

Research. – 2014. – Volume 4 No.7. – с. 284-287

2. Цегелик Г.Г. Апарат некласичних мажорант і діаграм Ньютона функцій, заданих таблично, та його використання в чисельному аналізі: монографія. – Львів: ЛНУ імені Івана Франка, 2013. – 190с.

3. Глебена М.І. Математичні моделі та числові методи мажорантного типу для аналізу дискретних оптимізаційних процесів: автор. дис. на здобуття наук. ступеня канд. фіз.- мат. наук: спец. 01.05.02 “Математичне моделювання та обчислювальні методи” / М.І. Глебена. – Івано-Франківськ, 2012. – 23 с.

4. Глебена М.І. Апарат некласичних мінорант Ньютона та його використання / М.І. Глебена, Г.Г. Цегелик // Наук. вісн. Ужгород. ун-ту. Сер. матем. та інформ.. – 2013. – Вип. 24.-N1. – С.16-21.

ЯХНО В.М., ПРОСТИБОВЕНКО С. С., РУБАН А. О..

ЕКСПЕРИМЕНТАЛЬНЕ ОБГРУНТУВАННЯ ЯКОСТІ ГРАДІЄНТНИХ МЕТОДІВ НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ

JAKHNO V.M., PROSTIBGENKO S.S., RUBAN A.O.

EXPERIMENTAL JUSTIFICATION OF THE QUALITY OF GRADIENT METHODS FOR LEARNING NEURON NETWORKS

Purpose and objectives. . The greatest interest in the gradient method in recent years is due to the fact that gradient descents and their stochastic or randomized variants underlie almost all modern learning algorithms developed in data analysis [1]. Most optimization algorithms come from propositions that have access to an exact gradient or hessian. In practice, there is usually only a noisy or even biased estimate of these values. Almost all deep learning algorithms are described on sample estimates, by extreme networks, in terms of the use of mini-packages of learning examples to calculate the gradient. It also happens that the objective function that minimizes is one that has no computational solution. In this case, there is usually no computational solution to the gradient calculation problem, and then only the approximate gradient remains. Such problems most often arise in complex models, for example, the algorithm of the compared distribution (contractual divergence) offers a method of approximation of gradient functions of logarithmic plausible mechanical engineering. Various neural network optimization algorithms have been developed to compensate for inaccurate gradient estimation. The problem can also be found by choosing a surrogate loss function that is even more approximate than true. In any case, gradient descents and their stochastic variants underlie almost all modern learning algorithms. The purpose of the study is to develop a software tool that allows you to investigate and compare experimentally the effectiveness of different options for software implementation of algorithms that use the gradient as the direction of descent. The options involve a variety of methods for constructing descent directions using a gradient direction..

Object and subject of research. The object of research is the features of software implementation of algorithms that use the gradient as the direction of descent. The subject of the research is the issues related to the comparative analysis of the most common and well-founded technologies of selection and calculation of directions and steps that use algorithms that correspond to the above scheme.

Research methods and tools. The main research method that determines the research technology is the method of computational experiments applied to nonlinear unconditional optimization problems. Important tools for the implementation of research methods are also .NET programming technologies, software models for building visual graphical representations and data aggregation. Research methods are implemented with the help of programming patterns based on .NET technologies.

Scientific novelty and practical significance of the obtained results. Methods of unconditional optimization is one of the important directions in engineering practice. The basic numerical method of unconditional optimization are methods that use the gradient of the function as the direction of descent.

The software product described in this paper allows to obtain and substantiate recommendations for the choice of a set of parameters of nonlinear optimization methods. The recommendations are based on experiments with the most complex functions for descent methods. Programs based on similar principles and having the above characteristics are not known.

Keywords: filament, system of serve of filament, directing elements, strainers of filament.

Вступ

Мета і завдання. Найбільший інтерес до градієнтного методу в останніх роках пов'язаний з тим, що градієнтні спуски та їх стохастичні або рандомізовані варіанти лежать в основі майже всіх сучасних алгоритмів навчання, розроблених в аналізі даних [1]. Більшість алгоритмів оптимізації походить із пропозицій, що має доступ до точного градієнту або гесіану. На практиці зазвичай є лише зашумлена або навіть зміщена оцінка цих величин. Майже всі алгоритми глибокого навчання описуються на вибіркових оцінках, за крайніми мережами, у тому, що торкається використання міні-пакетів навчальних прикладів для обчислення градієнта. Буває і так, що цільова функція, яка мінімізує, є такою що не має обчислювального розв'язання. У такому випадку не має обчислювального розв'язання звичайно і завдання підрахунку градієнта, а тоді залишається лише апроксимований градієнт. Такі проблеми частіше всього виникають в складних моделях, наприклад, алгоритм співставленого розподілу (контрактивна дивергенція) пропонує метод апроксимації градієнтних функцій логарифмічного правдоподібного машинобудування. Для компенсації неточної оцінки градієнта розроблені різні алгоритми оптимізації нейронних мереж. Проблема можна знайти також за допомогою вибору сурогатної функції втрат, яка ще більше апроксимує, чим істинна. В будь якому випадку градієнтні спуски та їх стохастичні варіанти лежать в основі майже всіх сучасних алгоритмів навчання. Мета дослідження розробити програмний засіб, що дозволяє дослідити і порівняти експериментально ефективність різних варіантів програмної реалізації алгоритмів, що використовують градієнт в якості напрямку спуску. Варіанти пов'язані з різноманітними методами побудови напрямків спуску з використанням напрямку, що визначає градієнт.

Об'єкт та предмет дослідження. Об'єктом дослідження є особливості програмної реалізації алгоритмів, що використовують градієнт в якості напрямку спуску. Предметом дослідження є питання пов'язані з порівняльним аналізом найбільш поширених та обґрунтованих технологій вибору та обчислення напрямків та кроків які використовують алгоритми, що відповідають наведеній схемі.

Методи та засоби дослідження. Основним методом дослідження, що визначає технологію дослідження, є метод обчислювальних експериментів, що застосовуються до задач нелінійної безумовної оптимізації. Важливі засоби для реалізації методів дослідження це також технології програмування .NET, програмні моделі побудови наочних графічних представлень та агрегація даних. Реалізуються методи дослідження з допомогою патернів програмування на базі технологій .NET.

Наукова новизна та практичне значення отриманих результатів. Методи безумовної оптимізації один із важливих напрямків в інженерній практиці. Базовим чисельним методом безумовної оптимізації є методи, що використовують градієнт функції як напрямок спуску.

Описаний в даній роботі програмний продукт дозволяє отримати та обґрунтувати рекомендації для вибору сукупності параметрів методів нелінійної оптимізації. Рекомендації базуються на експериментах з найскладнішими для методів спуску функціях. Програми, що базуються на подібних принципах і мають наведені характеристики не відомі.

Основна частина

Предметом дослідження є питання пов'язані з порівняльним аналізом найбільш поширених та обґрунтованих технологій вибору напрямків та кроків які використовують алгоритми, що відповідають схемі

$$x_{k+1} = x_k - h_k (v'(x_k) + v_2'(x_k)), k=0, 1, \dots$$

і напрямки $v'(x_k)$, та $v_2'(x_k)$ визначається з допомогою градієнту.

Ця проблема є найбільш важливою (визначає коло задач які можуть бути досліджені) в нелінійному програмуванні і, в наш час, не має теоретичного розв'язку. Теоретичну якість алгоритмів спуску для задач нелінійного програмування характеризують параметром що визначає наступна формальна залежність

$$\|x^{j+1} - x^*\| \leq g_j \|x^j - x^*\| \quad \text{або} \quad \|x^{j+1} - x^*\| \leq g \|x_j - x^*\|^2$$

Наведене співвідношення визначає класи швидкості збіжності алгоритмів. Теоретична оцінка (якість) алгоритмів не завжди збігається з практичними результатами і це підтверджено програмою, що була запропонована.

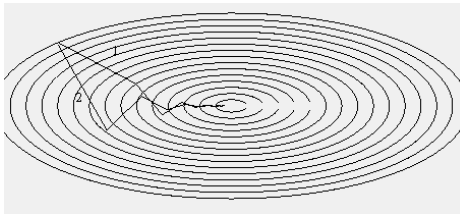
В [3] відзначається що для завдань безсумнівної мінімізації найпопулярніший зараз є наступний (двокроковий) варіант швидкого

градієнтного методу. Обчислення виконуються відповідно наведеним нижче формулам

$$x^{k+1} = y^k - \frac{1}{L} \nabla f(y^k),$$

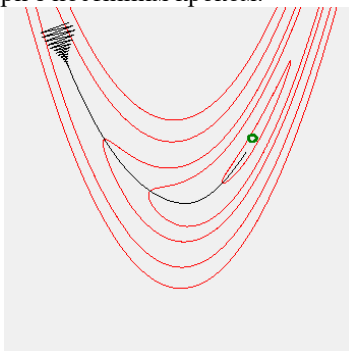
$$y^k = x^k + \frac{k-1}{k+2} (x^k - x^{k-1}),$$

Обчислювальні експерименти, що були проведені, не підтвердили це твердження.



На малюнку траєкторія 2 відповідає (двокроковому) варіанту швидкого градієнтного методу, а траєкторія 1 визначає траєкторію важкого шарика.

Для функції Розенброка ефективність (двокроковою) варіанту швидкого градієнтного методу залежить від початкової точки і в більшості випадків гірше ніж з дробленням кроку та регулюванням кроку, не значно відрізняється від траєкторії з постійним кроком.



Висновки

На основі аналізу даних експериментальних досліджень, можна стверджувати, що найбільш надійними засобами отримання результатів є алгоритм з регулюванням кроку або алгоритм з дробленням кроку якщо

використовується нормований градієнт. В будь якому випадку крок задовольняє правило Голдштейна - Армійо: знайти таке h що

$$x_{k+1} = x_k - hf'(x_k), \text{ та мають місце нерівності}$$

$$\alpha f'(x_k), x_k - x_{k+1} \leq f(x_k) - f(x_{k+1}),$$

$$\beta f'(x_k), x_k - x_{k+1} \geq f(x_k) - f(x_{k+1}),$$

де $\alpha, \beta, 0 < \alpha < \beta < 1$, — деякі фіксовані параметри.

Література

1. Гудфеллоу Я., БенДжио И., Курвиль А. Глибоке навчання. – Івано-Франківськ, 2017. - 652 с.
2. Shi Y., Eberhart R.C. (1998). Parameter selection in particle swarm optimization. Proceedings of Evolutionary Programming VII (EP98). с. 591–600.
3. Eberhart R.C., Shi Y. (2000). Comparing inertia weights and constriction factors in particle swarm optimization. Proceedings of the Congress on Evolutionary Computation 1. с. 84–88