsymbol{x}) \neq 0,$ whereby two approaches similar to Newton's are detailed in Section 9.

Despite this, a local super linear convergence rate is often enjoyed by the more costly Newton step. A resource-intensive precise Newton step is frequently necessary, particularly when calculating the exact Hessian is costly, notwithstanding its theoretical benefits. The well-known quasi-Newton approximation is based on the notion of approximating the Hessian, which is often used to avoid some of the related computing problems. So, let's take a quick look at the BFGS upgrade that gets close to the identical Hessian again.

Consequently, we may find that projecting the Newton step under the Euclidean norm is more suitable for optimizations across these domains. The famous two-metric projection technique really makes this decision; the name comes from the fact that it employs two separate matrices (metrics) to scale the gradient and calculate the projection. Here we will examine Bertsekas's projected Newton technique (Bertsekas, 1982) [7] as an example, specifically in the context when Ω has bound restrictions.

Literature Review

Lin, Zhoushan. (2019) [1]. Because ML relies on effectively solving mathematical models, optimization is an essential component of ML. However, fresh momenta and optimization ideas may also be provided via machine learning. Improving the efficiency of optimization and machine learning interactions is the primary goal of this research.

Khac hay, Michael *et al*. (2019) [2]. The 18th International Conference on Mathematical optimization Theory and Operations Research (MOTOR 2019) [2], which took place in Ekaterinburg, Russia, from July 8th to the 12th, 2019, is collected in this book. The proceedings of the conference were reviewed and accepted. Mathematical programming, discrete optimization, complexity theory, combinatorial algorithms, optimal control, games, and their applications to pertinent practical problems in data analysis, mathematical economy, and operations research are all part of the conference's scope.

Gambell a, Claudio *et al*. (2019) [3]. Machine learning as optimisation models is introduced in this work, which also provides a literature review on the topic. The development of numerical optimisation methods, which have had a significant impact in several machine learning contexts, may improve these models. In particular, we provide mathematical optimisation models for popular ML techniques like deep neural networks, clustering, classification, and regression, as well as for techniques that are only starting to gain traction in fields like machine education and empirical model learning. We highlight possible future research avenues and examine the benefits and drawbacks of various models.

Padamwar, Badri *et al*. (2019) [4]. Optimisation approaches in Operations Research (OR) are thoroughly examined in this review article, which aims to shed light on their relevance, current state of the art, obstacles, and potential future developments. Decisions in many different sectors may be improved with the use of mathematical and analytical tools developed in the field of operations research. The first part of the article provides a general introduction to OR, touching on its background, scope, and definition. From then, it moves on to the basics of optimisation, discussing various techniques and issues in the field.

Lynch, Matthew *et al*. (2019) [5]. Mechanical parts and systems may benefit greatly from the latest developments in design optimisation, which might greatly enhance their performance. One class of numerical approaches that is having an impact on the mechanical design of newly introduced hardware is topology optimisation, which, when combined with additive manufacturing, produces algorithmically created optimized designs. Tuning parameters govern algorithmic function and convergence; however, many of these algorithms need substantial human setup and control. To save users the trouble and expense of manual tuning, this study presents a system that uses machine learning techniques to suggest tuning settings. Using a Bayesian optimisation strategy, the method retrieves tuning parameters from a database of previously solved, comparable issues that were evaluated using a dissimilarity measure derived from problem information.

Research Methodology

Numerical solutions to equations can be effectively achieved using the Newton-Raphson method, sometimes known as the Newton Method. Its foundation is the elementary principle of linear approximation, as is true of most of differential calculus. When applied correctly, the Newton Method may often resolve complex problems by zeroing down on specific causes.

The function f(x) must be well-behaved, and the equation f(x) = 0 must have a root r. An initial guess of r, denoted as x0, is used. We derive a better (hopefully) estimate of x1 from x0. We derive a fresh estimate x2 from x1. A revised estimate, x3, is generated from x2. We keep going until we reach a point where we can turn around, or until it becomes obvious that we are not making any progress. Iterative describes the aforementioned broad approach to problem solving. The Newton-Raphson technique is the most used iterative root-finding process because it is both simple and powerful.

Offer your thoughts. Although most mathematicians prefer to begin counting at zero, the initial estimate is occasionally referred to as x1. The first approximation is sometimes referred to as a "guess." While the Newton Method performs admirably when x0 is near r, it can provide appalling results when it is far from it. It is important to carefully select the "guess" x0.

Data Analysis: A non-zero solution to the equation x = 2sin x can be found by applying the Newton Method. Prove that f(x) = x − 2sin x. At that point, the Newton-Raphson iteration is equal to 1 − 2cos x.

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n - 2\sin x_n}{1 - 2\cos x_n} = \frac{2(\sin x_n - x_n \cos x_n)}{1 - 2\cos x_n}.$$

With x0 = 1.1, we can. Here are six more estimates, all to the nearest three digits:

$$\begin{array}{llllll} x_1 = & 8.453 & x_3 = & 203.384 & x_5 = & -\ 87.471 \\ x_2 = & 5.256 & x_4 = & 118.019 & x_6 = & -203.637. \end{array}$$

Things don't look good, and they get worse. It turns out that x35 < −64000000. We could be stubborn and soldier on. Miracles happen-but not often. (One happens here, around n = 212.)

In order to have a sense of the problem, plot the equation y = x2sin x using a graphing application. Keep in mind that the point where the tangent line at xn meets the x-axis is xn+1. The bumps on the line y = x 2sin x severely muddle the Newton Method.

The choice of x0 = π/3 ≈ 1.0472 is disastrous since it results in 1 − 2cos x0 = 0, which means that x1 does not exist. A (relatively) flat section of the curve serves as our beginning point, with x0 = 1.1. Following the tangent line will lead us to x1, which is spatially far from x0. X2 is located far away from x1 since both x1 and x2 are on flat portions of the curve. The wild ride goes on, even if x2 is on a flatter section of the curve.

The selection of x0 was the root cause of the problem. Perhaps we can improve upon this. The lines y=x and y= 2sin x should be drawn. A cursory drawing reveals that they intersect little after π/2. However, we shall assume x0 = 1.5 carelessly. Our calculations were done up to 50 places, and here are the next six estimations, up to 19 places.

$$x_1 = 2.0765582006304348291 \qquad x_4 = 1.8954942764727706570$$
$$x_2 = 1.9105066156590806258 \qquad x_5 = 1.8954942670339809987$$
$$x_3 = 1.8956220029878460925 \qquad x_6 = 1.8954942670339809471$$

For the first 19 iterations, and even for the first 32, the value of x7 agrees with x6, and the real root is identical to x6 up to the 32nd iteration.

The data needed by the optimizations models outlined is collected during the pre-processing stage. Using a software called gdxrrw, the data will be converted into the suitable GAMS format called gdx (Jain and Dirkse, 2018) [9].

Following the selection of a technique, the implementation invokes one of two GAMS scripts that identify the partitioning variable and, subsequently, determine the best number of regions. After that, we bring the results back into R and format them so they're easier to understand. Additionally, a bespoke function that uses R only to make predictions on fresh data has been developed.

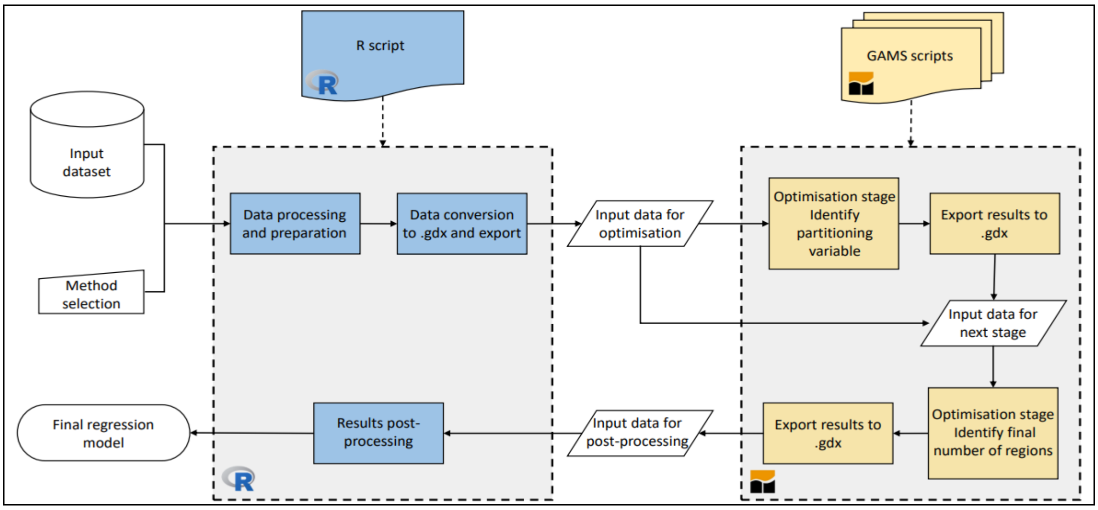

Fig 1: The implementation of the piecewise regression algorithm

Through the use of a literary example, the final form of the regression equations is demonstrated. This image is about gasoline's octane rating. This dataset looks at the octane rating of fuel in particular as it goes through refining. Wood (1973) [14] states that the three raw materials utilized to determine the fuel's rating are A1, A2, and A3, and that A3 is a variable that quantifies the refinery's production circumstances.

Table 1: Final regression functions for some of the proposed methods

Method	Regression functions
PRIA	$Y = \begin{cases} -5.13 \cdot A_1 + 0.11 \cdot A_2 - 0.71 \cdot A_3 + 1.95 \cdot Q + 95.71, & 0 \leq A_3 \leq 0.58 \\ -7.57 \cdot A_1 - 0.97 \cdot A_2 - 3.70 \cdot A_3 + 3.96 \cdot Q + 98.83, & 0.58 < A_3 \leq 0.71 \\ -8.13 \cdot A_1 - 1.80 \cdot A_2 - 1.82 \cdot A_3 + 2.05 \cdot Q + 99.00, & 0.71 < A_3 \leq 0.92 \\ 23 \cdot A_1 + 2.95 \cdot A_2 + 13.90 \cdot A_3 + 6.48 \cdot Q + 59.15, & 0.71 < A_3 \leq 1 \end{cases}$
PROA	$Y = \begin{cases} -5.13 \cdot A_1 + 0.11 \cdot A_2 - 0.71 \cdot A_3 + 1.95 \cdot Q + 95.71, & 0 \leq A_3 \leq 0.58 \\ -7.79 \cdot A_1 - 1.19 \cdot A_2 - 0.1.07 \cdot A_3 + 3.11 \cdot Q + 97.71, & 0.58 < A_3 \leq 1 \end{cases}$

The various regression models generated by the iterative and single-level techniques utilizing AIC are shown in Table 1.

Several real-world datasets have been utilized to evaluate the suggested methodologies. Table 2 displays datasets that have been collected from various internet sources. The datasets package in R provides access to the pharmacokinetics and seismic data, StatLib to the bodyfat and sensory data (Vlachos, 2005) [15], OpenMV.net to the distillation data, and the UCI machine learning repository to the remainder.

Table 2: Regression datasets examined in piecewise regression work

Dataset	No. samples	No. variables
Pharmacokinetics	132	4
Bodyfat	252	14
Distillation	253	26
Yacht Hydrodynamics	308	6
Sensory	576	11
Cooling efficiency	768	8
Heating efficiency	768	8
Earthquake	1000	4
Concrete	1030	8
White wine quality	4898	11

In order to assess the efficiency of sailing boats and determine the amount of propellant needed, the yacht hydrodynamics set makes predictions about their residuary resistance. According to the energy efficiency dataset (Tsanas and Xifara, 2012) [16], there are eight factors that determine how much heating and cooling a building need. Predicting the compressive strength of structural concrete is the goal of the concrete dataset (Yeh, 1998) [17]. Cortez *et al.*

(2009) [18] found that the wine dataset can use the features of white wine to make quality predictions.

Conclusion

The goal of this research is to fill a need in the existing literature on machine learning and data mining by creating new supervised learning algorithms that can handle classification and regression problems. Overarchingly, this work is defined by a number of suggested algorithms that build upon and supplement previous work in the field. Despite this, a local super linear convergence rate is often enjoyed by the more costly Newton step. A resource-intensive precise Newton step is frequently necessary, particularly when calculating the exact Hessian is costly, notwithstanding its theoretical benefits. Numerical solutions to equations can be effectively achieved using the Newton-Raphson method, sometimes known as the Newton Method. In order to incorporate information criteria, this study expands the mathematical formulation using an existing method from the literature

References

1. Lin Z. How Can Machine Learning and Optimization Help Each Other Better? Journal of the Operations Research Society of China. 2019;8:1-12. doi:10.1007/s40305-019-00285-6.
2. Khachay M, Kochetov Y, Pardalos P. Mathematical Optimization Theory and Operations Research. In: 18th International Conference, MOTOR 2019, Ekaterinburg, Russia, July 8–12, 2019 Proceedings. LNCS Vol. 11548. 2019.
3. Gambella C, Ghaddar B, Naoum-Sawaya J. Optimization Problems for Machine Learning: A Survey. 2019.
4. Padamwar B, Pandey H. Optimization Techniques in Operations Research: A Review. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 2019;10(1):746-752. doi:10.61841/turcomat.v10i1.14604.
5. Lynch M, Sarkar S, Maute K. Machine Learning to Aid Tuning of Numerical Parameters in Topology Optimization. Journal of Mechanical Design. 2019;141(1):1-11. doi:10.1115/1.4044228.
6. Liu R. Optimization Model of Mathematics Instructional Mode Based on Deep Learning Algorithm. Computational Intelligence and Neuroscience. 2022;2022:1817990. doi:10.1155/2022/1817990.
7. Roberts T, Tonna S. Optimizing Parameters for Machine Learning Models and Decisions in Production. In: Wiley Online Library; c2022. doi:10.1002/9781119824961.ch8.
8. Thouheed Ahmed SY, Basha SM. Mathematical Principles in Machine Learning. 2023.
9. Snyman J, Wilke D. Practical Mathematical Optimization: Basic Optimization Theory and Gradient-Based Algorithms. Springer Optimization and Its Applications. 2018;133:1-15. doi:10.1007/978-3-319-77586-9.
10. Almosa N, Al-Jilawi A. Developing Mathematical Optimization Models with Python. AIP Conference Proceedings. 2023;2591:020030. doi:10.1063/5.0119585.
11. Nagy-Bota S, Moldovan L, Nagy-Bota MC, Varga I. Mathematical Models Used in the Optimizations of Supply Chains. Acta Marisiensis Ser Technological. 2023;20:27-31. doi:10.2478/amset-2023-0005.
12. Yang L, Shami A. On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice. 2020. Available from: https://arxiv.org/abs/2007.15745.
13. Bertsimas D, Darnell C, Soucy R. Portfolio construction through mixed-integer programming at Grantham, Mayo, Van Otterloo and Company. Interfaces. 1999;29(1):49-66.
14. Wood FS. The use of individual effects and residuals in fitting equations to data. Technometrics. 1973;15(4):677-695.
15. Vlachos M, Yu P, Castelli V. On periodicity detection and structural periodic similarity. InProceedings of the 2005 SIAM international conference on data mining Society for Industrial and Applied Mathematics; c2005. p. 449-460.
16. Tsanas A, Xifara A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and buildings. 2012;49:560-567.
17. Yeh IC. Modeling of strength of high-performance concrete using artificial neural networks. Cement and Concrete research. 1998;28(12):1797-1808.
18. Cortez P, Cerdeira A, Almeida F, Matos T, Reis J. Modeling wine preferences by data mining from physicochemical properties. Decision support systems. 2009;47(4):547-553.