DOI: 10.18137/RNU.V9187.25.02.P.43 УДК 004.032.26

Сонина Светлана Дмитриевна
старший преподаватель кафедры автоматики, телемеханики, связи и вычислительной техники, Донецкий институт железнодорожного транспорта, город Донецк. ORCID: 0009-0009-6114-4386, AuthorID: 1037741, SPIN-код: 9626-1100,
Электронный адрес: soninadonigt@yandex.com
Svetlana D. Sonina
Senior Lecturer at the Department of automation, telemechanics, communications and computer technology, Donetsk Railway Transport Institute, Donetsk. ORCID: 0009-0009-6114-4386, AuthorID: 1037741, SPIN-code: 9626-1100,
E-mail address: soninadonigt@yandex.com

ОПТИМИЗАЦИЯ ВРЕМЕНИ ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ С АДАПТИВНЫМИ ПАРАМЕТРАМИ СКОРОСТИ ОБУЧЕНИЯ

Аннотация. В работе выполнен анализ классического метода градиентного спуска и предложен способ динамического изменения шага обучения на основе вычисляемых параметров τ и p. Основной акцент сделан на алгоритме, который позволяет вычислять оптимальные значения параметров τ и p для минимизации времени обучения. Эксперименты демонстрируют, как изменения этих параметров влияют на скорость обучения для различных топологий нейронных сетей и функций активации. Результаты моделирования показывают, что правильный выбор τ и p может значительно сократить временные затраты при обучении нейронных сетей с фиксированной структурой. Использование этих параметров позволяет улучшить процесс обучения, предотвращая застревание в локальных минимумах и обеспечивая баланс между скоростью обучения и точностью результата. Исследования продемонстрировали эффективность адаптивного подхода при различных топологиях нейронных сетей и функциях активации. Представленные графики и численные расчёты показывают зависимость средней скорости обучения от выбранных параметров.

Ключевые слова: нейронная сеть, градиентный спуск, оптимизация, скорость обучения, адаптивные параметры, корректировка

Для цитирования: *Сонина С.Д.* Оптимизация времени обучения нейронных сетей с адаптивными параметрами скорости обучения // Вестник Российского нового университета. Серия: Сложные системы: модели, анализ, управление. 2025. № 2. С. 43 – 54. DOI: 10.18137/RNU.V9187.25.02.P.43

OPTIMIZING THE LEARNING TIME OF NEURAL NETWORKS WITH ADAPTIVE LEARNING RATE PARAMETERS

Abstract. The paper analyzes the classical gradient descent method and suggests a method for dynamically changing the learning step based on the calculated parameters τ and p. The main focus is on an algorithm that allows calculating the optimal values of the parameters τ and p to minimize the training time. The experiments demonstrate how changes in these parameters affect the learning rate for various neural network topologies and activation functions. The simulation results show that the correct choice of τ and p can significantly reduce the time required for training neural networks with a fixed structure. Using these parameters allows to improve the learning process, preventing getting stuck in local minima and ensuring a balance between the learning rate and the accuracy of the result. Research has demonstrated the effectiveness of an

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление», выпуск 2 за 2025 год

adaptive approach for various neural network topologies and activation functions. The presented graphs and numerical calculations show the dependence of the average learning rate on the selected parameters.

Keywords: neural network, gradient descent, optimization, learning rate, adaptive parameters, correction.

For citation: Sonina S.D. (2025) Optimization of training time of neural networks with adaptive learning rate parameters. *Vestnik of Russian New University. Series: Complex Systems: Models, analysis, management.* No. 2. Pp. 43 – 54. DOI: 10.18137/RNU.V9187.25.02.P.43 (In Russian).

В современном мире нейронные сети (далее – НС) используются при решении сложных задач в области распознавания изображения, обработки естественного языка, финансового прогнозирования и медицинской диагностики. Стремительный рост объемов данных приводит к необходимости быстрого и точного обучения модели. Одной из основных задач при этом является оптимизации скорости обучения НС. Традиционный алгоритм обратного распределения ошибки выполняет корректировку весовых коэффициентов, тем самым позволяя минимизировать функцию потерь [1]. Хотя данный подход и эффективен, но имеет значительные ограничения при обучении сети с большим количеством слоев, поскольку требует значительных вычислительных ресурсов и времени. По этой же причине ограничивается применение данного алгоритма моделей, требующих частых обновлений или переобучения с использованием новых данных.

В работах [2–4] рассматривается возможность решения задачи скорости обучения с применением различных модификаций метода градиентного спуска, а также их эффективность при сохранении точности обучения модели. При неправильном выборе скорости обучения возможно замедление сходимости для достижения приемлемого уровня точности к колебаниям вокруг локального минимума и, как следствие, невозможность нахождения оптимального решения. При решении задач минимизации функции потерь методы Adam, RMSprop и Adagrad [5; 6] находят эффективное применение за счет использования адаптивных скоростей обучения. Несмотря на то, что данные методы успешно применяются при решении задач оптимизации, необходимо учитывать их ограничения, такие как повышенная чувствительность к начальным значениям или вероятность преждевременной сходимости.

Скорость обучения η – значение шага изменения весовых коэффициентов, которое влияет на приближение (удаление) к минимуму функции. В работе [7] описан классический метод градиентного спуска с фиксированным значением скорости обучения $\eta = 0{,}1$. Определено, что константное значение не всегда обеспечивает наилучший результат и тем самым ограничивает применение метода в реальной практике. Однако в результате проведенных исследований выявлено, что существует возможность уменьшения времени обучения, если параметр η на итерации n изменять в соответствии с выражением [2]

$$\eta(n) = \begin{cases} \eta(n-1)\tau, \ \textit{если} \ \sum\limits_{i=1}^{w}\left(\left|w_{ji}^{k}(n)\right| - \left|w_{ji}^{k}(n-1)\right|\right) < p, \\ \dfrac{\eta(n-1)}{\tau}, \ \textit{если} \ \sum\limits_{i=1}^{w}\left(\left|w_{ji}^{k}(n)\right| - \left|w_{ji}^{k}(n-1)\right|\right) \geq p, \end{cases} \tag{1}$$

где $\eta(n)$ – значение η на итерации n; $\eta(n-1)$ – значение на предыдущей итерации; τ – коэффициент скорости изменения $\eta(n)$, $\tau \in [0,1]$, устанавливается перед началом обучения; $w_{ji}^{k}(n)$ – значение i -го весового коэффициента из общего их числа W в j -м нейроне k -го слоя; p – пороговое значение, в соответствии с которым производится корректировка $\eta(n)$.

В выражении (1) $\sum_{i=1}^{w}\left(\left|w_{ji}^{k}(n)\right|-\left|w_{ji}^{k}(n-1)\right|\right)$ представляет собой сумму разностей абсолютных значений всех весовых коэффициентов j -го нейрона k -го слоя на текущей и предыдущей итерации. Как показано в работе [8], при приближении значения этого выражения к нулю, итерационный алгоритм градиентного спуска приближается к минимуму. Однако если при этом суммарная энергия ошибки не удовлетворяет условию, то он является локальным. В связи с этим в выражении (1) введены параметры p и τ, устанавливаемые перед началом процесса обучения. Их смысл заключается в обеспечении необходимого шага изменения весовых коэффициентов в процессе градиентного спуска при приближении алгоритма к минимуму.

Таким образом, имеются две зависимые переменные τ и p, определить значение которых аналитически не представляется возможным. Для их оценки численным методом разработан алгоритм и программное обеспечение, которое производит расчет времени выполнения процесса обучения НС при изменении параметров $\hat{o}$ и p в некотором диапазоне, устанавливаемом априорно.

Алгоритм (Рисунок 1) реализован в программном обеспечении и проведен ряд экспериментов. При этом установлены следующие значения переменных: локальный цикл (используемый при обучении одного примера) – $local\_delta\_min = 0,1$ и $local\_count = 1000$; глобальный $total\_delta\_min = 0,01$ и $counter = 100$; $\acute{a} = 0,1$; параметры функции нормального распределения при рандомизации весовых коэффициентов и значений параметра наклона функции возбуждения нейронов – среднее $m = 0$, стандартное отклонение $s = 0,01$. Расчеты проводились для 100 дискретных значений $\tau \in [0 \ldots 1]$ и 100 значений $p \in \left[10^{-10} \ldots 1\right]$.

Так, для сигмоидальной функции возбуждения нейронов в сети минимальной конфигурации (три нейрона, топология 2-1) отмечается следующая тенденция: поверхность значений возрастает от точки $\tau = 0$ и $p = 10^{-10}$ до точек диапазона $\tau = [0,9 \ldots 1,0]$ и $p = \left[10^{-9} \ldots 10^{-1}\right]$ и характеризуется значительной неравномерностью (Рисунок 2). Минимальное значение $s = 0,47$ (мс) получено при $\tau = 0,61$ и $p = 1$. Следует отметить, что возрастание значений s не зависит от топологии НС с сигмоидальной функцией возбуждения нейронов.

В работах [9; 10] рассматривалось влияние архитектуры на эффективность обучения НС. Данные исследования показали, что увеличение количества скрытых слоев может как ускорить сходимость, так и привести к переобучению при неоптимальном подборе параметров.

В качестве примера исследованы трехслойные НС вида 2-5-1 (Рисунок 3, *а*) и 5-10-2 (Рисунок 3, *б*).

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление», выпуск 2 за 2025 год

Рисунок 1. Алгоритм расчета времени процесса обучения НС в зависимости от параметров τ и p
Источник: здесь и далее рисунки выполнены автором.

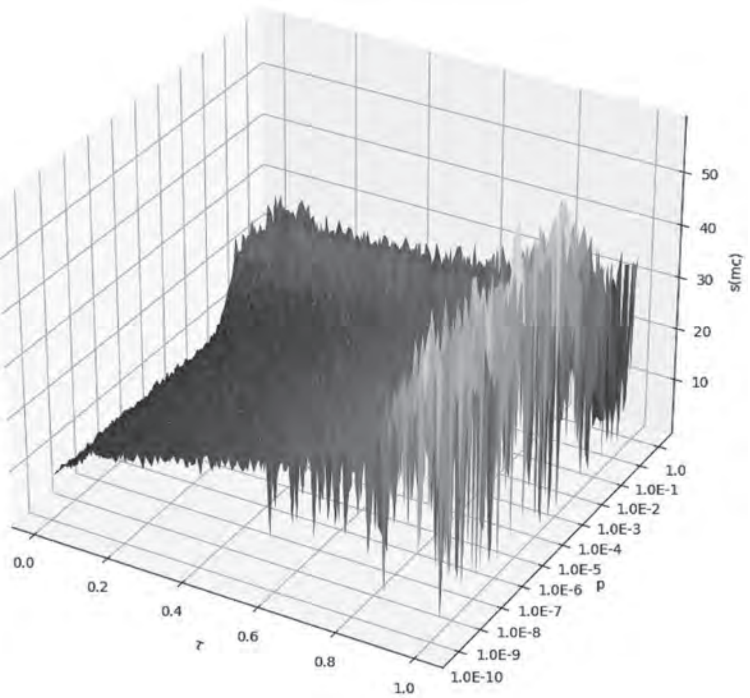
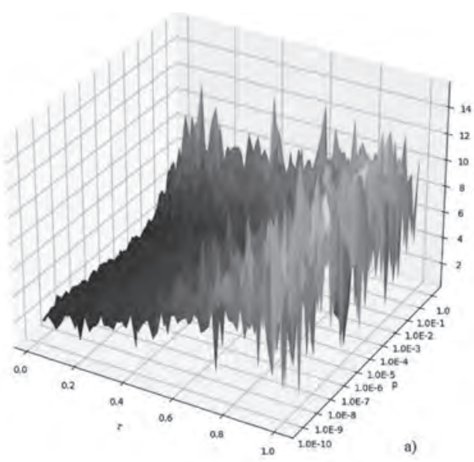


Рисунок 2. Зависимость скорости выполнения локального цикла обучения s (мс) от значений τ и p при сигмоидальной функции возбуждения нейронов



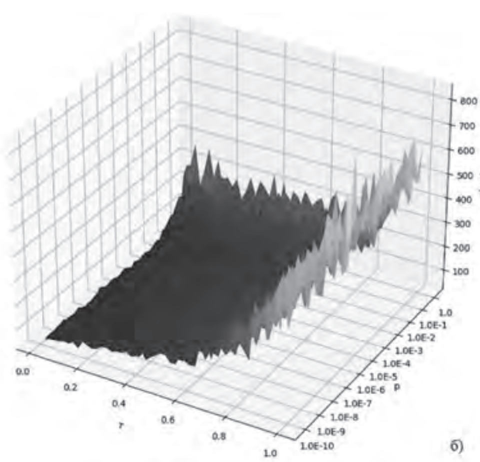


Рисунок 3. Зависимость s (мс) от значений τ и p при сигмоидальной функции возбуждения нейронов и топологиях НС: *а* – 2-5-1; *б* – 5-10-2

При применении гиперболического тангенса в качестве функции возбуждения нейронов зависимость среднего значения скорости выполнения локального цикла обучения s (мс) от значений τ и p имеет вид, представленный на Рисунке 4.

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление», выпуск 2 за 2025 год

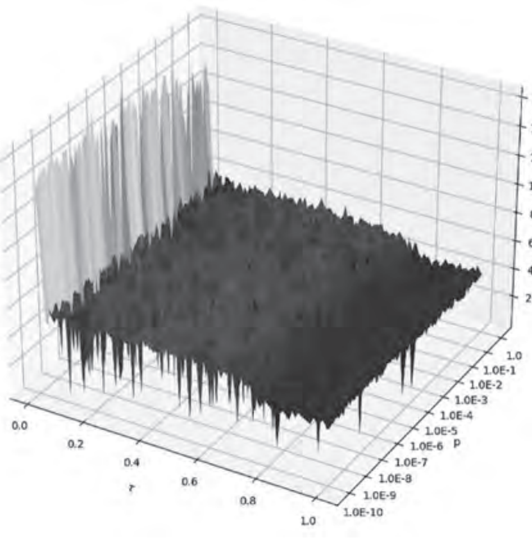


Рисунок 4. Зависимость среднего значения скорости выполнения локального цикла обучения s (мс) от значений τ и p при функции возбуждения нейронов – гиперболический тангенс

В этом случае (Рисунок 4) также сохраняется значительная неравномерность, но, в отличие от предыдущего случая (сигмоидальная функция) поверхность значений убывает от точки $\tau = 0$ и $p = 10^{-10}$ до точек диапазона $\tau = [0,9 \ldots 1,0]$ и $p = \left[10^{-9} \ldots 10^{-1}\right]$. Получено минимальное значение $s = 0,4163$ (мс) при $\tau = 0,64$ и $p = 10^{-1}$.

Также исследованы многослойные НС, в качестве примера приведены НС с топологиями 2-5-1 (Рисунок 5, *а*) и 2-3-3-3-1 (Рисунок 5, *б*). В связи с увеличением неравномерности тренд на убывание выражен значительно меньше.

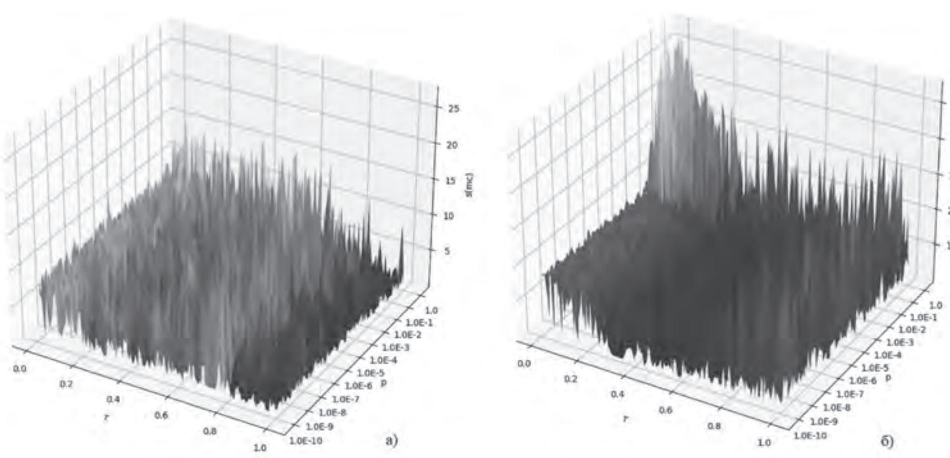


Рисунок 5. Зависимость s (мс) от значений τ и p при топологиях НС:
а – 2-5-1; *б* – 2-3-3-3-1 при функции возбуждения нейронов – гиперболический тангенс

Из результатов проведенных исследований становится понятно, что среднее значение скорости выполнения локального цикла обучения можно представить в виде двумерной случайной величины $s(\tau, p)$ и вычислить ее численные характеристики [11], такие как среднее значение $M\left[s(\tau, p)\right]$ и стандартное отклонение $\sigma\left[s(\tau, p)\right]$. Используя возможности библиотеки NumP языка программирования Python [12] для выполнения расчетов, получим следующую подпрограмму:

```
import numpy as np
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d import Axes3D

def calc_3D_mean(data, column_name, row_name):
    plt.figure()
    plt.subplot(121)
    axis_0 = np.mean(data, axis=0)
    plt.plot(axis_0)
    plt.xlabel("N")
    plt.ylabel("S(mc)")
    plt.title('Среднее по '+column_name)
    plt.subplot(122)
    axis_1 = np.mean(data, axis=1)
    plt.plot(axis_1)
    plt.title('Среднее по '+row_name)
    plt.xlabel("N")
    plt.ylabel("S(mc)")
    plt.suptitle('M[s('+column_name+', '+row_name+')] =
                 '+str(np.mean(data))+' $\sigma$[s('+column_name+', '+row_name+')] =
                 '+str(np.std(data)))
    plt.show()
```

На Рисунке 6 представлены результаты расчетов среднего значения и стандартного отклонения для нейронной сети с конфигурацией 2-1 (см. Рисунок 2). Поскольку данные представляют собой двумерный массив, то также рассчитаны средние значения по τ и p.

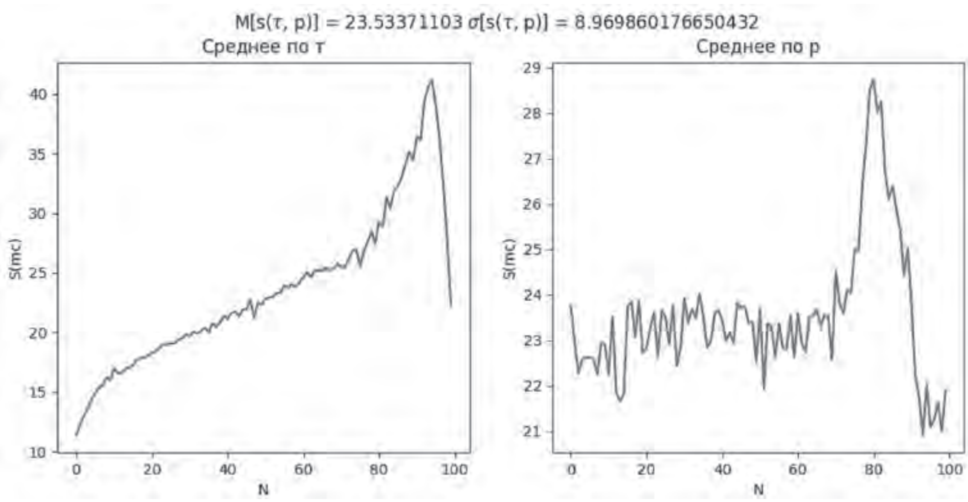


Рисунок 6. Результаты расчетов $M\left[s(\tau, p)\right]$ и $\sigma\left[s(\tau, p)\right]$ для зависимости, представленной на Рисунке 2

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление», выпуск 2 за 2025 год

Для зависимостей, представленных на Рисунке 3, *а*, *б*, результаты расчетов характеристик изображены на Рисунках 7, 8.

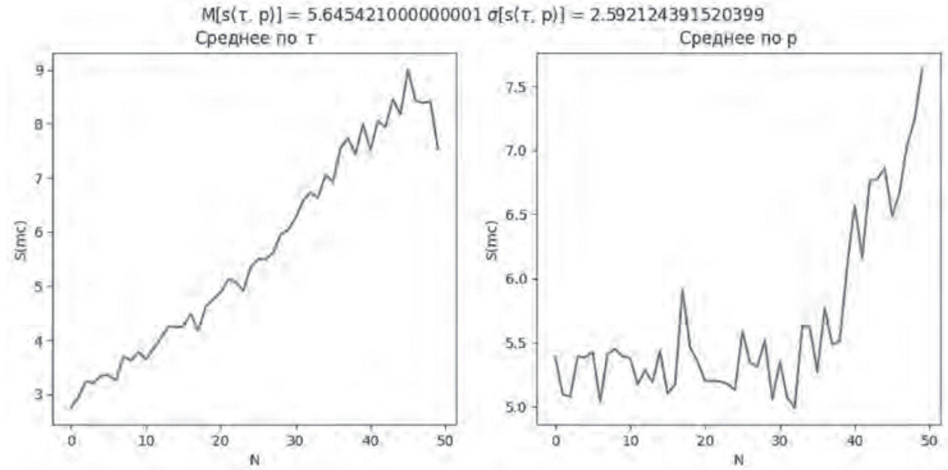


Рисунок 7. Результаты расчетов $M\left[s(\tau, p)\right]$ и $\sigma\left[s(\tau, p)\right]$ для зависимости, представленной на Рисунке 3, *а*

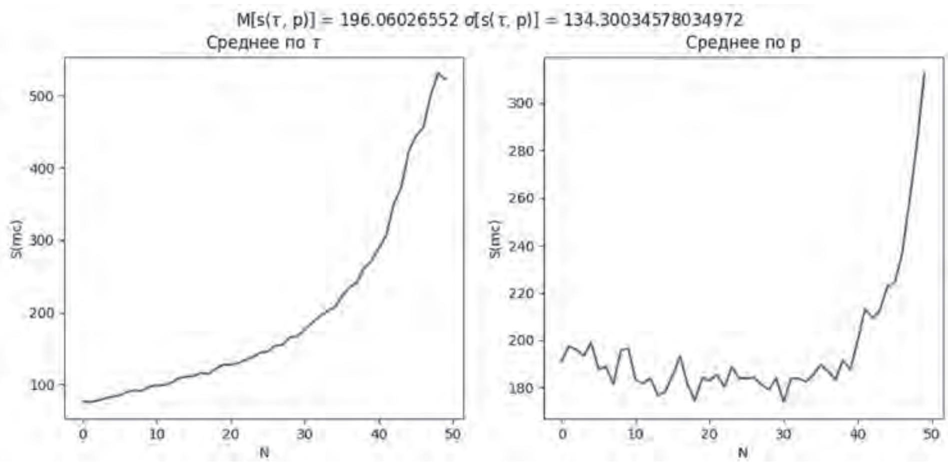


Рисунок 8. Результаты расчетов $M\left[s(\tau, p)\right]$ и $\sigma\left[s(\tau, p)\right]$ для зависимости, представленной на Рисунке 3, *б*

Как было установлено ранее, использование гиперболического тангенса приводит к значительному увеличению неравномерности полученных результатов. Поэтому расчет численных характеристик двумерной случайной величины $s(\tau, p)$ может быть полезным инструментом для анализа результатов исследований. В качестве примера на Рисунке 9 представлены результаты, полученные для нейронной сети, схема которой показана на Рисунке 4. Аналогичным образом результаты, соответствующие схемам, изображенным на Рисунке 5, *а*, *б*, представлены на Рисунках 10, 11 соответственно.

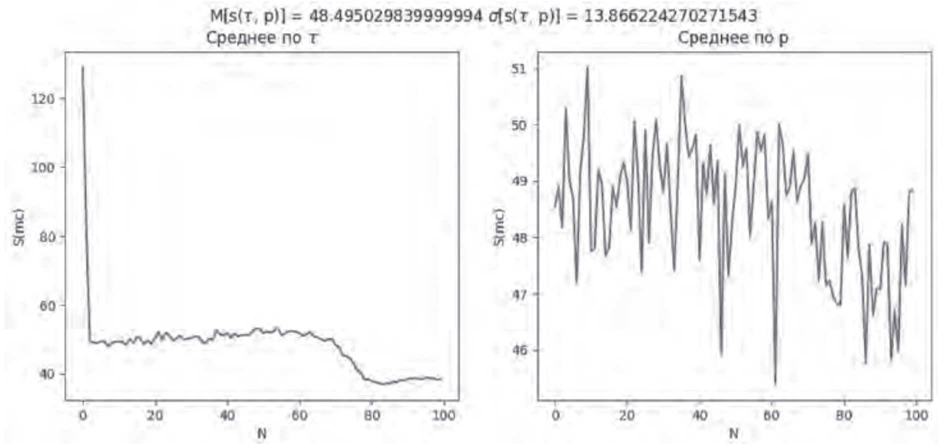


Рисунок 9. Результаты расчетов $M\left[s(\tau, p)\right]$ и $\sigma\left[s(\tau, p)\right]$ для зависимости, представленной на Рисунке 4

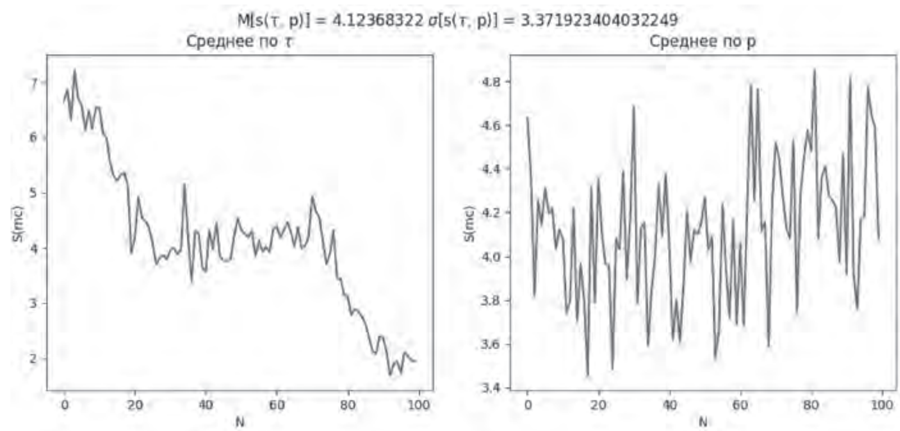


Рисунок 10. Результаты расчетов $M\left[s(\tau, p)\right]$ и $\sigma\left[s(\tau, p)\right]$ для зависимости, представленной на Рисунке 5, *а*

На основании приведенных результатов моделирования можно сформулировать основные положения метода уменьшения времени выполнения процесса обучения НС с учетом оценки значений коэффициента τ и порога p, в соответствии с которыми производится корректировка параметра скорости $\eta(n)$.

1. Суть метода заключается в адаптивной настройке шага дискретного изменения весовых коэффициентов НС в процессе ее обучения. Корректировка параметра скорости производится по мере приближения (удаления) значения функции потерь к минимуму функции в соответствии с выражением (1).

2. Перед началом процесса в соответствии с выражением (1) устанавливаются параметры τ и p, $\tau \in [0,1]$ и $p \in [0,1]$. Значения этих параметров существенно зависят от топологии НС и функции возбуждения нейронов.

3. Для расчета значений численным методом разработан алгоритм (Рисунок 1) и соответствующее программное обеспечение. Результат – получение значений τ и p для минимальной скорости обучения и оценка поверхности решений (например, Рисунок 2).

4. Используя представленное программное обеспечение и рассмотренный подход для НС заданной структуры, можно выполнить оценку длительности одного цикла обучения на основе расчета среднего значения $M\left[s(\tau, p)\right]$, стандартного отклонения $\sigma\left[s(\tau, p)\right]$ и изменения среднего по τ и по p, что позволит оценить временные затраты на обучение проектируемой сети.

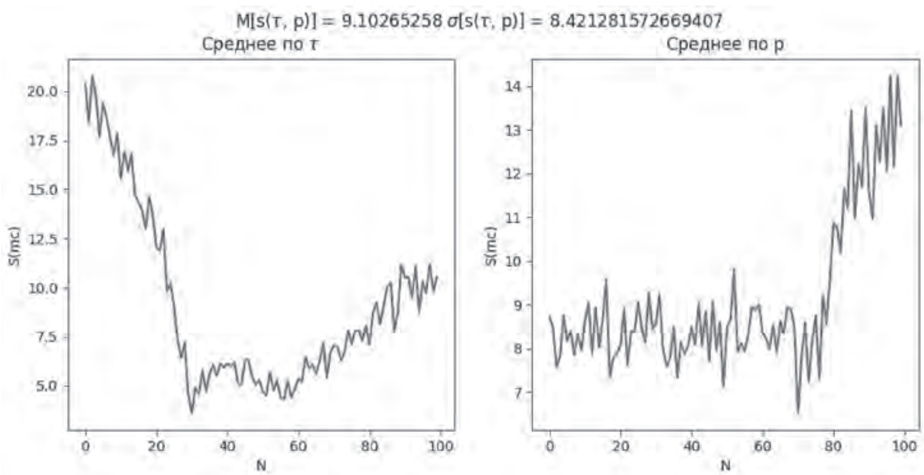


Рисунок 11. Результаты расчетов $M\left[s(\tau, p)\right]$ и $\sigma\left[s(\tau, p)\right]$ для зависимости, представленной на Рисунке 5, *б*

Следует отметить, что целесообразность применения метода может наблюдаться в тех случаях, когда проектируемая НС предназначена для многократного обучения на различных входных данных без изменения ее структуры.

Литература

1. Rumelhart D., Hinton G., Williams R. Learning representations by back-propagating errors // Nature. 1986. Vol. 323. P. 533–536. DOI: https://doi.org/10.1038/323533a0

2. *Каширина И.Л., Демченко М.В.* Исследование и сравнительный анализ методов оптимизации, используемых при обучении нейронных сетей // Вестник ВГУ. Серия: Системный анализ и информационные технологии. 2018. № 4. С. 123–132. DOI: 10.17308/sait.2018.4/1262. EDN YTRTOP.

3. *Перков А.С., Жангиров Т.Р., Лисс А.А., Григорьева Н.Ю. Чистякова Л.В.* Сравнение методов обучения нейронных сетей в задаче классификации // Известия СПбГЭТУ «ЛЭТИ». 2019. № 6. С. 53–61. EDN XETRHR.

4. *Толстых А.А., Голубинский А.Н.* Сравнение эффективности методов изменения скорости обучения искусственных нейронных сетей в различных задачах классификации // Международный научно-исследовательский журнал. 2022. № 7 (121). С. 102–106. DOI: 10.23670/IRJ.2022.121.7.013. EDN IDFXSS.

5. *Kingma D.P., Ba J.* Adam: A Method for Stochastic Optimization // ArXiv. 2014. DOI: https://doi.org/10.48550/arXiv.1412.6980

6. *Sun R.* (2020). Optimization for deep learning: an overview // Journal of the Society of Operations Research of China. Vol. 8. No. 2. P. 249–294. DOI: 10.1007/s40305-020-00309-6. EDN RLWRBK.

7. *Хайкин С.* Нейронные сети : полный курс. / Пер. с англ. Н.Н. Куссуль, А.Ю. Шелестова. 2-е изд. М. : Вильямс, 2006. 1104 с. ISBN 978-5-8459-0890-2.

8. *Чепцов М.Н., Сонина С.Д.* Модель оптимизации параметра скорости обучения нейронной сети // Сборник научных трудов Донецкого института железнодорожного транспорта. 2021. № 62. С. 28–32. EDN CQVNKA.

9. *Momot A., Galagan R.* Influence of architecture and training dataset parameters on the neural networks efficiency in thermal nondestructive testing // Sciences of Europe. 2019. No. 44-1 (44). Pp. 20–25. URL: https://core.ac.uk/download/323534566.pdf (дата обращения: 12.12.2024).

10. Shapovalova S., Moskalenko Yu. (2020). Methods for increasing the classification accuracy based on modifications of the basic architecture of convolutional neural networks // ScienceRise. No. 6 (71). Pp. 10–16. DOI: 10.21303/2313-8416.2020.001550. EDN GEMBIM.

11. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике. Часть вторая. Случайные величины. М. : Высшая школа, 2004. 400 с. ISBN 5-06-004212-X.

12. *Harris C.R., Millman K.J., van der Walt S.J., et al.* (2020) Array programming with NumPy // Nature. Vol. 585. P. 357–362. DOI: 10.1038/s41586-020-2649-2

References

1. Rumelhart D., Hinton G., Williams R. (1986) Learning representations by back-propagating errors. *Nature*. Vol. 323. Pp. 533–536. DOI: https://doi.org/10.1038/323533a0

2. Kashirina I.L., Demchenko M.V. (2018) Research and comparative analysis of optimization methods used in the teaching of neural networks. *Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*. No. 4. Pp. 123–132. DOI:10.17308/sait.2018.4/1262

3. Perkov A.S., Zhangirov T.R., Liss A.A., Grigor'eva N.Yu., Chistyakova L.V. (2019) Comparison of neural network training methods in the classification problem. *LETI Transactions on Electrical Engineering & Computer Science.* No. 6. Pp. 53–61. ISSN: 2071-8985

4. Tolstykh A.A., Golubinskiy A.N. (2022) Comparison of the effectiveness of methods for changing the learning rate of artificial neural networks in various classification tasks. *International Research Journal.* No. 7 (121). Pp. 102–106. DOI: 10.23670/IRJ.2022.121.7.013

5. Kingma D.P., Ba J. (2014) Adam: A Method for Stochastic Optimization. *ArXiv.* DOI: https://doi.org/10.48550/arXiv.1412.6980

6. Sun R. (2020). Optimization for deep learning: An overview. *Journal of the Society of Operations Research of China*. Vol. 8. No. 2. Pp. 249–294. DOI: 10.1007/s40305-020-00309-6

7. Haykin S. (1999) *Neural Networks*. Prentice Hall, NJ : Upper Saddle River. 872 p. ISBN 0132733501. (Russian edition: transl. by N.N. Kussul', A.Yu. Shelestov. Moscow : Vil'yams Publ. 1104 p.)

8. Cheptsov M.N., Sonina S.D. (2021) Optimization model of the neural network learning rate parameter. *Collection of scientific papers of the Donetsk Railway Transport Institute*. No. 62. Pp. 28–32.

9. Momot A., Galagan R. (2019) Influence of architecture and training dataset parameters on the neural networks efficiency in thermal nondestructive testing. *Sciences of Europe.* No. 44-1 (44). Pp. 20–25. URL: https://core.ac.uk/download/323534566.pdf (accessed 12.12.2024).

Вестник Российского нового университета
Серия «Сложные системы: модели, анализ и управление», выпуск 2 за 2025 год

10. Shapovalova S., Moskalenko Y. (2020). Methods for increasing the classification accuracy based on modifications of the basic architecture of convolutional neural networks. *ScienceRise.* No. 6 (71). Pp. 10–16. DOI: 10.21303/2313-8416.2020.001550

11. Gmurman V.E. (2004) *Rukovodstvo k resheniyu zadach po teorii veroyatnostey i matematicheskoy statistike* [Guide to solving problems in probability theory and mathematical statistics]. Part two. Random values. Moscow : Vysshaya shkola Publ. 400 p. ISBN 5-06-004212-X.

12. Harris C.R., Millman K.J., van der Walt S.J., et al. (2020) Array programming with NumPy. *Nature.* Vol. 585. Pp. 357–362. DOI: 10.1038/s41586-020-2649-2

Поступила в редакцию: 18.04.2025 Received: 18.04.2025
Поступила после рецензирования: 15.05.2025 Revised: 15.05.2025
Принята к публикации: 30.05.2025 Accepted: 30.05.2025