

МЕТОД ГРАДІЄНТНОГО СПУСКУ В ЗАДАЧАХ НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ

Вінницький національний технічний університет

Анотація

У роботі описано математичні засади методу градієнтного спуску, який є одним із найважливіших алгоритмів оптимізації у задачах навчання штучних нейронних мереж. Розглянуто принципи обчислення градієнта функції втрат, механізм ітераційного коригування параметрів моделі та особливості використання різних модифікацій алгоритму у сучасних системах машинного навчання.

Ключові слова: метод градієнтного спуску, функція втрат, частинні похідні, машинне навчання.

Abstract

The paper describes the mathematical foundations of the gradient descent method, which is one of the most important optimization algorithms used in training artificial neural networks. The principles of gradient computation, the iterative process of updating model parameters, and the practical aspects of applying different variations of the algorithm in modern machine learning systems are discussed.

Keywords: gradient descent method, loss function, partial derivatives, machine learning.

Вступ

Сучасні методи машинного навчання ґрунтуються на задачах багатовимірної оптимізації. Одним із таких методів є метод градієнтного спуску. Ідея цього методу полягає у поступовому наближенні до точки мінімуму функції шляхом руху в напрямку, протилежному до градієнта. Навчання нейронної мережі формально зводиться до мінімізації функції втрат (loss function), яка характеризує відхилення прогнозованих значень від еталонних.

Результати дослідження

Нехай задано функцію

$$L(w) \rightarrow \min,$$

де $w = (w_1, w_2, \dots, w_n)$ – вектор параметрів або ваг нейронної мережі, а $L(w)$ – функція втрат [1].

Для знаходження напрямку, в якому значення функції зменшується, використовується градієнт цієї функції. Градієнт функції – це вектор частинних похідних по кожному із параметрів моделі, який спрямований у бік максимальної зміни функції:

$$\Delta L(w) = \left(\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \dots, \frac{\partial L}{\partial w_n} \right).$$

Градієнт функції втрат допомагає визначити, як зміна того чи іншого параметра впливає на значення функції. Тобто він вказує, як потрібно змінити параметри моделі, щоб втрата була меншою.

В контексті нейронних мереж ітераційна формула методу градієнтного спуску має вигляд [2]:

$$w^{k+1} = w^k - \eta \nabla L(w^k),$$

де $\eta > 0$ – коефіцієнт швидкості навчання (learning rate), що визначає величину кроку при переході від однієї ітерації до іншої.

Геометрично метод градієнтного спуску інтерпретується як послідовний рух по багатовимірній поверхні функції втрат в напрямку мінімуму, тобто протилежному до градієнта функції. Параметри моделі оновлюються на кожній ітерації та рухаються у напрямку, протилежному градієнту функції втрат, поки ми не знайдемо точку у просторі параметрів, де втрата на навчальних даних буде мінімальною. Вибір коефіцієнта η суттєво впливає на швидкість збіжності алгоритму. Чим меншим ми його

візьмемо, тим повільніше відбуватиметься процес навчання мережі. Він вибирається експериментально (наприклад, 0.1, 0.01, 0.001 і так далі), поки результат не буде задовільним. Мета алгоритму градієнтного спуску – мінімізувати критерій якості нейронної мережі – суму квадратів втрат навчальної вибірки:

У задачах навчання нейронних мереж обчислення градієнтів виконується за допомогою алгоритму зворотного поширення втрат (backpropagation), який ґрунтується на правилі диференціювання складеної функції. Завдяки цьому стає можливим навчання багат шарових нейронних мереж з великою кількістю параметрів.

Залежно від способу використання навчальної вибірки розрізняють пакетний (Batch Gradient Descent), стохастичний (Stochastic Gradient Descent) та міні-пакетний (Mini-batch Gradient Descent) градієнтний спуск. Стохастичні модифікації дозволяють зменшити обчислювальні витрати та пришвидшити навчання моделей.

Пакетний градієнтний спуск передбачає знаходження градієнта на основі всієї навчальної вибірки перед кожним оновленням параметрів. Такий підхід забезпечує точний напрям оптимізації, але може бути не ефективним для великих обсягів даних.

Стохастичний градієнтний спуск оновлює параметри моделі після обробки кожного окремого прикладу з навчальної вибірки. Цей підхід значно пришвидшує навчання та дозволяє уникати деяких локальних мінімумів, проте може спричиняти значні коливання значень функції втрат.

Міні-пакетний градієнтний спуск є проміжним варіантом між двома іншими. У цьому випадку градієнт обчислюється для невеликих підмножин навчальних даних. Такий підхід поєднує стабільність пакетного методу та швидкість стохастичного, тому широко застосовується на практиці.

У сучасних системах машинного навчання застосовуються удосконалені алгоритми (Momentum, RMSProp, Adam, які враховують інформацію про попередні значення градієнтів, що дозволяє прискорювати навчання.

Однією з важливих проблем оптимізації у багатовимірних просторах є наявність локальних мінімумів, плато та сідлових точок. У таких випадках класичний метод градієнтного спуску може сповільнюватися або тимчасово зупинятися. Використання ж стохастичних методів та адаптивних оптимізаторів дозволяє частково подолати ці труднощі та покращити процес навчання.

Висновки

Метод градієнтного спуску є фундаментальним інструментом оптимізації у машинному навчанні. Завдяки простоті реалізації та універсальності, метод став базовим інструментом для розв'язання задач мінімізації функцій багатьох змінних. Попри відносну повільність у класичному вигляді, численні модифікації дозволили зробити його основою сучасних алгоритмів машинного навчання. То ж сучасні дослідження спрямовані на розробку нових методів оптимізації, які спрямовані на підвищення швидкості збіжності, стійкості до локальних мінімумів та ефективності в умовах великої розмірності простору параметрів.

СПИСОК ВИКОРИСТАНИХ ЛІТЕРАТУРИ

1. Ruder S. An overview of gradient descent optimization algorithms [Electronic resource]. – 2016. – Available at: <https://arxiv.org/abs/1609.04747>.
2. Bishop C. Pattern recognition and machine learning. – New York : Springer, 2006. 798 p.

Кабаровський Ілья Олександрович – студент групи ЗПІ-25б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця, email: kabarovskiyilia@gmail.com.

Прозор Олена Петрівна – к.пед.н., доцент кафедри вищої математики, Вінницький національний технічний університет, Вінниця, email: prozor@vntu.edu.ua.

Kabarovskyi Iliia O. – student of group ZPI-25b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email: kabarovskiyilia@gmail.com.

Prozor Olena P. – PhD (in Pedagogical Sciences), Docent, Associate Professor of the Department of Higher Mathematics, Vinnytsia National Technical University, Vinnytsia, email: prozor@vntu.edu.ua.