

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДИФИКАЦИЙ МЕТОДА СТОХАСТИЧЕСКОГО ГРАДИЕНТА ПРИ РЕШЕНИИ ЗАДАЧИ РЕГРЕССИИ

**Воробьев Иван Григорьевич**, студент, направление подготовки 01.03.02 Прикладная математика и информатика, Оренбургский государственный университет, Оренбург  
e-mail: ivanexplay2000@gmail.com

**Гришина Любовь Сергеевна**, аспирант, направление подготовки 02.06.01 Компьютерные и информационные науки, Оренбургский государственный университет, Оренбург  
e-mail: zabrodina97@inbox.ru

Научный руководитель: **Болодурина Ирина Павловна**, доктор технических наук, профессор, заведующий кафедрой прикладной математики, Оренбургский государственный университет, Оренбург  
e-mail: prmat@mail.osu.ru

***Аннотация.** В настоящее время регрессионный анализ данных является одним из наиболее мощных методов выявления характера зависимостей для построения математических моделей. Одним из наиболее актуальных и эффективных методов её решения являются оптимизационные методы машинного обучения. Точность получаемого решения во многом зависит как от выбранного метода, так и от его параметров. Данное исследование направлено на реализацию некоторых стохастических методов оптимизации и рассмотрения различных темпов обучения для них с целью анализа их влияния на точность решения регрессионной задачи. Рассмотревались следующие методы: SGD, SAG, Momentum, NAG. Результаты вычислительных экспериментов на синтетических данных показали, что наиболее точное решение регрессионной задачи обеспечивает метод NAG. Разработанное программное обеспечение может быть использовано для дальнейших исследований, направленных на поиск оптимизационных методов, дающих наиболее точное решение регрессионных задач.*

***Ключевые слова:** регрессия, оптимизация, стохастический градиентный спуск, темп обучения.*

***Для цитирования:** Воробьев И. Г., Гришина Л. С. Сравнительный анализ модификаций метода стохастического градиента при решении задачи регрессии // Шаг в науку. – 2021. – № 4. – С. 46–51.*

## COMPARATIVE ANALYSIS OF MODIFICATIONS OF THE STOCHASTIC GRADIENT METHOD FOR SOLVING THE REGRESSION PROBLEM

**Vorobiev Ivan Grigorievich**, student, training program 01.03.02 Applied Mathematics and Computer Science, Orenburg State University, Orenburg  
e-mail: ivanexplay2000@gmail.com

**Grishina Lyubov Sergeevna**, postgraduate student, training program 02.06.01 Computer and Information Sciences, Orenburg State University, Orenburg  
e-mail: zabrodina97@inbox.ru

Research advisor: **Bolodurina Irina Pavlovna**, Doctor of Technical Sciences, Professor, Head of the Department of Applied Mathematics, Orenburg State University, Orenburg  
e-mail: prmat@mail.osu.ru

***Abstract.** Currently, regression data analysis is one of the most powerful methods for identifying the nature of dependencies for building mathematical models. One of the most relevant and effective methods for solving it is machine learning optimization methods. The accuracy of the obtained solution largely depends on both the chosen method and its parameters. This study is aimed at implementing some stochastic optimization methods and considering different learning rates for them in order to analyze their impact on the accuracy of solving the regression problem. The following methods were considered: SGD, SAG, Momentum, NAG. The results of computational experiments on synthetic data have shown that the NAG method provides the most accurate solution to the regression problem. The developed software can be used for further research aimed at finding optimization*

methods that provide the most accurate solution to regression problems.

**Key words:** regression, optimization, stochastic gradient descent, learning rate.

**Cite as:** Vorobiev, I. G., Grishina, L. S. (2021) [Comparative analysis of modifications of the stochastic gradient method for solving the regression problem]. *Shag v nauku* [Step into science]. Vol. 4, pp. 46–51.

### Введение

Область решения задач регрессионного анализа данных связана с выявлением характера зависимостей некоторых свойств изучаемого процесса между собой. Регрессионный анализ используется для решения проблем в экономике, физике и химии [4], эпидемиологии [1] и др.

Благодаря развитию вычислительных технологий задачи регрессионного анализа стали решать, используя методы машинного обучения. Применение методов интеллектуального анализа данных позволяет строить математические модели с настроенными параметрами для исследуемых процессов. Используя различные функционалы качества, осуществляется оценка полученной модели и её релевантность на эмпирических данных.

Таким образом, задача регрессионного анализа сводима к задаче оптимизации функционала качества от некоторой параметрической модели исследуемого процесса. Данный функционал вычисляется на некоторых, заранее подготовленных эмпирических данных. Теория решения оптимизационных задач начала интенсивно развиваться, начиная с середины прошлого века, однако классические алгоритмы неэффективны для решения регрессионных задач.

В силу большого числа рассматриваемых признаков, по которым производится регрессия, вычисление полного градиента функционала качества, необходимого для методов оптимизации первого порядка, становится трудоёмкой вычислительной задачей. Для преодоления этой проблемы были разработаны метод стохастического градиента и его модификации. Однако до возникновения проблем при обработке больших массивов данных (Big Data) данные алгоритмы были не востребованы.

Модификации метода стохастического градиента на данный момент получили широкое распространение в решении задач регрессионного анализа при обработке больших данных с использованием алгоритмов машинного обучения.

Таким образом, исследование различных модификаций алгоритма стохастического градиента является на данный момент актуальной задачей, так как выбор наиболее эффективного подхода может обеспечить как значительно более высокую точность, так и сниженное время выполнения.

### 1. Обзор исследований

Исследованиями стохастических методов оптимизации в различных прикладных задачах, а также анализом их свойств занимаются учёные по всему миру.

Авторы исследования [5] рассматривают обучение нейронных сетей, как оптимизационную задачу. В данной работе выделены специфические черты, исходя из которых определены допустимые методы решения. Данные методы представляют собой различные модификации стохастического градиентного спуска. Однако данное исследование не содержит примеров практической реализации алгоритмов на реальных или синтетических данных.

В исследовании [10] авторы анализируют различные, в основном, адаптивные модификации стохастического градиентного спуска. Производится сравнение точности их решения обыкновенной задачи двумерной оптимизации при фиксированном темпе обучения и количестве итераций. По результатам анализа авторы выделяют метод Adadelta, как наиболее оптимальный среди рассматриваемых.

В публикации [7] рассматриваются пакетные и стохастические модификации классического градиентного спуска. Анализируются решения регрессионной задачи вычисления  $Z$ -фактора углекислого газа различными модификациями через сравнение среднего квадратичного отклонения каждого из них от реального значения  $Z$ -фактора. Наиболее точным из рассматриваемых авторами методов для решения данной задачи является классический градиентный спуск.

Автор исследования [6] проанализировал скорость сходимости классического градиентного спуска и некоторых модификаций метода стохастического градиента на задачах обыкновенной двумерной оптимизации, логической регрессии и задачи распознавания образов. Сравнение результатов решения рассматриваемыми модификациями задачи обыкновенной оптимизации не позволяет явно определить наиболее эффективный метод. Однако на всех остальных задачах метод S2GD продемонстрировал наилучшие результаты.

Таким образом, обзор исследований показал, что сравнительный анализ различных модификаций алгоритма стохастического градиентного спуска с целью выявления наиболее точных и быстро сходящихся из них для решения регрессионных задач – один из актуальных вопросов на настоящий момент.

Данное исследование направлено на сравнение четырёх модификаций метода стохастического градиента: обыкновенный стохастический градиентный спуск (SG), средний стохастический градиент (SAG), стохастический градиент с импульсом (Momentum), стохастический градиент Нестерова (NAG). Сравнительный анализ проводится на осно-

ве их времени выполнения и значения функционала качества на полученной в результате оптимизации модели. Также в рамках данной статьи рассматривается изменение поведения представленных методов при изменении темпа обучения.

## 2. Постановка задачи

Рассмотрим некоторое конечное подмножество  $X_n = \{x_1, x_2, \dots, x_n\}$  множества исследуемых объектов произвольной природы  $X$  такое, что ему соответствует множество  $Y_n = \{y_1, y_2, \dots, y_n\} = \{y(x_1), \dots, y(x_n)\} \subseteq \mathbb{R}^n$  – множество заранее известных значений некоторого искомого вещественнозначного признака  $y(x)$  на  $X_n$ .

Для решения регрессионной задачи необходимо найти некоторый вектор параметров  $w = (w_1, \dots, w_m) \in \mathbb{R}^m$  такой, что для заранее выбранного параметрического алгоритма  $a(x, w)$  будет верно следующее соотношение:

$$a(x, w) \approx y(x) \leftrightarrow Q(a, y) < \varepsilon,$$

где

$Q(a, y)$  – функционал качества, демонстрирующий величину ошибки приближения, а вещественное  $\varepsilon > 0$  – требуемая точность.

Тогда для решения соответствующей оптимизационной задачи требуется найти вектор параметров  $w$  такой, что  $w = \arg \min_v \sum_{i=1}^n Q(a(x_i, v), y_i)$ .

В данном исследовании будут использоваться следующие функции  $a(x, w)$  и  $Q(a, y)$ :

$$a(x, w) = \sum_{i=1}^m w_i f_i(x),$$

где

$f_i(x), i = 1, \dots, m$  – некоторые численные признаки объекта  $x$ , в рамках данного исследования  $f_i(x) = x^i, i = 1, \dots, m$ ;  
 $Q(a, y) = (a - y)^2$  – квадратичный функционал качества.

## 3. Численное решение задачи оптимизации

### 3.1. Метод стохастического градиента

Стохастический градиентный спуск [8] – итеративный алгоритм поиска локального минимума функции. На итерации с номером  $h$  вектор параметров  $w$  обновляется в направлении функции  $-\nabla_w Q_i(w)$ , где  $Q_i(w) = Q(a(x_i, w), y_i)$  и  $i$  – случайно выбранное целое число такое, что  $1 \leq i \leq n$ . Принцип работы данного метода состоит в том, что математическое ожидание  $-\nabla_w Q_i(w)$  равняется антиградиенту функционала качества на  $X_n$ :  $\frac{1}{n} \nabla_w \sum_{i=1}^n Q_i(w)$ , который показывает локальное направление наибольшего убывания функции.

Его алгоритм имеет следующий вид:

*Вход:* выборки  $X_n, Y_n$ , темп обучения  $\alpha$ , темп забывания  $\lambda$ ;

*Выход:* вектор параметров  $w$ ;

1. Инициализируем  $w$  и  $\bar{Q} := \frac{1}{n} \sum_{i=1}^n Q_i(w)$
2. Повторять
3. | Случайно выберем  $i$  от 1 до  $n$

4. |  $w := w - \alpha \nabla_w Q_i(w)$
5. |  $\bar{Q} := \lambda Q_i(w) + (1 - \lambda) \bar{Q}$
6. Пока  $w$  и/или  $\bar{Q}$  не сойдутся

### 3.2. Метод стохастического среднего градиента

Данная модификация работает по тому же принципу, что и SGD, однако для ускорения сходимости итерационный шаг осуществляется в направлении среднего арифметического от антиградиентов  $Q_i(w), i = 1, \dots, n$ .

Алгоритм SGD имеет следующий вид [9]:

*Вход:* выборки  $X_n, Y_n$ , темп обучения  $\alpha$ , темп забывания  $\lambda$ ;

*Выход:* вектор параметров  $w$ ;

1. Инициализируем  $w$  и  $\bar{Q} := \frac{1}{n} \sum_{i=1}^n Q_i(w)$
2. Инициализируем градиенты  $G_i := \nabla_w Q_i(w)$
3. Повторять
4. | Случайно выберем  $i$  от 1 до  $n$
5. |  $G_i := \nabla_w Q_i(w)$
6. |  $w := w - \alpha \nabla_w Q_i(w)$
7. |  $\bar{Q} := \lambda Q_i(w) + (1 - \lambda) \bar{Q}$
8. Пока  $w$  и/или  $\bar{Q}$  не сойдутся

### 3.3. Метод стохастического градиента с импульсом

Данный метод представляет собой модификацию SAG, где прямое вычисление среднего антиградиентов заменено вычислением скользящего среднего от них (импульса), что позволяет ускорить вычисления.

Алгоритм Momentum имеет следующий вид [3]:

*Вход:* выборки  $X_n, Y_n$ , темп обучения  $\alpha$ , темп забывания  $\lambda$ , силу импульса  $\gamma$ ;

*Выход:* вектор параметров  $w$ ;

1. Инициализируем  $w$  и  $\bar{Q} := \frac{1}{n} \sum_{i=1}^n Q_i(w)$
2. Инициализируем импульс  $v$
3. Повторять
4. | Случайно выберем  $i$  от 1 до  $n$
5. |  $v := \gamma v + (1 - \gamma) \nabla_w Q_i(w)$
6. |  $w := w - \alpha v$
7. |  $\bar{Q} := \lambda Q_i(w) + (1 - \lambda) \bar{Q}$
8. Пока  $w$  и/или  $\bar{Q}$  не сойдутся

### 3.4. Метод стохастического градиента Нестерова

Модификация Momentum, где сначала осуществляется шаг в направлении импульса, а лишь затем вычисляется градиент.

Алгоритм NAG имеет следующий вид [2]:

*Вход:* выборки  $X_n, Y_n$ , темп обучения  $\alpha$ , темп забывания  $\lambda$ , силу импульса  $\gamma$ ;

*Выход:* вектор параметров  $w$ ;

1. Инициализируем  $w$  и  $\bar{Q} := \frac{1}{n} \sum_{i=1}^n Q_i(w)$
2. Инициализируем импульс  $v$
3. Повторять
4. | Случайно выберем  $i$  от 1 до  $n$

5.  $v := \gamma v + (1 - \gamma) \nabla_w \mathcal{L}_i(w - \alpha \gamma v)$  и предельно допустимым числом итераций.
6.  $w := w - \alpha v$
7.  $\bar{Q} := \lambda Q_i(w) + (1 - \lambda) \bar{Q}$
8. Пока  $w$  и/или  $\bar{Q}$  не сойдутся

### 3.5. Достоинства и недостатки рассматриваемых методов

Отметим, что завершение выполнения метода может обеспечиваться не только сходимостью, но и на основании исследованных работ можно выделить следующие характерные черты данных методов.

Таблица 1. Характеристика исследуемых модификаций

|             | SGD   | SAG   | Momentum  | NAG |
|-------------|---|---|---|-----|
| Достоинства | Уменьшение вычислительной нагрузки по сравнению с градиентным спуском | Уменьшение вычислительной нагрузки при сохранении скорости сходимости | Уменьшение вариации шага относительно градиентного спуска |     |
| Недостатки  | Низкая скорость сходимости  | Высокое потребление памяти  | Необходимость подбора дополнительного параметра           |     |

Источник: разработано авторами на основе [2, 3, 8, 9]

### 4. Результаты исследования

В ходе выполнения работы были реализованы исследуемые методы на языке программирования Python и проведена серия вычислительных экспериментов.

Анализируются решения регрессионных задач при следующих значениях параметров:

- $n = 1000$ ;
- $X_n = \{a, a + \frac{b-a}{n}, \dots, b\}$ ,

где

$$a, b \in \mathbb{R};$$

$$y(x) = f(x) + \text{uniform\_noise},$$

где

*uniform\_noise* – равномерный шум;

- $\alpha(x, w) = \sum_{i=0}^m w_i x^i$ ;
- $Q(a, y) = (a - y)^2$ ;
- $\varepsilon = 10^{-6}$ ;
- $\lambda = 0.35$ ;
- $\gamma = 0.35$ ;
- $\alpha \in \{\frac{1}{h}, \frac{1}{\sqrt{h}}\}$ ,

где

$h$  – номер итерации;

$h_{max} = 10^6$  – предельно допустимое число итераций,

где

$f(x)$ ,  $a$ ,  $b$ , *uniform\_noise* и  $m$  принимают одно из следующих значений:

$$1) a = 0, b = 1, f(x) = x^2, \text{uniform\_noise} \in [-\frac{1}{2}, \frac{1}{2}], m = 2;$$

$$2) a = -6, b = 6, f(x) = x^3, \text{uniform\_noise} \in [-15, 15], m = 3;$$

$$3) a = 0, b = 10, f(x) = \sqrt{x}, \text{uniform\_noise} \in [-\frac{1}{2}, \frac{1}{2}], m = 2;$$

$$4) a = 0, b = 6, f(x) = \sin(x), \text{uniform\_noise} \in [-\frac{1}{2}, \frac{1}{2}], m = 3.$$

На основе полученных решений были составлены две таблицы.

Таблица 2. Результаты решения регрессионных задач при  $\alpha = 1 / h$

| $\alpha = 1 / h$ | $f(x) = x^2$  | $f(x) = x^3$  | $f(x) = \sqrt{x}$ | $f(x) = \sin(x)$ |
|------------------|---------------|---------------|-------------------|------------------|
| SGD              | 0.026; 423483 | 0.020; 297274 | 0.002; 162749     | 0.078; $h_{max}$ |
| SAG              | 0.065; 6773   | 0.031; 39617  | 0.01; 10935       | 0.138; 32424     |
| Momentum         | 0.011; 234544 | 0.017; 153008 | 0.002; 92590      | 0.066; $h_{max}$ |
| NAG              | 0.011; 190518 | 0.012; 148019 | 0.004; 87734      | 0.053; $h_{max}$ |

Источник: разработано авторами

Таблица 3. Результаты решения регрессионных задач при  $\alpha = 1/\sqrt{h}$ 

| $\alpha = 1/\sqrt{h}$ | $f(x) = x^2$      | $f(x) = x^3$      | $f(x) = \sqrt{x}$ | $f(x) = \sin(x)$  |
|-----------------------|-------------------|-------------------|-------------------|-------------------|
| SGD                   | 0.045; $h_{\max}$ | 0.017; $h_{\max}$ | 0.008; $h_{\max}$ | 0.087; $h_{\max}$ |
| SAG                   | 0.044; 12966      | 0.020; 12652      | 0.010; 7653       | 0.091; 106911     |
| Momentum              | 0.016; $h_{\max}$ | 0.013; $h_{\max}$ | 0.006; $h_{\max}$ | 0.076; $h_{\max}$ |
| NAG                   | 0.015; $h_{\max}$ | 0.006; $h_{\max}$ | 0.003; $h_{\max}$ | 0.052; $h_{\max}$ |

Источник: разработано авторами

В таблицах 2 и 3 представлены для каждого метода значения функционала качества на полученном векторе параметров  $w$  и число использованных итераций.

В результате эксперимента метод NAG показал наилучшие результаты по точности почти во всех рассмотренных задачах. Несколько менее точным оказался Momentum при аналогичном быстродействии. Наихудшей модификацией по точности ответа является SAG во всех рассматриваемых задачах, однако он показывает наилучшее быстродействие. Обыкновенный стохастический градиент имеет более низкую точность и скорость сходимости по отношению к импульсным модификациям, при этом не давая преимуществ во времени выполнения.

Сравнение решений при различных  $\alpha$  показало, что  $\alpha = 1/\sqrt{h}$  показывает как более низкие темпы сходимости, так и менее точные решения.

### Заключение

В рамках данного исследования была рассмотрена регрессионная задача в качестве оптимизационной, для решения которой исследовались численные методы. Был осуществлен выбор моди-

фикаций метода стохастического градиента среди прочих классов численных алгоритмов и их программная реализация.

Полученные методы решения продемонстрировали хорошие результаты при решении регрессионных задач на синтетических данных, наибольшее из полученных значений функционала качества 0.138.

Проведенный при решении регрессионных задач сравнительный анализ некоторых модификаций стохастического градиентного спуска показал, что метод NAG позволяет получать наиболее точное решение, однако, при значимых временных затратах. Наиболее быстродейственный метод из рассмотренных – SAG, который при этом демонстрирует наименьшую точность. Однако если SAG не успевае найти решение после малого числа итераций, то скорость его выполнения заметно снижается из-за необходимости хранения градиентов для всех элементов выборки и вычисления их среднего. Также по итогам вычислительных экспериментов установлено, что при  $\alpha = 1/h$  и точность, и скорость сходимости возрастает по отношению к ситуации, в которой темп обучения  $\alpha = 1/\sqrt{h}$ .

### Литература

1. Евстегнеева В. А., Честнова Т. В., Смольянинова О. Л. Анализ временных рядов в прогнозировании природно-очаговых инфекций // Вестник новых медицинских технологий. Электронное издание. – 2015. – №. 4. – С. 9.
2. Нестеров Ю. Е. Метод решения задачи выпуклого программирования со скоростью сходимости  $O(1/k^2)$  // Доклады Академии наук. – Российская академия наук, 1983. – Т. 269. – №. 3. – С. 543–547.
3. Поляк Б. Т. О некоторых способах ускорения сходимости итерационных методов // Журнал вычислительной математики и математической физики. – 1964. – Т. 4. – №. 5. – С. 791–803.
4. Померанцев А. Л. Методы нелинейного регрессионного анализа для моделирования кинетики химических и физических процессов: дис. ... д-р физ.-мат. наук: 01.04.17. – Москва. – 2003. – 304 с.
5. Самолук Т. А. Исследование стохастических градиентных методов оптимизации алгоритмов обучения искусственных нейронных сетей // Комп'ютерні засоби, мережі та системи. – 2017. – №. 16. – С. 97–105.
6. Papamakarios G. Comparison of modern stochastic optimization algorithms // CiteSeerX: Библиогр. база данных. – URL: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.716.6367&rep=rep1&type=pdf> (дата обращения: 29.04.2021).
7. Riza L. S. et al. gradDescentR: An R package implementing gradient descent and its variants for regression tasks // 2016 1st International Conference on Information Technology, Information Systems and Electrical



Engineering (ICITISEE). – IEEE, 2016. – pp. 125–129.

8. Robbins H., Monro S. A stochastic approximation method //The annals of mathematical statistics. – 1951. – Vol. 22 . – No. 3 – pp. 400–407.

9. Schmidt M., Le Roux N., Bach F. Minimizing finite sums with the stochastic average gradient // Mathematical Programming. – 2017. – Vol. 162. – №. 1–2. – pp. 83–112.

10. Yazan E., Talu M. F. Comparison of the stochastic gradient descent based optimization techniques // 2017 International Artificial Intelligence and Data Processing Symposium (IDAP). – IEEE, 2017. – pp. 1–5.

Статья поступила в редакцию: 19.05.2021; принята в печать: 08.11.2021.

Авторы прочитали и одобрили окончательный вариант рукописи.