ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ

УДК 519.7 Научная статья
DOI: 10.35330/1991-6639-2025-27-5-159-167
EDN: DASTQK

Робастный оптимизатор Adam на основе усредняющих агрегирующих функций

М. А. Казаков

Институт прикладной математики и автоматизации –
филиал Кабардино-Балкарского научного центра Российской академии наук
360000, Россия, г. Нальчик, ул. Шортанова, 89 А

Аннотация. Обучение на загрязненных данных (выбросы, тяжелые хвосты, шум меток, артефакты предобработки) делает арифметическое усреднение в эмпирическом риске неустойчивым: несколько аномалий смещают оценки, дестабилизируют шаги оптимизации и ухудшают обобщающую способность. Требуется способ повысить робастность без изменения функции потерь и архитектуры модели.

Цель исследования. Разработать и продемонстрировать вариант Adam, в котором усреднение по партии (batch) заменено на робастную усредняющую агрегирующую функцию на основе штрафа, позволяющую ослабить влияние выбросов при сохранении преимуществ момента и покоординатной адаптации шага.

Методы исследования. Используются усредняющие агрегирующие средние на базе штрафной функции. В качестве функций несходства используется функция Хубера. Для нахождения робастного центра и весов элементов партии используется метод Ньютона. Эффективность оценивается на синтетическом эксперименте с контролируемыми выбросами через сравнение со стандартным Adam по устойчивости обучения.

Результаты. Робастный Adam показал более устойчивое обучение на синтетической линейной регрессии: при наличии до 20 % выбросов итоговая модель сохраняет устойчивость. Метод сохраняет вычислительную эффективность и совместимость, добавляются лишь несколько итераций поиска робастного центра на партию, асимптотика не меняется. При квадратичной штрафной функции он вырождается в обычный Adam, что подтверждает корректность обобщения.

Выводы. При помощи М-средних произведена модификация алгоритма оптимизации Adam. Данный алгоритм сохраняет робастность при линейной регрессии при наличии выбросов по крайней мере до 20 %. Точные ограничения еще подлежат исследованию. Накладные вычислительные расходы связаны с вычислением оптимального значения u^* для каждой партии. Однако в силу быстрой сходимости (около трех итераций по методу Ньютона) замедление алгоритма несущественное.

Ключевые слова: робастная оптимизация, выбросы, тяжелые хвосты, шум в данных, агрегирующие функции, эмпирический риск, Adam, адаптивные методы градиентной оптимизации, стохастический градиентный спуск по мини-партиям, глубокое обучение, робастная статистика, машинное обучение

Поступила 14.07.2025, одобрена после рецензирования 26.08.2025, принята к публикации 25.09.2025

Для цитирования. Казаков М. А. Робастный оптимизатор Adam на основе усредняющих агрегирующих функций // Известия Кабардино-Балкарского научного центра РАН. 2025. Т. 27. № 5. С. 159–167. DOI: 10.35330/1991-6639-2025-27-5-159-167

MSC: 68T20 Original article

Robust ADAM optimizer based on averaging aggregation functions

M.A. Kazakov

Institute of Applied Mathematics and Automation –
branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences
89 A Shortanov street, Nalchik, 360000, Russia

Abstract. Training on contaminated data (outliers, heavy tails, label noise, preprocessing artifacts) makes arithmetic averaging in empirical risk unstable: multiple anomalies bias estimates, destabilize optimization steps, and degrade generalization ability. There is a need for a way to improve the robustness without changing the loss function or model architecture.

Aim. The paper aims to develop and demonstrate an alternative approach to batch averaging in ADAM, replacing it with a robust penalty-based averaging aggregation function, which mitigates the influence of outliers, while still maintaining the benefits of moment-based and coordinate- wise step adaptation.

Methods. Penalized, averaging aggregation means are used. The Huber dissimilarity function is used. Newton's method is used to find the optimal center and weights for batch elements. Performance is evaluated in a controlled experiment with synthetic outliers, by comparing it to the standard ADAM algorithm for training stability.

Results. Robust ADAM showed more robust training for synthetic linear regression, with the resulting model remaining stable even with up to 20% of outliers. The method keeps providing computational efficiency and compatibility by adding only a small number of iterations of the robust center search to each batch, while sustaining the same asymptotic behavior. With a quadratic penalty function, it degenerates into standard Adam, confirming the validityof the generalization.

Conclusion. A modification of the Adam optimization algorithm has been made using M-means. This method ensures the stability of linear regression, with outliers even up to 20%. The exact limitations are still to be determined. Computational overhead is associated with calculating the optimal value for each batch. However, due to the rapid convergence (approximately three iterations using Newton's method), the algorithm slowdown is not significant.

Keywords: robust optimization, outliers; heavy-tailed distributions, data noise, aggregation functions, empirical risk, Adam, adaptive gradient optimization methods, mini-batch stochastic gradient descent, deep learning, robust statistics, machine learning

Submitted on 14.07.2025, approved after reviewing on 26.08.2025, accepted for publication on 25.09.2025

For citation. Kazakov M.A. Robust ADAM optimizer based on averaging aggregation functions. *News of the Kabardino-Balkarian Scientific Center of RAS.* 2025. Vol. 27. No. 5. Pp. 159–167. DOI: 10.35330/1991-6639-2025-27-5-159-167

Введение

Адаптивные методы стохастической оптимизации [1–4], прежде всего Adam [1], стали стандартом при обучении глубоких моделей благодаря устойчивости к нестационарному шуму градиентов и эффективной покоординатной адаптации шага. Однако их эффективность существенно снижается в присутствии выбросов, тяжелохвостого шума и систематических загрязнений данных (например, случайных или недобросовестных меток). Причина в классической постановке минимизации эмпирического риска, где усреднение по выборке выполняется арифметическим средним: несколько крупных отклонений вносят непропорционально большой вклад и смещают решение.

Робастная статистика предлагает широкий арсенал средств для подавления влияния выбросов: от замены среднего медианой и квантилями до М-оценивания с гладкими штрафными функциями, такими как функция Хубера или Тьюки [4–10]. Вместо изменения самой функции потерь для каждого наблюдения или агрессивного клиппинга градиентов в работе рассматривается более общий и согласованный с оптимизацией подход: арифметическое усреднение в эмпирическом риске заменяется на усредняющую агрегирующую функцию, построенную по штрафной функции [11–14].

В данной работе предлагается модификация Adam, в которой роль эмпирического риска по партии (batch) играет М-среднее от значений функции потерь, а градиент по параметрам вычисляется как взвешенная сумма градиентов по объектам. Веса v_i вычисляются на основе второй производной выбранной функции несходства и зависят от отклонения $L_i(\theta) - u^*$ относительно робастного центра u^*, который вычисляется как минимум штрафной суммы. Такая конструкция обладает рядом желательных свойств: веса неотрицательны и автоматически нормируются в единицу, убывают при росте остатка и тем самым подавляют вклад выбросов; при квадратичной функции несходства метод вырождается в стандартное арифметическое среднее и совпадает с обычным Adam.

Практически значимо, что предлагаемая модификация минимально вмешивается в архитектуру Adam, поэтому сохраняются преимущества инерции (момента) и покоординатной адаптации шага, а устойчивость к выбросам достигается за счет корректного агрегирования сигналов из выборки. Дополнительные вычисления ограничиваются несколькими итерациями оценки u^* и не меняют асимптотику по размеру партии.

Взвешенный эмпирический риск

Функция потерь – это отображение $L: Y \times Y \to R_{\geq 0}$, количественно измеряющее несоответствие между целевым значением $y_i \in Y$ и прогнозом $\hat{y}_i \in Y$ для объекта $x_i \in X$ выборки.

Пусть $\theta \in \Theta$ – параметры модели $f_\theta: X \to Y$. Тогда прогнозируемое значение

$$\hat{y}_i = \hat{y}(\theta, x_i) = f_\theta(x_i).$$

Пусть задана выборка $S = \{(x_i, y_i)\}_{i=1}^n$. *Эмпирический риск* (ER) представляет собой среднее арифметическое значений функций потерь на S [15]:

$$\mathrm{ER}(\theta) = \frac{1}{n}\sum_{i=1}^{n} L(\hat{y}(\theta, x_i), y_i) = \frac{1}{n}\sum_{i=1}^{n} L_i(\theta).$$

При наличии выбросов в выборке S среднее арифметическое приводит к большим искажениям, в результате которых решение смещается в сторону выбросов. Одним из способов решения является использование медианы или квантилей вместо среднего арифметического.

В рамках данной работы представляет интерес *взвешенный эмпирический риск* с весами $\mathrm{v} = (v_1, \ldots, v_n)$:

$$\mathrm{ER}_\mathrm{v}(\theta) = \frac{1}{n}\sum_{i=1}^{n} v_i L_i(\theta),$$

в частности, с нормированными весами $\sum v_i = 1$. При низких значениях v_i для выбросов эмпирический риск будет устойчив по отношению к наличию выбросов.

Задача обучения формулируется как минимизация эмпирического риска (иногда с регуляризацией $\lambda\Omega(\theta)$):

$$\min_{\theta\in\Theta} \mathrm{ER}_v(\theta) \quad \text{или} \quad \min_{\theta\in\Theta} \mathrm{ER}_v(\theta)+\lambda\Omega(\theta).$$

Оптимизатор Adam позволяет эффективно находить параметры θ, при которых достигается минимум эмпирического риска (по крайней мере локальный), а М-средние используются для поиска v_i, снижающих влияние выбросов.

Оптимизатор ADAM

Полный градиентный спуск является стабильным, но при этом довольно дорогим: один шаг требует прохода по всей выборке. Предпочтительнее использовать стохастический градиентный спуск, например, по мини-партиям. Пусть выборка состоит из N элементов. На каждом временно́м этапе производится семплирование индексов $B^{(t)} \subset \{1,\ldots,N\}$, $b = |B^{(t)}|$ – размер мини-партии. Тогда градиент по мини-партии имеет вид

$$\mathrm{g}^{(t)} = \frac{1}{b}\sum_{i\in B^{(t)}} \nabla L_i(\theta^{(t)}).$$

Во взвешенном эмпирическом риске слагаемые домножаются на соответствующие веса v_i. При минимизации эмпирического риска на каждой итерации производится обновление

$$\theta^{(t+1)} = \theta_t - \alpha \mathrm{g}^{(t)}.$$

где α – скорость обучения. На ландшафте функции потерь часто возникают анизотропные области в виде узких долин, из-за которых при стохастическом градиентном спуске движение может стать зигзагообразным (быстрые колебания поперек долины и медленное продвижение вдоль нее). Для решения этой проблемы Поляк ввел импульс (по аналогии с физическим импульсом), призванный аккумулировать инерцию в направлении стабильного спуска и ускорять тем самым движение вдоль долины [3]. Кроме того, инерция позволяет снизить чувствительность к масштабу признаков за счет временно́го сглаживания. Поляк определял эту инерцию через приращение параметров, но в глубоком обучении чаще используется вариант с экспоненциальным средним градиентов:

$$\mathrm{m}^{(t)} = \beta_1 \mathrm{m}^{(t-1)} + (1-\beta_1)\mathrm{g}^{(t)}, \qquad \theta^{(t+1)} = \theta^{(t)} - \alpha \mathrm{m}^{(t)}.$$

Даже с импульсом стохастический градиентный спуск остается чувствительным к разным масштабам и дисперсиям по координатам: там, где градиенты шумные и велики по модулю, шаги стоит уменьшать и, наоборот, увеличивать шаги для небольших, но стабильных градиентов. Исторически первая систематическая адаптация была произведена в алгоритме AdaGrad, но в алгоритме Adam используется вариант из алгоритма RMSProp, предложенного Телеманом и Хинтоном [2]. Здесь используется экспоненциальное сглаживание квадратов градиентов:

$$\mathrm{v}^{(t)} = \beta_2 \mathrm{v}^{(t-1)} + (1-\beta_2)(\mathrm{g}^{(t)} \odot \mathrm{g}^{(t)}), \quad \mathrm{v}^{(0)} = 0,$$

$$\theta^{(t+1)} = \theta^{(t)} - \alpha \frac{\mathrm{g}^{(t)}}{\sqrt{\mathrm{v}^{(t)}} + \varepsilon}.$$

где $\odot$ – оператор поэлементного умножения матриц, коэффициент $\beta_2 \in [0,1)$ задает горизонт забывания масштаба; обычно коэффициент выбирается в диапазоне от 0,9 до 0,99. Такое сглаживание стабилизирует шаги в нестационарном шуме и подавляет колебания в узких долинах за счет покоординатной нормализации.

В практическом плане RMSProp хорошо сочетается с импульсом:

$$\mathrm{m}^{(t)} = \beta_1 \mathrm{m}^{(t-1)} + (1-\beta_1)\mathrm{g}^{(t)}, \qquad \theta^{(t+1)} = \theta^{(t)} - \alpha \frac{\mathrm{m}^{(t)}}{\sqrt{\mathrm{v}^{(t)}} + \varepsilon}.$$

Оптимизатор Adam строится на базе этих двух алгоритмов, но с одной важной поправкой [1]. Стартовые значения $\mathrm{m}^{(0)}$ и $v^{(0)}$ инициализируются тензорами, заполненными нулями. Это приводит к смещению к нулю, в частности, на ранних итерациях. Для учета этого в статье было предложено производить коррекцию смещения

$$\mathrm{m}^{(t)} = \frac{\mathrm{m}^{(t)}}{1-\beta_1^t}, \quad \mathrm{v}^{(t)} = \frac{\mathrm{v}^{(\mathrm{t})}}{1-\beta_2^t}$$

и обновление

$$\theta^{(t+1)} = \theta^{(t)} - \alpha \frac{\mathrm{m}^{(t)}}{\sqrt{\mathrm{v}^{(t)}} + \varepsilon}.$$

Это сочетает инерцию направления с адаптацией масштаба без деградации шага со временем.

M-СРЕДНИЕ

Для того чтобы при оптимизации минимизировать влияние выбросов, представим эмпирический риск в виде усредняющей агрегирующей функции:

$$\mathrm{ER}(\mathrm{w}) = \mathrm{M}\{L_1(\theta), \ldots, L_n(\theta)\}.$$

Будем использовать агрегирующую функцию на базе штрафной функции

$$P(z_1, \ldots, z_n, u) = \sum_{k=1}^{n} p(z_k, u),$$

где $p(z_k, u)$ – функция несходства. Тогда усредняющая агрегирующая функция определяется как

$$\mathrm{M}_p\{z_1, \ldots, z_n\} = \arg\min_u P(z_1, \ldots, z_n, u),$$

если минимум штрафной функции является синглетон, или

$$\mathrm{M}_p\{z_1, \ldots, z_n\} = \frac{a+b}{2},$$

если минимум представляет собой интервал с концами a и b.

Уникальность минимума $P(z_1, \ldots, z_n, u)$ и монотонность $\mathrm{M}_p\{z_1, \ldots, z_n\}$ гарантированы, когда

$$p(z,u) = K(h(z) - h(u)),$$

где $K: \mathbb{R}^2 \to \mathbb{R}$ – непрерывная строго монотонная функция; $h(u)$ – строго монотонная функция [11, 12].

Будем использовать функции несходства вида $p(z,u) = \rho(z-u)$. Агрегирующую функцию M_p на основе штрафной функции с такой функцией несходства будем называть М-средним [14]. Если ρ имеет частные производные по u, то $u^* = \bar{z} = \mathrm{M}_p\{z_1,\ldots,z_n\}$ является решением уравнения

$$\sum_{k=1}^{n} \rho_u(z_k - \bar{z}) = 0 .$$

Из этого следует, что сумма производных функции ρ по аргументу также равна нулю:

$$\sum_{k=1}^{n} \rho'(z_k - \bar{z}) = 0 .$$

Если ρ дважды дифференцируема, то $\mathrm{M}_\rho\{z_1,\ldots,z_n\}$ имеет все частные производные по z_k:

$$\frac{\partial \mathrm{M}_\rho}{\partial z_k} = \frac{\rho''(z_k - \bar{z})}{\rho''(z_1 - \bar{z}) + \ldots + \rho''(z_n - \bar{z})} .$$

В этом случае $\frac{\partial \mathrm{M}_\rho}{\partial z_k} \geq 0$ и $\frac{\partial \mathrm{M}_\rho}{\partial z_1} + \ldots + \frac{\partial \mathrm{M}_\rho}{\partial z_n} = 1$.

В частности, если используется функция несходства $(z-u)^2/2$, то М-среднее представляет собой среднее арифметическое [14].

М-средние от функций потерь

Оптимальные значения θ^* параметров модели могут быть найдены путем минимизации М-средних от функций потерь:

$$\mathrm{ER}(\theta) = \mathrm{M}_p\{L_1(\theta),\ldots,L_n(\theta)\} .$$

Штрафная функция, на основе которой строится такая агрегирующая функция, имеет вид

$$P(L_1(\theta),\ldots,L_n(\theta),u) = \sum_{k=1}^{n} \rho(L_k(\theta) - u) .$$

Условие оптимальности:

$$\mathrm{M}_p\{L_1(\theta),\ldots L_n(\theta)\} = \arg\min_u \sum_{k=1}^{n} \rho(L_k(\theta) - u) = u^* .$$

Если функция несходства ρ дважды дифференцируема, то

$$\frac{\partial \mathrm{M}_p}{\partial L_k} = \frac{\rho''(L_k(\theta) - u^*)}{\rho''(L_1(\theta) - u^*) + \ldots + \rho''(L_n(\theta) - u^*)} .$$

Тогда, обозначив $v_k = \frac{\partial \mathrm{M}_p}{\partial L_k}$, для градиента от функции потерь по параметрам можно записать:

$$\text{grad}\mathcal{Q}(\theta) = \sum_{k=1}^{n} v_k(\theta)\text{grad}L_k(\theta).$$

Поиск оптимального u^* можно осуществлять градиентным методом или методом Ньютона. В качестве функции несходства можно взять функцию Хубера или функцию Тьюки [6, 7]. В общем случае функция несходства должна быть гладкой и строго выпуклой, чтобы гарантировать единственность оптимального значения u.

Алгоритм

Ниже приведен псевдокод робастного алгоритма оптимизации Adam с использованием М-среднего.

Инициализация при $t = 0$

repeat

1. $L_i(\theta) \leftarrow L(f(x_i), y_i), \quad i \in B^{(t)}$

2. Схема Ньютона для поиска u^*:

   a. $u^{(0)} = \frac{1}{b}\sum_{i \in B} L_i$

   b. $u^{(r)} \leftarrow u^{(r-1)} - \frac{\sum_i \rho'(L_i - u^{(r-1)})}{\sum_i \rho''(L_i - u^{(r-1)}) + \delta}$ для $r = 1, \ldots, K$

   c. $u^* \leftarrow u^{(K)}$

3. $v_i = \frac{\rho''(L_k(\theta) - u^*)}{\rho''(L_1(\theta) - u^*) + \ldots + \rho''(L_n(\theta) - u^*) + \delta}, \quad i \in B^{(t)}$

4. $\mathrm{g}^{(t)} = \sum_{k=1}^{n} v_k(\theta)\text{grad}L_k(\theta)$

5. Оптимизация Adam

   a. $\mathrm{m}^{(t)} = \beta_1 \mathrm{m}^{(t-1)} + (1-\beta_1)\mathrm{g}^{(t)}$

   b. $\mathrm{v}^{(t)} = \beta_2 \mathrm{v}^{(t-1)} + (1-\beta_2)(\mathrm{g}^{(t)} \odot \mathrm{g}^{(t)})$

   c. $\hat{\mathrm{m}}^{(t)} = \frac{\mathrm{m}^{(t)}}{1-\beta_1^t}, \qquad \hat{\mathrm{v}}^{(t)} = \frac{\mathrm{v}^{(t)}}{1-\beta_2^t}$

   d. $\theta^{(t+1)} = \theta^{(t)} - \alpha \frac{\hat{\mathrm{m}}^{(t)}}{\sqrt{\hat{\mathrm{v}}^{(t)}} + \varepsilon}$

until

обучение не стабилизируется.

Для демонстрации работы алгоритма были синтезированы данные для задачи линейной регрессии, в которых 20 % точек являются выбросами. На рисунке 1 представлены прямые, полученные при помощи оптимизатора Adam и робастного оптимизатора Adam с применением М-средних. В качестве функции несходства ρ использовалась функция Хьюбера.

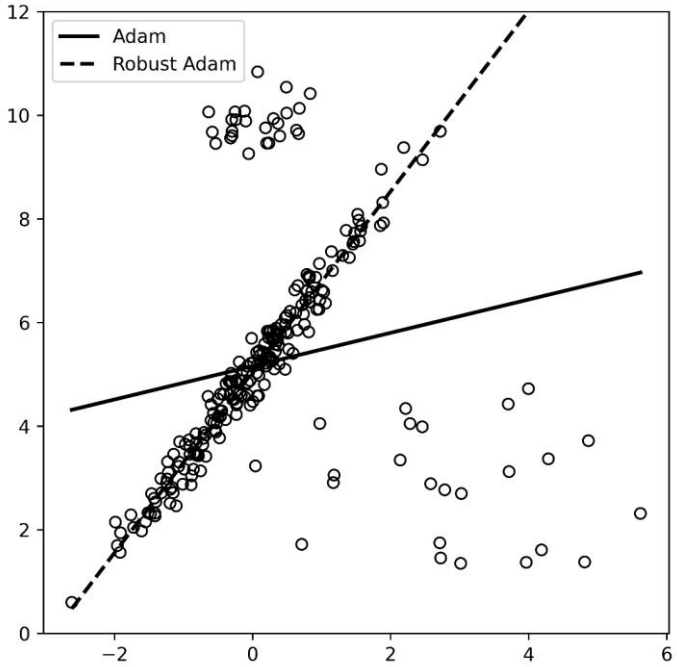

Рис. 1. *Линейная регрессия при наличии выбросов*

Fig. 1. *Linear regression in the presence of outliers*

Заключение

При помощи М-средних произведена модификация алгоритма оптимизации Adam. Данный алгоритм сохраняет робастность при линейной регрессии при наличии выбросов по крайней мере до 20 %. Точные ограничения еще подлежат исследованию. Накладные вычислительные расходы связаны с вычислением оптимального значения u^* для каждой партии. Однако в силу быстрой сходимости (около трех итераций по методу Ньютона) замедление алгоритма несущественное.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Kingma D.P., Ba J. Adam. A method for stochastic optimization. international conference on learning representations (ICLR 2015). San Diego, 2015. 15 p., available at: https://arxiv.org/abs/1412.6980

2. Tieleman T. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*. 2012. Vol. 4. No. 2. Pp. 26. DOI: https://cir.nii.ac.jp/crid/1370017282431050757

3. Поляк Б. Т. О некоторых способах ускорения сходимости итерационных методов // Журнал вычислительной математики и математической физики. 1964. Т. 4. № 5. С. 791–803. DOI: 10.1016/0041-5553(64)90137-5

Polyak B.T. Some methods of speeding up the convergence of iteration methods. *Computational Mathematics and Mathematical Physics*. 1964. Vol. 4. No. 5. Pp. 1–17. DOI: 10.1016/0041-5553(64)90137-5. (In Russian)

4. Duchi J., Hazan E., Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*. 2011. Vol. 12. Pp. 2121–2159.

5. Koenker R. Quantile Regression. Cambridge: Cambridge University Press, 2005. 366 p. ISBN: 0-521-60827-9

6. Huber P.J. Robust Statistics. New York: Wiley, 1981. 308 p. ISBN: 0-471-41805-6

7. Tukey J.W. A survey of sampling from contaminated distributions. in: contributions to probability and statistics: essays in honor of Harold Hoteling. Stanford: Stanford University Press, 1960. Pp. 448–485. DOI: https://cir.nii.ac.jp/crid/1570291226404846720

8. Rousseeuw P.J., Leroy A.M. Robust regression and outlier detection. New York: Wiley, 1987. 329 p. ISBN: 9780471852339

9. Rousseeuw P.J. Least median of square regression. *Journal of the American Statistical Association*. 1984. Vol. 79. Pp. 871–880. DOI: 10.1080/01621459.1984.10477105.

10. Vapnik V. The nature of statistical learning theory. New York: Springer-Verlag, 2000. 314 p. ISBN: 978-1-4419-3160-3

11. Beliakov G., Sola H., Calvo T. A practical guide to averaging functions. Berlin: Springer-Verlag, 2016. 371 p. ISBN: 978-3319247519

12. Calvo T., Beliakov G. Aggregation functions. *Fuzzy Sets and Systems*. 2010. Vol. 161. No. 10. Pp. 1420–1436. DOI: 10.1016/j.fss.2009.05.012

13. Mesiar R., Kolesárová A., Calvo T., Komorníková M. A review of aggregation functions. fuzzy sets and their extensions: representation, aggregation and models. studies in fuzziness and soft computing. 2008. Vol. 220. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-540-73723-0_7

14. Shibzukhov Z.M. Principle of minimizing empirical risk and averaging aggregate functions. *Journal of Mathematical Sciences*. 2021. Vol. 253. No. 4. Pp. 571–583. DOI: 10.1007/s10958-021-05256-y

15. Vapnik V. Principles of risk minimization for learning theory. *Advances in Neural Information Processing Systems* (NeurIPS). 1991. Vol. 4. Pp. 831–838. DOI: https://cir.nii.ac.jp/crid/1571698599429734144

Финансирование. Исследование проведено без спонсорской поддержки.

Funding. The study was performed without external funding.

Информация об авторе

Казаков Мухамед Анатольевич, мл. науч. сотр. отдела нейроинформатики и машинного обучения, Институт прикладной математики и автоматизации – филиал Кабардино-Балкарского научного центра Российской академии наук;

360000, Россия, г. Нальчик, ул. Шортанова, 89 А;

kasakow.muchamed@gmail.com, ORCID: https://orcid.org/0000-0002-5112-5079, SPIN-код: 6983-1220

Information about the author

Mukhamed A. Kazakov, Junior Researcher of the Department of Neuroinformatics and Machine Learning, Institute of Applied Mathematics and Automation – branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

89 A Shortanov street, Nalchik, 360000, Russia;

kasakow.muchamed@gmail.com, ORCID: https://orcid.org/0000-0002-5112-5079, SPIN-code: 6983-1220