Taylor Series and Applications in ML

Mathematics and Statistics Group Project

Karina Bakhtiyarova

Hongtao Li

16/05/2024

1 Introduction to Taylor Series

A Taylor series is a way to represent a function using an infinite series around a particular point. It's called the Taylor series because it was developed by a mathematician named Brook Taylor. The formula for a Taylor series of a function f(x) centered at a is a general form that can be used to approximate the function.

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$$

Here, f(a) is the value of the function at the center a, f'(a) is the first derivative of the function evaluated at a, f''(a) is the second derivative evaluated at a, and so on. The terms involving derivatives capture the behavior of the function at and around point a.

The Taylor series allows us to approximate a function using a polynomial, which can be useful for various mathematical calculations and analyses, particularly in areas like calculus, differential equations, and numerical analysis.

2 Historical context of Taylor Series

Brook Taylor created the Taylor series in 1715, expanding on the contributions of mathematicians such as Fermat, Newton, and Leibniz in the field of calculus. This series allowed functions to be represented as infinite series of variable powers, which proved to be a crucial tool in both mathematics and science.

2.1 Ancient Roots

The idea of approximating functions using polynomial series can be traced back to ancient Greek mathematicians like Ptolemy, who used series expansions for trigonometric functions.

2.2 Madhava and Kerala School

In the 14th century, the Indian mathematician Madhava of Sangamagrama and the Kerala School of Mathematics developed methods for expressing trigonometric functions as infinite series, including what would later be recognized as the Taylor series for sine, cosine, and arctangent functions.

2.3 Renaissance Europe

In Europe during the Renaissance, scholars like Isaac Newton and Gottfried Wilhelm Leibniz independently developed calculus. Newton used series expansions in his work on calculus, including what is now known as the binomial series, which is a special case of the Taylor series.

2.4 Brook Taylor

Taylor's contribution was to formalize and generalize the concept of series expansions in his 1715 work "Methodus Incrementorum Directa et Inversa" (Direct and Inverse Methods of Incrementation). He presented a method for expressing a wide range of functions as infinite series, now known as Taylor series, which became fundamental in calculus and mathematical analysis.

The Taylor series has since become a cornerstone of mathematical analysis, providing a powerful tool for approximating and studying functions in various mathematical contexts.

3 Importance and applications in Mathematics

The Taylor series holds significant importance in mathematics due to its wide range of applications.

3.1 Function Approximation

Taylor series are used to approximate complex functions with simpler polynomial expressions. This is particularly useful in calculus and numerical analysis for estimating function values and solving differential equations.

3.2 Error Analysis

In numerical methods, Taylor series are employed to analyze the error of approximation techniques. By comparing the Taylor series expansion with the actual function, mathematicians can quantify and understand the accuracy of numerical computations.

3.3 Derivatives and Integrals

Taylor series provide a systematic way to compute derivatives and integrals of functions. By taking derivatives of the Taylor series representation, one can obtain expressions for higher-order derivatives of the original function.

3.4 Physics and Engineering

Taylor series are super important in physics and engineering because they're used to model all sorts of cool stuff that happens in the real world. They're especially handy when it comes to making complicated systems simpler and easier to understand. By using Taylor series, scientists and engineers can figure out how things will behave when they're not quite in balance, and they can even predict how things will change over time. It's like having a crystal ball for math and science!

3.5 Probability and Statistics

The importance of Taylor series in statistics consists in helping us figure out the properties of probability distributions and statistical estimation techniques. We use them in moments generating functions, which are like the secret sauce that helps us understand and analyze data.

3.6 Machine Learning and Optimization

Taylor series are used in optimization algorithms because they help us estimate objective functions and gradients. This estimation is crucial for iterative optimization methods such as gradient descent.

3.7 Signal Processing

The Taylor series are super useful in signal processing because they help us estimate signals and study their frequency parts. They come in handy for Fourier analysis and making digital filters. Taylor series are so versatile that they're essential in math, science, and engineering, giving us awesome tools for analyzing, modeling, and calculating.

4 Theoretical Foundations of Taylor Series

4.1 Basic concepts: Polynomial approximation

The Taylor series is a mathematical tool that allows us to approximate functions using polynomials. Specifically, it represents a function $f(x)$ as an infinite sum of terms involving powers of $(x-a)$, where a is a point around which we want to approximate the function.

The basic idea is to start with a known function $f(x)$ and its derivatives at the point $x=a$. These derivatives give us information about how the function changes at that point. The Taylor series then constructs a polynomial that matches the function and its derivatives up to a certain order near $x=a$.

The general form of the Taylor series for a function $f(x)$ around $x=a$ is:

$$f(x)=f(a)+f'(a)(x-a)+\frac{f''(a)}{2!}(x-a)^2+\frac{f'''(a)}{3!}(x-a)^3+\cdots$$

Here, $f'(a)$, $f''(a)$, $f'''(a)$, and so on, represent the derivatives of $f(x)$ evaluated at $x=a$. The terms $(x-a)^n$ are powers of the difference between x and a, which capture how the function changes as x moves away from a.

By including more terms in the series, we get a more accurate polynomial approximation of the function around x=a. This is particularly useful in calculus, physics, engineering, and other areas where precise approximations of functions are needed.

4.2 Derivation of Taylor Series

The Taylor series is obtained using the concept of the Maclaurin series, which being the specific case of the Taylor series with the center point at 0. Here's a basic outline of the derivation.

4.2.1 Maclaurin Series

Start with function f(x) and assume it has derivatives of all orders at x=0. The Maclaurin series of f(x) is given by:

$$f(x)=f(0)+f'(0)x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\cdots$$

This series is essentially the Taylor series with a=0.

4.2.2 General Taylor Series

To derive the Taylor series for a function f(x) around point a, we use a similar approach but with x−a as the variable. Let h=x−a, so x=a+h. Rewrite the Maclaurin series using h:

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \frac{f'''(a)}{3!}h^3 + \cdots$$

This series represents $f(x)$ in terms of h (which is $x - a$) around $x = a$. It's called the Taylor series of $f(x)$ about $x = a$.

4.2.3 Simplification

Often, we rewrite the Taylor series using sigma notation for brevity:

$$f(a+h) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}h^n$$

Here, $f^{(n)}(a)$ denotes the n-th derivative of $f(x)$ evaluated at $x = a$.

This derivation shows how the Taylor series is obtained by expressing a function f(x) as an infinite sum involving its derivatives evaluated at a particular point a, multiplied by powers of x−a (or h, as in this derivation).

5 Taylor series in machine learning

Taylor series plays an important role in machine learning, particularly in optimization algorithms and function approximation.

5.1 Gradient Descent Optimization

One of the key applications of Taylor series in machine learning is in the Gradient Descent Optimization Algorithm. This optimization technique is vital for training machine learning models, as it helps in minimizing the loss function, which is a measure of how far the model's predictions are from the actual results. General gradient descent formula comes from applying the Taylor series expansion to the loss function. Specifically, the gradient descent formula is derived by taking the first-order Taylor series approximation of the loss function around the current parameter values.

Taylor series provides a way to approximate the loss function around a given point using a linear function of the parameters. By taking the derivative of this linear approximation, the gradient descent algorithm can estimate the direction and magnitude of the gradient, guiding the optimization process towards the minimum of the loss function (Howell, 2023). The accuracy of this approximation is crucial for the convergence and efficiency of the optimization algorithm.

5.2 Function Approximation

The Taylor series is also implemented in function approximation which is a fundamental task in machine learning. The power of Taylor series in function approximation lies in its ability to capture the local behavior of a function around a specific point. By using the derivatives of the function, Taylor series can provide a close approximation of the function's value in the vicinity of the chosen point. This property makes Taylor series particularly useful in scenarios where the function is not easily differentiable or when the function's global behavior is complex and difficult to model directly.

Taylor series expansion allows for the approximation of complex functions using a series of simpler polynomial terms. This technique is widely used in numerical computations when estimates of a function's values at different points are required, as demonstrated in the examples provided by Mehreen Saeed (2022). Furthermore, Taylor series can be used to approximate functions that are not easily representable in closed form. By expanding the function around a point of interest and truncating the series at an appropriate order, Taylor series can provide a practical and efficient way to estimate the function's value, even for functions that do not have a simple analytical expression.

5.3 Robustness and Stability

A research conducted by Daniel Pfrommer et al. in 2023 introduced "Taylor Series Imitation Learning" (TaSIL) as a method to enhance the stability and robustness of machine learning models, particularly in imitation learning. TaSIL leverages higher-order Taylor series terms to penalize deviations between learned and expert policies, thereby promoting closer alignment and resulting in more stable policies. By incorporating higher-order terms into the TaSIL framework, the method captures intricate details of the policy space, leading to more nuanced adjustments during the learning process (ibid).

TaSIL extends beyond traditional behavior cloning approaches by considering not only the first-order differences but also higher-order terms, enabling the model to learn from more comprehensive error signals. This approach enhances the model's ability to generalize across diverse scenarios, contributing to improved performance across various tasks in the MuJoCo environment (ibid). The success of TaSIL underscores the significance of integrating mathematical principles into machine learning methodologies. By embracing concepts like Taylor series, researchers can unlock new avenues for enhancing model robustness and performance.

6 Taylor Series in Deep Learning

Taylor series also plays a crucial role in the development and optimization of deep learning models, contributing to neural network architectures, training, and activation function performance.

6.1 Neural Network Architectures

Taylor series is applied in the development of innovative neural network architectures. The work presented by Balduzzi, McWilliams and Butler-Yeoman introduced the concept of neural Taylor approximations, which applies Taylor series to capture the local behavior of the network's activations and gradients, enabling a more stable and interpretable optimization process (2018). This approach provides the first convergence guarantee applicable to modern convolutional networks, which are neither smooth nor convex (ibid). By incorporating Taylor series into the network architecture, the researchers were able to address the challenge of shattered gradients which is a common issue in training deep neural networks with rectifier activations.

Another example is TaylorNet, a generic neural architecture that leverages Taylor series expansion to introduce inductive bias to deep neural networks, which parameterizes Taylor polynomials using DNNs without using non-linear activation functions, aiming to mitigate the curse of dimensionality and improve the stability of model training (Singh, 2023). The key idea behind TaylorNet is to take advantage of the properties of Taylor series, such as its ability to capture local function behavior and its linear structure, to enhance the literal capacity and training stability of deep neural networks, potentially leading to improved performance and generalization in deep learning tasks.

6.2 Activation Function Optimization

Taylor series is also applied to improve the performance of deep learning activation functions. Taylor series can be used to approximate the activation functions, providing a more efficient and differentiable representation. This can lead to improved numerical stability, faster convergence, and potentially better generalization in deep learning models. By incorporating Taylor series-based activation functions, deep learning architectures can benefit from the mathematical properties of Taylor expansion, further advancing the state-of-the-art in deep learning.

For example, Alex discussed in one of his YouTube videos about deriving Taylor polynomials for common deep learning activation functions like the Hyperbolic Tangent, Sigmoid, and Exponential Linear Unit, suggesting that using these Taylor polynomial approximations of the activation functions can help improve the training performance and convergence of neural networks, by addressing issues like the vanishing gradient problem (2022).

7 Limitations and Challenges

While Taylor series is a powerful tool in machine learning and deep learning, it does have some limitations and challenges. One of the primary challenges is the issue of convergence, as Taylor series may not converge for all functions or over the entire domain of the function. Additionally, the accuracy of the approximation can be sensitive to the choice of the expansion point and the order of Taylor polynomial used.

To address these limitations, researchers and developers in machine learning have explored various techniques, such as the use of piecewise polynomial approximations (e.g., splines) and the incorporation of Taylor series into more sophisticated function approximation methods, like neural networks (Pfrommer, et al. 2023). These approaches aim to improve the flexibility and robustness of Taylor series-based approximations, making them more suitable for a wider range of machine learning applications.

References

Alex. (2022) Taylor series for deep learning activation functions. Available at: https://youtu.be/VWnZK2w8pCc (Accessed: 10 May 2024)

Balduzzi, D., McWilliams, B. & Butler-Yeoman, T. (2018) "Neural Taylor Approximations: Convergence and Exploration in Rectifier Networks." Available at: https://doi.org/10.48550/arxiv.1611.02345.

Howell, E. *Why do we even have neural networks?* Available at: https://towardsdatascience.com/why-do-we-even-have-neural-networks-72410cb9348e (Accessed: 08 May 2024)

Pfrommer, D., Zhang, T., Tu, S. & Matni, N. (2023) "TaSIL: Taylor Series Imitation Learning." Available at: https://doi.org/10.48550/arxiv.2205.14812.

Saeed, M. A Gentle Introduction to Taylor Series. Available at: https://machinelearningmastery.com/a-gentle-introduction-to-taylor-series/ (Accessed: 08 May 2024)

Singh, A. Taylor Series Approximation: Definition, Formula, and Example. Available at: https://www.shiksha.com/online-courses/articles/taylor-series-approximation-definition-formula-and-example/ (Accessed: 08 May 2024)