На правах рукописи

Перминов Андрей Игоревич

Доверенный байесовский классификатор для данных малой размерности на основе многослойного персептрона

Специальность 2.3.5 —
«Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва — 2026

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте системного программирования им. В.П. Иванникова Российской Академии Наук.

Научный руководитель: кандидат физико-математических наук **Турдаков Денис Юрьевич**

Официальные оппоненты: **Яроцкий Дмитрий Александрович,** доктор физико-математических наук, Автономная некоммерческая образовательная организация высшего образования «Сколковский институт науки и технологий», профессор

Никитин Николай Олегович, кандидат технических наук, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО», руководитель группы научно-технического развития

Ведущая организация: Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук»

Защита состоится 23 апреля 2026 г. в 14:45 на заседании диссертационного совета 24.1.120.01 при Федеральном государственном бюджетном учреждении науки Институте системного программирования им. В.П. Иванникова Российской Академии Наук по адресу: 115035, г. Москва, ул. Садовническая, д. 41, ст. 2.

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного учреждения науки Института системного программирования им. В. П. Иванникова Российской академии наук.

Автореферат разослан «___» _____ 2026 г.

Ученый секретарь
диссертационного совета
24.1.120.01,
кандидат физико-математических
наук Турдаков Д.Ю.

Общая характеристика работы

Актуальность темы.

Исследования в области доверенного искусственного интеллекта направлены на обеспечение возможности применения методов машинного обучения в критически важных областях, включая государственное управление, критическую инфраструктуру, медицину и финансовые системы. В рамках этой области исследований модели должны не только демонстрировать высокое качество предсказаний, но и обладать формализованными механизмами оценки собственной уверенности, определения границ применимости и принятия статистически обоснованных решений. Реализация этих требований в общем виде нуждается в строгой математической теории, обеспечивающей формальные гарантии корректности работы моделей.

Однако в настоящее время такая теория для современных методов машинного обучения в целом отсутствует. На практике преобладают эмпирические подходы, ориентированные главным образом на оптимизацию стандартных метрик качества. Несмотря на это, нейросетевые модели получили широкое распространение благодаря своей универсальности и способности эффективно решать широкий круг прикладных задач. Но, как правило, они не обеспечивают строгого статистического обоснования принимаемых решений и контроля области своей применимости.

В то же время в математической статистике разработан развитый теоретический аппарат для задач классификации, позволяющий получать интерпретируемые результаты и строгие вероятностные гарантии. Однако область применимости классических статистических методов существенно уже по сравнению с методами машинного обучения и, в частности, нейросетевыми моделями, что ограничивает их использование в современных прикладных задачах.

Исследование основано на положениях российских и зарубежных научных школ теории распознавания образов. Методологическую базу составляют труды Ю. И. Журавлёва, К. В. Рудакова и К. В. Воронцова, а также исследования М. И. Забежайло, А. А. Грушо и А. К. Горшенина. Значительное влияние оказали фундаментальные работы по теории статистического обучения В. Н. Вапника, А. Я. Червоненкиса и Л. Девроя и вероятностным моделям К. Бишопа.

Настоящая работа делает шаг в направлении построения математической теории нейросетевых моделей на основе методов математической статистики. Исследование сосредоточено на задачах классификации в пространствах малой размерности, что позволяет сохранить формальную строгость получаемых результатов. В рамках работы предлагается статистически обоснованный доверенный классификатор на основе многослойного персептрона, обеспечивающий формализованную оценку уверенности

и определение границ компетенции модели, тем самым закладывая основу для дальнейшего развития теории доверенного искусственного интеллекта.

Целью данной работы является разработка методов построения доверенных классификаторов на основе многослойного персептрона для данных малой размерности, обеспечивающей способность к отказу от классификации вне носителя распределения, устойчивость к дисбалансу классов и интерпретируемость принимаемых решений.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Разработать метод построения доверенного объяснимого классификатора на основе многослойного персептрона, обеспечивающего статистически обоснованное оценивание апостериорных вероятностей и устойчивость к дисбалансу классов.
2. Разработать метод генерации синтетических данных, сохраняющих геометрические и статистические свойства исходного распределения, на основе разработанного метода.
3. Провести экспериментальное исследование разработанных методов для оценки устойчивости классификатора к дисбалансу классов, корректности работы вне носителя обучающего распределения и качества генерируемых синтетических данных.
4. Разработать интеллектуальную систему машинного обучения, реализующую предложенные методы и обеспечивающую решение задач классификации данных малой размерности в условиях дисбаланса классов и высокой неопределённости вне носителя распределения.

Основные положения, выносимые на защиту:

1. Теоретическая база непараметрического оценивания в условиях дисбаланса классов и малой размерности. Сформулированы и доказаны теоремы, обосновывающие асимптотическую связь между нейросетевой и гистограммной оценками апостериорной вероятности.
2. Метод построения статистически обоснованного объяснимого байесовского классификатора на основе многослойного персептрона и дерева решений.
3. Метод построения унарного классификатора, устойчивого к дисбалансу классов и позволяющего генерировать синтетические данные.
4. Интеллектуальная система машинного обучения, реализующая предложенные методы и обеспечивающая решение задач классификации данных малой размерности в условиях дисбаланса классов и высокой неопределённости вне носителя распределения.

Перечисленные положения относятся к направлениям исследований 4, 7, 8 и 9 паспорта специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»:

- п. 4. Интеллектуальные системы машинного обучения, управления базами данных и знаний, инструментальные средства разработки цифровых продуктов.
- п. 7. Модели, методы, архитектуры, алгоритмы, форматы, протоколы и программные средства человеко-машинных интерфейсов, компьютерной графики, визуализации, обработки изображений и видеоданных, систем виртуальной реальности, многомодального взаимодействия в социокиберфизических системах.
- п. 8. Модели и методы создания программ и программных систем для параллельной и распределенной обработки данных, языки и инструментальные средства параллельного программирования.
- п. 9. Модели, методы, алгоритмы, облачные технологии и программная инфраструктура организации глобально распределенной обработки данных.

Научная новизна: разработан метод построения доверенного классификатора на основе многослойного персептрона, обеспечивающего формальные гарантии корректного поведения модели. Предложен подход, позволяющий трактовать выход персептрона как статистически обоснованную оценку апостериорной вероятности и реализующий механизм осознанного отказа от классификации для объектов вне носителя обучающего распределения, что отличает его от стандартных нейросетевых методов, не имеющих подобного теоретического обоснования. Предложен метод унарной классификации, устраняющий проблему дисбаланса классов без искажающих процедур балансировки и позволяющий генерировать синтетические данные, сохраняющие геометрические и статистические свойства исходной выборки. Предложен инструмент объяснения решений классификатора – дерево eXBTree, обеспечивающее интерпретируемость модели за счёт анализа правил принятия решений и схожих прецедентов. Теоретической основой подхода является доказанная теорема о корректном поведении классификатора вне носителя распределения, что вносит вклад в развитие математических основ доверенного искусственного интеллекта для нейросетевых моделей.

Теоретическая и практическая значимость

Теоретическая значимость работы заключается в развитии статистических основ доверенной классификации на основе многослойного персептрона и в формировании формализованного подхода к оценке уверенности предсказаний нейросетевых моделей. В работе сформулирован и доказан ряд теорем, позволяющих статистически обосновать построение доверенных классификаторов, определить границы компетенции модели и описать её поведение вне носителя распределения, включая механизм

отказа от классификации. Установлена асимптотическая связь между нейросетевой и гистограммной оценками апостериорной вероятности, что подтверждает состоятельность предложенного подхода. Полученные результаты расширяют теоретическую базу непараметрического оценивания в условиях дисбаланса классов и малой размерности и формируют основу для построения математически строгой теории доверенного искусственного интеллекта.

Практическая значимость работы заключается в использовании предложенных методов при разработке инструментов доверенного искусственного интеллекта в Исследовательском Центре Доверенного Искусственного Интеллекта (ИЦДИИ) ИСП РАН. Разработанный классификатор применяется для анализа данных в условиях дисбаланса классов, обеспечивая интерпретируемость решений и повышение надёжности за счёт механизма автоматического отказа от классификации в недостоверных областях. Метод генерации синтетических данных, сохраняющих статистическую структуру оригинала, используется для безопасного расширения обучающих выборок. Реализованная система обеспечивает воспроизводимость и практическое применение подхода в задачах, требующих доверенного принятия решений.

Апробация работы. Основные результаты работы были представлены на следующих конференциях и семинарах:

- Форум «Цифровая экономика. Технологии доверенного искусственного интеллекта», Москва, 25 мая 2023 г.
- 32-я научно-техническая конференция «Методы и технические средства обеспечения безопасности информации» (МиТСОБИ), Санкт-Петербург, 26-29 июня 2023 г.
- WAIT: Workshop on Artificial Intelligence Trustworthiness, Almaty, Kazakhstan, 24 апреля 2024 г.
- Международная конференция «Иванниковские чтения», Великий Новгород, 17-18 мая 2024 г.
- II форум «Технологии доверенного искусственного интеллекта», Москва, 27 мая 2024 г.
- 33-я научно-техническая конференция «Методы и технические средства обеспечения безопасности информации» (МиТСОБИ), Санкт-Петербург, 24-27 июня 2024 г.
- MathAI 2025 The International Conference dedicated to mathematics in artificial intelligence, March 24-28, 2025 г.
- III форум «Технологии Доверенного Искусственного Интеллекта», Москва, 20 мая 2025 г.
- 34-я всероссийская конференция «Методы и технические средства обеспечения безопасности информации» (МиТСОБИ), Санкт-Петербург, 23-26 июня 2025 г.

– Международная конференция «Иванниковские чтения», Иркутск, 26-27 июня 2025 г.

Личный вклад. Все выносимые на защиту результаты получены лично автором.

Публикации. Основные результаты по теме диссертации изложены в 9 печатных изданиях, 6 из которых изданы в журналах, рекомендованных ВАК, 4 — в периодических научных журналах, индексируемых Web of Science и Scopus, 3 — в тезисах докладов. Зарегистрированы 4 программы для ЭВМ.

Личный вклад в совместные публикации является определяющим. Из 6 основных публикаций по теме диссертации одна работа [1] выполнена без соавторов. В основных публикациях по теме диссертации автору принадлежат: метод статистически обоснованного объяснимого байесовского классификатора на основе многослойного персептрона и дерева решений eXBTree и соответствующая формулировка теоремы, а также разработка системы визуализации DenseNetworkVisualizer [2], теорема, обосновывающая асимптотическую связь между нейросетевой и гистограммной оценками апостериорной вероятности, и метод построения унарного классификатора, устойчивого к дисбалансу классов [3], метод генерации синтетических данных [4; 5], метод обучения классификатора на основе многослойного персептрона на данных с пропусками [6].

Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, приводится обзор научной литературы по изучаемой проблеме, формулируется цель, ставятся задачи работы, излагается научная новизна и практическая значимость представляемой работы.

Первая глава Первая глава содержит обзор существующих подходов к решению задачи классификации в контексте требований доверенного искусственного интеллекта. Рассматриваются фундаментальные ограничения традиционных методов машинного обучения, связанные с отсутствием формализованных механизмов оценки неопределённости, определения области компетенции модели и отказа от принятия решения в условиях статистической необоснованности предсказаний.

В главе анализируются основные классы методов классификации, включая непараметрические методы, линейные модели, деревья решений, ансамблевые алгоритмы и нейросетевые модели. Показано, что историческое развитие методов классификации характеризуется переходом от статистически интерпретируемых моделей к более гибким, но менее прозрачным алгоритмам, ориентированным преимущественно на достижение высокой предсказательной точности. Установлено, что между

интерпретируемостью, теоретической обоснованностью и выразительной способностью моделей существует фундаментальный компромисс, который существенно ограничивает их применение в задачах с повышенными требованиями к надёжности и безопасности принимаемых решений.

Особое внимание уделяется проблеме неопределённости предсказаний. Рассматриваются алеаторная и эпистемическая составляющие неопределённости и их роль в формировании доверенных решений. Показано, что большинство современных методов оценки неопределённости реализуются в виде внешних процедур по отношению к базовой модели классификации и требуют дополнительной калибровки или усложнения архитектуры, что снижает их практическую применимость и не обеспечивает статистически согласованного поведения модели.

В главе подробно рассматриваются подходы к обнаружению объектов вне носителя обучающего распределения (OOD-детекция). Анализируются методы, основанные на вероятностных выходах модели, на исследовании внутренних представлений нейросетей, а также на модификации процедуры обучения. Показано, что существующие методы OOD-детекции обладают аддитивным характером по отношению к классификатору, зависят от эмпирического выбора порогов и наличия репрезентативных данных для калибровки, что ограничивает их надёжность в условиях априорно неизвестных возмущений распределения данных.

Отдельно рассматривается проблема дисбаланса классов и её влияние на устойчивость и корректность решений классификаторов. Проанализированы методы уровня данных и уровня алгоритма, включая недовыборку, перевыборку, синтетическую генерацию примеров и взвешивание классов. Установлено, что данные подходы либо искажают исходное распределение данных, либо вводят дополнительные гиперпараметры, что может приводить к нестабильности обучения и снижению качества обобщения.

В результате проведённого анализа показано, что существующие решения задач оценки неопределённости, обнаружения выхода за распределение и обработки дисбаланса классов носят фрагментарный характер и не образуют единого статистически обоснованного подхода. Большинство методов не содержит встроенного механизма определения границ применимости модели и формализованного отказа от классификации, что приводит к избыточной уверенности моделей и некоректным решениям в условиях ограниченности данных и неопределённости.

Сделан вывод о необходимости разработки интегрированного подхода к построению классификаторов, в котором обработка неопределённости, дисбаланса классов и выхода за носитель распределения является не внешней надстройкой, а внутренним свойством математической модели. Это обстоятельство определяет актуальность и направленность настоящего исследования, ориентированного на создание статистически обоснованных

и доверенных методов классификации в условиях ограниченного объёма данных.

Вторая глава посвящена разработке модифицированного байесовского классификатора, обеспечивающего возможность отказа от принятия решения в условиях неопределённости, а также методам его аппроксимации и интерпретации (положения 1 и 2).

В разделе 2.1 сформулирована постановка задачи бинарной классификации в рамках вероятностной модели. Пусть (X, Y) – случайный вектор с распределением P, где $X \in [0,1]^d$, $Y \in \{\pm 1\}$, а P_X – отвечающее P распределение X. Требуется построить дискриминантную функцию $f : [0,1]^d \to \{\pm 1\}$, минимизирующую вероятность ошибки $\mathbb{P}(Y \neq f(X))$. Оптимальным решением является байесовский классификатор $s(x)$:

$$s(x) = \begin{cases} 1, & \text{если } g(x) > 0 \text{ и } x \in \mathbb{S}, \\ \text{любое из значений } \pm 1, & \text{если } g(x) = 0 \text{ или } x \notin \mathbb{S}, \\ -1, & \text{если } g(x) < 0 \text{ и } x \in \mathbb{S}, \end{cases} \tag{1}$$

где $g(x) = \mathbb{E}(Y|X = x)$, $\mathbb{S}$ – носитель распределения P_X. Вне $\mathbb{S}$ и при $g(x) = 0$ классификатор определён неоднозначно.

Как показано в разделе 2.2, ключевая проблема существующих методов заключается в том, что они формируют решающие правила на всём компакте $[0,1]^d$, включая области вне $\mathbb{S}$, что приводит к необоснованным решениям при сдвигах распределения.

В разделе 2.3 предложена модификация байесовского классификатора, позволяющая преодолеть указанную проблему. Модификация заключается в добавлении к обучающей выборке искусственных наблюдений, компоненты которых равномерно распределены на всём компакте $[0,1]^d$, а метки классов фиксированы и равны нулю. В результате формируется смешанное распределение $P_\alpha = (1-\alpha)P + \alpha\hat{P}$, где $\alpha \in (0,1)$, P – исходное распределение, $\hat{P}$ – распределение фонового класса с равномерной плотностью на $[0,1]^d$ и тождественно нулевой меткой. Маргинальное распределение признаков принимает вид $\lambda_\alpha = (1-\alpha)P_X + \alpha\lambda$, где λ – мера Лебега на $[0,1]^d$. Пусть ρ – неотрицательная интегрируемая функция на $[0,1]^d$ и борелевское множество $A \subseteq \mathbb{S}$ нулевой лебеговой меры такие, что для всех борелевских множеств B в $[0,1]^d$ верно $P_X(B) = \int_B \rho(x)dx + P_X(A \cap B)$.

Теорема 1. *Для всякого $\alpha \in (0,1)$ решение g_α задачи регрессии*

$$\mathbb{E}_\alpha \left(Y - f(X)\right)^2 \to \min_f \tag{2}$$

существует, это решение единственно P_X- и λ-п. н. и может быть задано формулой

$$g_\alpha(x) = \begin{cases} g(x), & \text{если } x \in A, \\ \dfrac{(1-\alpha)g(x)\rho(x)}{\alpha + (1-\alpha)\rho(x)}, & \text{если } \rho(x) > 0 \text{ и } x \in \mathbb{S} \setminus A, \\ 0, & \text{если или } \rho(x) = 0 \text{ и } x \in \mathbb{S} \setminus A, \text{ или } x \notin \mathbb{S}, \end{cases} \tag{3}$$

здесь минимум берется по всем (борелевским) функциям f и

$$g(x) = \mathbb{E}\,(Y|X = x) \;\; \text{на } [0,1]^d.$$

При этом классификатор $s_\alpha = s_\alpha(x)$, $x \in [0,1]^d$, заданный формулой (1) с заменой g на любое решение g_α задачи (2) и $\mathbb{S}$ на $[0,1]^d$, обладает следующими свойствами:

- *(i). s_α реализует минимум в задаче классификации*

$$\mathbb{P}(Y \neq f(X)) \to \min_f,$$

  где минимум берется по всем (борелевским) функциям со значениями ± 1;
- *(ii). зоной неопределённости s_α является множество $\{x \in [0,1]^d : g_\alpha(x) = 0\}$, которое покрывает λ-п.н. множество $[0,1]^d \setminus \mathbb{S}$, где $\mathbb{S}$ – носитель распределения P_X.*

В эмпирической постановке вводится гиперпараметр $\beta > 0$, задающий порог доверия: отказ от классификации происходит при $|f(x)| \leq \beta$, что позволяет управлять шириной зоны неопределённости.

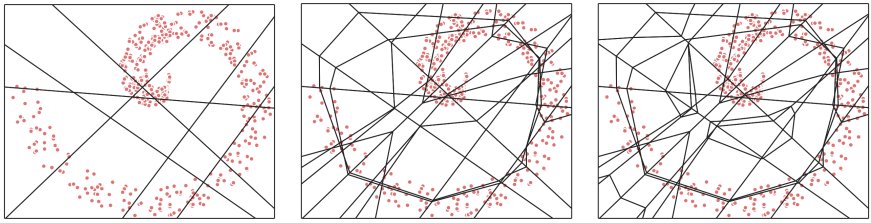
Раздел 2.4 посвящён аппроксимации байесовского классификатора. В разделе 2.4.1 рассмотрены классические методы (гистограммы, kNN, ядерные оценки, SVM) и показаны их ограничения. В разделе 2.4.3 в качестве аппроксиматора предлагается многослойный персептрон c_n с кусочно-линейной функцией активации скрытых слоёв и выходным слоем с одним нейроном с линейной активацией. Для выборки, дополненной n фоновыми наблюдениями, решается задача

$$\sum_{i=1}^{2n} (c_n(X_i) - Y_i)^2 \to \min_{c_n \in C(L,k)},$$

где $C(L,k)$ – множество персептронов с L скрытыми слоями по k нейронов. Персептрон $c_n^*(X)$, являющийся решением задачи, порождает иерархическое (по слоям) разбиение $[0,1]^d$ на $O(k^{dL})$ линейных регионов (ячеек, рис. 1).

В разделе 2.4.4 на основе этого разбиения строится кусочно-постоянная функция гистограммной регрессии, имеющая вид:

$$h_n^*(X) = \frac{n_{+1}(X) - n_{-1}(X)}{n_{-1}(X) + n_0(X) + n_{+1}(X)},$$



а) Ячейки первого слоя: 18 б) Ячейки второго слоя: 121 в) Ячейки выходного слоя: 148

Рис. 1 — Пример разбиения некоторым персептроном с $L = 2$, $k = 6$

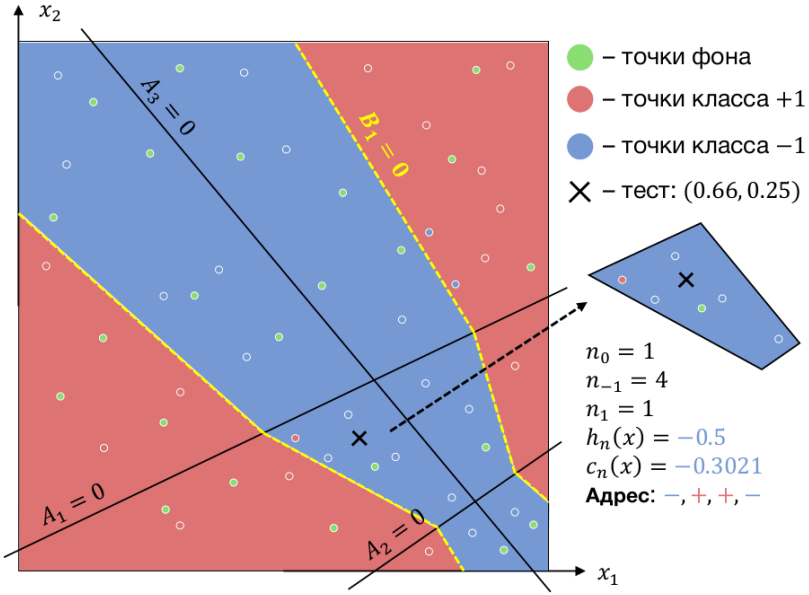
где n_{-1}, n_{+1} и n_0 – количества объектов классов -1, $+1$ и фона в ячейке, содержащей наблюдение X.

В разделе 2.5 разработан метод построения объясняющего двоичного дерева **eXBTree**, которое точно представляет функцию персептрона с кусочно-линейной активацией. Каждый нейрон разбивает входное пространство на две области знаком входной суммы, а последовательное раскрытие функций активаций по слоям формирует дерево, где узлы содержат линейные неравенства $w^\top x + b \geq 0$, а листья – линейные функции выходного слоя (рис. 2).

Теорема 2 (О верхней оценке сложности построения полного объясняющего дерева для многослойного персептрона)**.** *Рассмотрим многослойный персептрон с d-мерным входом, L скрытыми слоями, каждый из которых содержит k нейронов ($k > d$), и одним выходным нейроном. Пусть все скрытые нейроны используют кусочно-линейную функцию активации $|\cdot|$: $\sigma(x) = |x|$. Тогда временная сложность алгоритма построения полного объясняющего двоичного дерева (eXBTree), которое точно представляет функцию, вычисляемую персептроном, в худшем случае составляет $\mathcal{O}(k^{dL})$.*

Теорема 3 (О временной сложности получения прогноза по дереву eXBTree)**.** *Пусть T – полное объясняющее дерево (eXBTree), построенное для многослойного персептрона с L скрытыми слоями по k нейронов и входной размерностью d. Тогда временная сложность получения прогноза для нового наблюдения $x \in \mathbb{R}^d$ по дереву T составляет $\mathcal{O}(d \cdot (kL + 1))$.*

Для практических задач достаточно хранить только уникальные пути, соответствующие реальным наблюдениям, вместо построения дерева целиком. Построенное дерево позволяет интерпретировать результаты классификации, оценивать локальную уверенность, а при недостатке данных – агрегировать информацию с соседними ячейками или подниматься на вышестоящий уровень.



$$A_1 = -0.4x_1 + 0.8x_2 \qquad A_1(x) = -0.064 < 0 \quad \to -$$
$$A_2 = -1.2x_1 + 1.8x_2 + 0.8 \qquad A_2(x) = 0.458 > 0 \quad \to +$$
$$A_3 = -1.5x_1 - 1.3x_2 + 1.5 \qquad A_3(x) = 0.185 > 0 \quad \to +$$

$$B_1 = 0.6|A_1| - 0.5|A_2| + 2.1|A_3| - 0.5 \qquad B_1(x) = -0.3021 < 0 \to -$$

Рис. 2 — Пример eXBTree на основе персептрона с $d = 2$, $L = 1$, $k = 3$

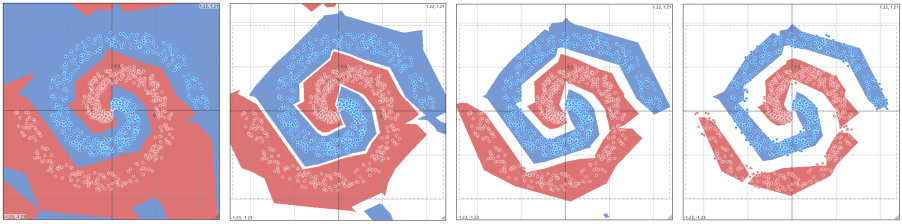
В разделе 2.6 установлена связь между аппроксимациями

$$\mathbb{E}\,(h_n^*(X) - c_n^*(X))^2 \to 0, \quad n \to \infty, \tag{4}$$

обосновывающая замену гистограммной аппроксимации на вычислительно эффективную нейросетевую.

В разделе 2.7 рассмотрен случай многих классов. Показано, что стратегии «один против всех» и попарной классификации обладают существенными недостатками: дисбаланс, квадратичный рост числа моделей, неоднозначность агрегации. Разработанный метод ориентирован на бинарную классификацию.

Раздел 2.8 содержит экспериментальное исследование. На модельных данных установлено, что модифицированный классификатор достигает точности 97–98%, а при исключении отказов – 99–100% при доле отказов 0.5–1%. Показано корректное поведение вне носителя распределения и повышенная устойчивость к backdoor-атакам. Исследовано влияние порога доверия β: увеличение β расширяет область отказа, что позволяет гибко настраивать систему под требования прикладной задачи (рис. 3).



а) $\beta = 0$ б) $\beta = 0.1$ в) $\beta = 0.3$ г) $\beta = 0.5$

Рис. 3 — Влияние порога доверия β на пространственное распределение классификационных решений

В разделе 2.9 исследована устойчивость к состязательным атакам. Предложен метод SLAP (Simple Linear Attack for Perceptron), основанный на решении систем линейных уравнений и квадратичной оптимизации. Эксперименты на датасетах MNIST и CIFAR-10 показали, что для стандартного персептрона атака успешно строит примеры с малым искажением.

Поскольку модифицированный классификатор предназначен для работы с данными малой размерности, проверка устойчивости проводилась на двумерном наборе «две спирали». Для модифицированного классификатора установлено, что без ограничений на входные значения атакующие точки выходят за границы компакта $[0, 1]^d$ и попадают в зону отказа, а с ограничениями допустимое решение в большинстве случаев не существует. Данное поведение свидетельствует о высокой устойчивости доверенного классификатора к SLAP-атаке.

В разделе 2.10 сформулированы выводы. Разработана математическая модель модифицированного байесовского классификатора с гарантированным отказом вне носителя. Предложены методы аппроксимации, доказана их асимптотическая эквивалентность. Разработан метод построения объясняющего дерева eXBTree для интерпретации решений и оценки локальной уверенности. Экспериментально подтверждены эффективность подхода и устойчивость к состязательным атакам.

Третья глава посвящена унарной классификации — подходу, устраняющему критическую зависимость доверенного бинарного классификатора от сбалансированности данных (положения 1 и 3). Дисбаланс классов приводит к смещённым оценкам и снижению надёжности; предлагаемый метод полностью исключает это влияние, фокусируясь только на целевом классе и удаляя данные противостоящего класса из обучения.

В разделе 3.1 вводится постановка задачи нейросетевой регрессии для единственного класса. Пусть имеется выборка $\{X_i\}_{i=1}^{n}$ – наблюдения целевого распределения с неизвестной ограниченной плотностью $f(X)$ на $[0, 1]^d$, которым сопоставляются метки $Y_i = 1$. Дополнительно формируется фоновый набор $\{X_i\}_{i=n+1}^{2n}$ – независимые равномерно распределённые на $[0, 1]^d$

наблюдения с метками $Y_i = 0$. На объединённом сбалансированном наборе мощности $2n$ решается задача минимизации среднеквадратичной ошибки:

$$\sum_{i=1}^{2n} \left(c_n(X_i) - Y_i\right)^2 \to \min_{c_n \in C(L,k)}.$$

Полученная модель $c_n^*(X)$ называется *нейросетевым унарным классификатором*. Решение о принадлежности наблюдения целевому классу принимается при превышении выходом модели порога доверия $\beta \in [0,1)$: $c_n^*(X) > \beta$.

В разделе 3.2 на основе разбиения компакта $[0,1]^d$, порождаемого персептроном, строится кусочно-постоянная функция гистограммной регрессии $h_n^*(X)$. В каждой ячейке K_r оптимальное значение определяется соотношением:

$$h_n^*(X) = \frac{n_1(X)}{n_1(X) + n_0(X)},$$

где $n_1(X)$ и $n_0(X)$ – количества целевых и фоновых наблюдений в ячейке, содержащей точку X. Функция $h_n^*(X)$ представляет собой гистограммную оценку апостериорной вероятности принадлежности к целевому распределению.

В разделе 3.3 сформулирована теорема, обосновывающая статистическую состоятельность предложенного подхода.

Теорема 4 (О сходимости нейросетевой и гистограммной регрессий)**.** *Пусть задана последовательность многослойных персептронов, обученных на выборке из n наблюдений из распределения с ограниченной плотностью на* $[0,1]^d$*, с модульной функцией активации, архитектура которых состоит из первого слоя ширины* r_n*,* L_n *слоёв ширины* k_n *и одноэлементного последнего слоя с линейной активацией, при условии, что весовые коэффициенты инициализируются независимо из непрерывного распределения, а параметры первого слоя заморожены при обучении, если целевая функция является кусочно-гладкой и выполнены ограничения:*

(i) Число ненулевых параметров:

$$S_{\mathrm{nnz},n} \;=\; c_1' \cdot \max\Big\{ n^{\frac{d}{2\beta+d}},\, n^{\frac{d-1}{\alpha+d-1}} \Big\}. \tag{5}$$

(ii) Ограничение на величины параметров:

$$B_n \;\le\; c_1\, n^s. \tag{6}$$

(iii) Количество слоёв после первого:

$$L_n \;\le\; c_1 \Big(1 + \max\big\{ \tfrac{\beta}{d},\, \tfrac{\alpha}{2(d-1)} \big\}\Big). \tag{7}$$

(iv) Число нейронов в первом слое:

$$r_n \;\geq\; 2d. \tag{8}$$

(v) Ограничение на скорость роста архитектуры:

$$k_n^{L_n \min(d,k_n)}\, r_n^d \;=\; o(n), \tag{9}$$

$$k_n^{L_n \min(d,k_n)} r_n^d \, d\,(k_n L_n + r_n)\,\log(k_n L_n + r_n)\,\frac{\log n}{n} \;\to\; 0. \tag{10}$$

(vi) Ограничения на ширину первого слоя (для произвольного $\gamma > 0$):

$$\sum_{n=1}^{\infty} e^{\frac{-c\,\gamma^4 r_n}{d^2}} \;<\; \infty, \tag{11}$$

$$r_n \;\geq\; C\,\gamma^{-12}\,\frac{d^7}{2\pi}. \tag{12}$$

Тогда

$$(c_n(X) - h_n(X)) \xrightarrow{P} 0. \tag{13}$$

В разделе 3.4 рассматривается обобщение на случай многих классов. Для каждого класса $c = 1, \ldots, C$ строится независимый унарный классификатор, обученный выделению носителя данного класса от равномерного фона. В отличие от схемы «один против одного», требующей $\frac{C(C-1)}{2}$ моделей, предложенный подход масштабируется линейно и не нуждается в сложных стратегиях агрегации.

Раздел 3.5 систематизирует преимущества унарной классификации: полная устойчивость к дисбалансу классов, возможность отказа от классификации при недостаточной уверенности, модульность архитектуры, допускающая использование различных конфигураций для разных классов, и естественная векторная интерпретация выходов как апостериорных вероятностей.

В разделе 3.6 предложен оригинальный набор метрик для оценки качества унарных классификаторов. *Мощность* $p^{(i)}$ характеризует долю объектов целевого класса, принимаемых классификатором (превышающих порог β). *Эффективность* $e^{(i)}$ отражает долю объектов, корректно распознанных своим классификатором и отвергнутых чужими. *Мера неразделимости* $g^{(i)}$ оценивает степень пересечения областей, принимаемых разными классификаторами. Для каждой пары классов вводятся интегральные показатели на основе гармонического среднего:

$$P_{12} = \frac{2p^{(1)}p^{(2)}}{p^{(1)} + p^{(2)}}, \quad E_{12} = \frac{2e^{(1)}e^{(2)}}{e^{(1)} + e^{(2)}}, \quad G_{12} = \frac{2g^{(1)}g^{(2)}}{g^{(1)} + g^{(2)}}.$$

На модельных примерах (раздел 3.6.4) продемонстрировано, что предложенные метрики позволяют различать ситуации, в которых классические показатели (accuracy, precision, recall, F_1) дают неразличимые значения.

Раздел 3.7 иллюстрирует работу унарных классификаторов на модельных данных. Для одного класса модель успешно выделяет область высокой плотности. Для двух классов каждый классификатор независимо определяет свою область, итоговая классификация осуществляется по максимуму выхода (рис. 4). Наиболее показателен эксперимент с четырьмя классами при искусственно созданном дисбалансе (соотношения 1:7, 1:5, 1:3): благодаря независимому обучению области принятия решений не искажены дисбалансом, что подтверждает устойчивость метода (рис. 5.

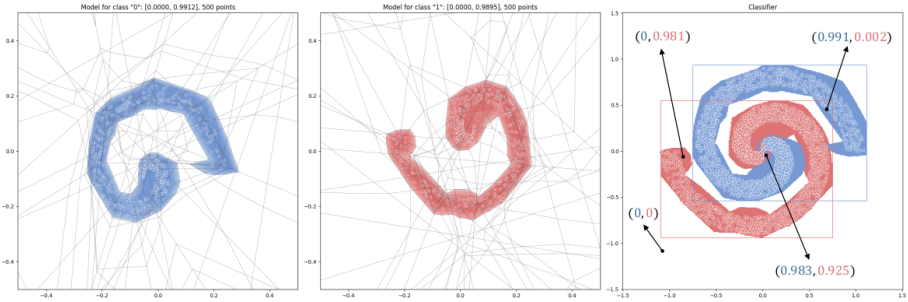


Рис. 4 — Унарная классификация для двух классов

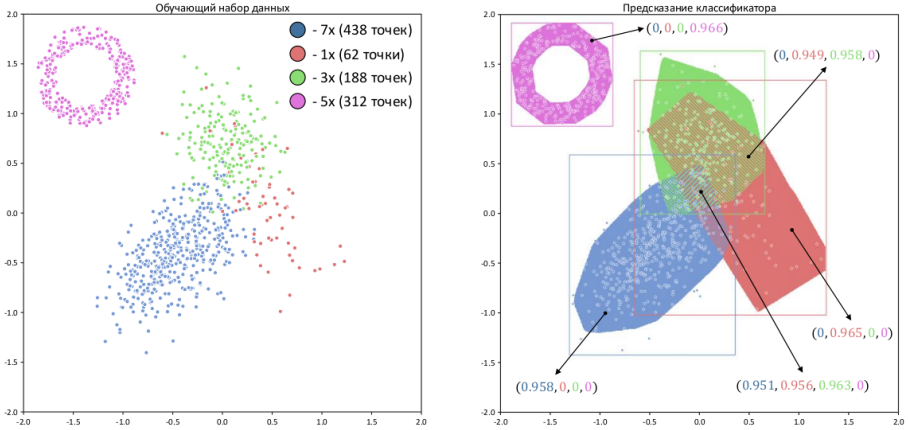


Рис. 5 — Унарная классификация для четырёх классов с дисбалансом

Раздел 3.8 содержит экспериментальное исследование на реальных данных. Для оценки предложенного метода были выбраны наборы из репозитория UCI: Iris ($d = 4$), tic-tac-toe ($d = 9$), liver disease ($d = 10$), wine quality ($d = 11$) и heart disease ($d = 13$). Наборы различаются размерностью и степенью дисбаланса – от сбалансированной выборки Iris до wine quality с соотношением классов 1:2. В качестве базового сравнения использовался XGBoost как эффективный и широко применяемый алгоритм для табличных данных.

Таблица 1 — f_1 мера на реальных наборах данных

Набор данных	d	**Unary**	**XGBoost**
Iris	4	**0.941**	0.933
tic tac toe	9	**0.992**	0.967
Liver disease	10	0.815	**0.824**
Wine quality	11	0.789	**0.797**
Heart disease	13	**0.798**	0.788

Результаты, представленные в таблице 1, показывают, что унарный классификатор демонстрирует конкурентоспособное качество. На сбалансированном наборе Iris с малой размерностью метод превосходит XGBoost. На наборе tic-tac-toe с заметным дисбалансом (1:1.7) и умеренной размерностью ($d = 9$) предложенный алгоритм также показывает более высокое качество. На наборах большей размерности результаты близки, а на heart disease унарный классификатор вновь демонстрирует небольшое преимущество. Это подтверждает как сохранение предсказательной силы на реальных данных, так и ожидаемую устойчивость к дисбалансу.

В разделе 3.9 рассматривается применение унарной классификации для обработки некомплектных данных. Предложен итеративный метод обучения персептрона, в котором на каждой эпохе пропуски временно заполняются случайными значениями, а полученный объект включается в обучающую выборку с вероятностью, равной выходу модели. Метод позволяет обучать классификатор без фиксированного искажения данных, избегая систематического смещения, характерного для классических методов заполнения.

Раздел 3.10 раскрывает связь предложенного подхода с современными архитектурами. Показано, что свёрточный слой может быть представлен как полносвязное линейное преобразование с разреженной матрицей весов, что открывает возможность переноса механизмов оценки носителя и отказа на свёрточные нейросети. Установлена концептуальная близость унарной классификации и обучения дискриминатора в генеративно-состязательных сетях: оба подхода решают задачу выделения эмпирического

распределения данных из шумового. Это указывает на потенциал обобщения предложенного метода на более сложные архитектуры, включая рекуррентные и трансформерные модели.

В разделе 3.11 сформулированы выводы. Разработанный метод унарной классификации полностью устраняет проблему чувствительности к дисбалансу классов, обеспечивая статистическую обоснованность решений в условиях неравномерного распределения объектов. Доказана теорема о состоятельности подхода. Предложены специализированные метрики оценки качества, учитывающие специфику унарной схемы. Экспериментально подтверждена эффективность метода на реальных данных и его устойчивость к дисбалансу. Установлена связь с современными архитектурами, определяющая направления дальнейшего развития.

Четвёртая глава посвящена разработке метода генерации синтетических табличных данных на основе унарной классификации (положение 3). Актуальность задачи обусловлена требованиями доверенного искусственного интеллекта в части обеспечения конфиденциальности обучающих данных. Прямая передача обученных моделей в регулируемых областях (медицина, финансы) сопряжена с рисками обратного восстановления чувствительных выборок по параметрам модели, что делает генерацию синтетических данных, сохраняющих статистические свойства оригиналов при гарантии отсутствия утечек, критически важным направлением.

В разделе 4.1 сформулирована постановка задачи. Пусть имеется выборка $X = \{x_1, \ldots, x_n\} \in [0,1]^d$ из неизвестного распределения. Требуется построить синтетическую выборку $\tilde{X} = \{\tilde{x}_1, \ldots, \tilde{x}_m\}$, сохраняющую геометрические и статистические свойства исходного распределения. В отличие от традиционных генеративных подходов, ориентированных на точное восстановление плотности, предлагаемый метод фокусируется на сохранении геометрической структуры (носителя) данных.

В разделе 4.2 представлен метод создания синтетических (репродукционных) данных, формальное описание которого представлено в алгоритме 1. Процесс состоит из двух этапов. На первом этапе унарно обучается персептрон $c_n(x) : [0,1]^d \to [0,1]$. На втором этапе из равномерного распределения сэмплируется множество точек $\tilde{B}$, и каждая точка $\tilde{b} \in \tilde{B}$ включается в синтетическую выборку с вероятностью, равной выходу обученной модели $c_n(\tilde{b})$.

Раздел 4.3 содержит экспериментальное исследование. На модельных данных (спираль, два квадрата, двумерное и десятимерное нормальные распределения) визуально подтверждена способность метода точно воспроизводить форму, плотность и ковариационную структуру оригиналов. В сравнительном анализе с VAE и GAN на задаче генерации спирали предложенный метод продемонстрировал более высокую устойчивость: VAE размывает глобальную структуру, GAN искажает геометрию и усиливает

Метод 1: Создание синтетических табличных данных

Вход : исходная выборка $X = \{x_1, \ldots, x_n\} \in [0,1]^d$
архитектура модели $(L,\ k)$
мощность синтетической выборки m

Выход : синтетическая выборка $\tilde{X}$

Инициализация: создать случайный персептрон $c_n(X)$

for *epoch* $\leftarrow 1$ **to** E **do**
  Сгенерировать фоновую выборку $B = \{b_1, \ldots, b_n\} \sim U[0,1]^d$
  Обучить $c_n(x)$ на $X \cup B$ с функцией потерь L:
    $L = \sum_{x \in X}(1 - c_n(x))^2 + \sum_{b \in B}(0 - c_n(b))^2$

$\tilde{X} \leftarrow \emptyset$

while $|\tilde{X}| < m$ **do**
  Сгенерировать точку $\tilde{b} \sim U[0,1]^d$
  Сгенерировать $\xi \sim U(0,1)$
  if $\xi < c_n(\tilde{b})$ **then**
    $\tilde{X} \leftarrow \tilde{X} \cup \{\tilde{b}\}$

Вернуть $\tilde{X}$

локальные плотности, тогда как унарный подход сохраняет равномерность распределения вдоль траектории.

Для верификации на реальных данных использованы наборы из репозитория UCI: Iris ($d = 4$), tic-tac-toe ($d = 9$), liver disease ($d = 10$), wine quality ($d = 11$), heart disease ($d = 13$). Оценка проводилась по двум метрикам: *utility* (F_1-мера XGBoost, обученного на синтетических данных и протестированного на реальных) и *fidelity* (усредненная статистика Колмогорова–Смирнова, пакет SDMetrics). Базовым методом сравнения выбран CTGAN – специализированная GAN для табличных данных.

Таблица 2 — Оценка полезности (utility) методов генерации синтетических данных

Набор данных	d	F_1 мера на тестовом наборе		
		train	**Unary**	**CTGAN**
Iris	4	0.933	0.973	0.920
tic tac toe	9	0.967	0.957	0.667
Liver disease	10	0.824	0.731	0.682
Wine quality	11	0.797	0.698	0.718
Heart disease	13	0.788	**0.838**	0.727

Результаты экспериментов показывают, что предложенный метод демонстрирует высокую utility на данных малой и средней размерности ($d \leq$

Таблица 3 — Оценка верности (fidelity) методов генерации синтетических данных

Набор данных	d	**Unary**	**CTGAN**
Iris	4	**0.901**	0.880
tic tac toe	9	0.858	**0.935**
Liver disease	10	0.875	**0.922**
Wine quality	11	0.857	**0.887**
Heart disease	13	0.852	**0.882**

13), превосходя CTGAN на четырех из пяти наборов, включая случай с заметным приростом качества на heart disease (табл. 2). По метрике fidelity метод лидирует на низкоразмерном Iris, однако с ростом размерности CTGAN ожидаемо выигрывает за счет более сложной архитектуры, ориентированной на точное моделирование многомерных плотностей (табл. 3).

В разделе 4.4 сформулированы выводы. Разработанный метод генерации синтетических данных на основе унарной классификации сочетает концептуальную простоту, теоретическую интерпретируемость и способность точно воспроизводить геометрическую структуру низкоразмерных распределений. Определена область эффективного применения — задачи малой и средней размерности, где критично сохранение топологии носителя данных. Выявлено ограничение, связанное с проклятием размерности: при увеличении числа признаков точность статистического воспроизведения снижается, что делает метод менее предпочтительным для высокоразмерных задач.

Пятая глава посвящена разработке интеллектуальной системы машинного обучения для визуализации и исследования методов классификации, обеспечивающей интерактивную среду для экспериментальной проверки теоретических результатов глав 2–4 и анализа доверенных классификаторов в условиях ограниченных вычислительных ресурсов (положение 4).

В разделе 5.1 сформулированы требования к разрабатываемой системе: автономность и кроссплатформенность (работоспособность без GPU и доступа к сети), интерактивная визуализация (обучения, архитектуры, разбиения компакта), динамическая модификация параметров в реальном времени без прерывания обучения, численная корректность (эквивалентность эталонной реализации PyTorch), а также полная реализация разработанных методов – модифицированной бинарной классификации, унарной классификации, генерации синтетических данных и объясняющего дерева eXBTree.

Раздел 5.2 описывает архитектуру системы. Система представляет собой автономное клиентское веб-приложение на JavaScript, не требующее установки или интернет-соединения. Архитектура построена на модульном принципе с применением паттерна Observer (EventEmitter), что обеспечивает дифференциальное обновление интерфейса: при изменении состояния

компонента генерируется событие, и только подписанные на него модули выполняют перерисовку(рис. 6). Это минимизирует вычислительные накладные расходы по сравнению с полной перерисовкой всего интерфейса. Взаимодействие компонентов организовано по принципу публикации-подписки: каскадное распространение событий позволяет обновлять только зависимые визуализации (рис. 7).

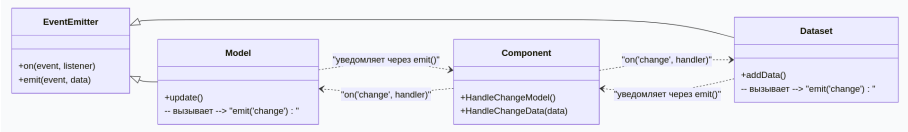


Рис. 6 — Архитектура на основе EventEmitter

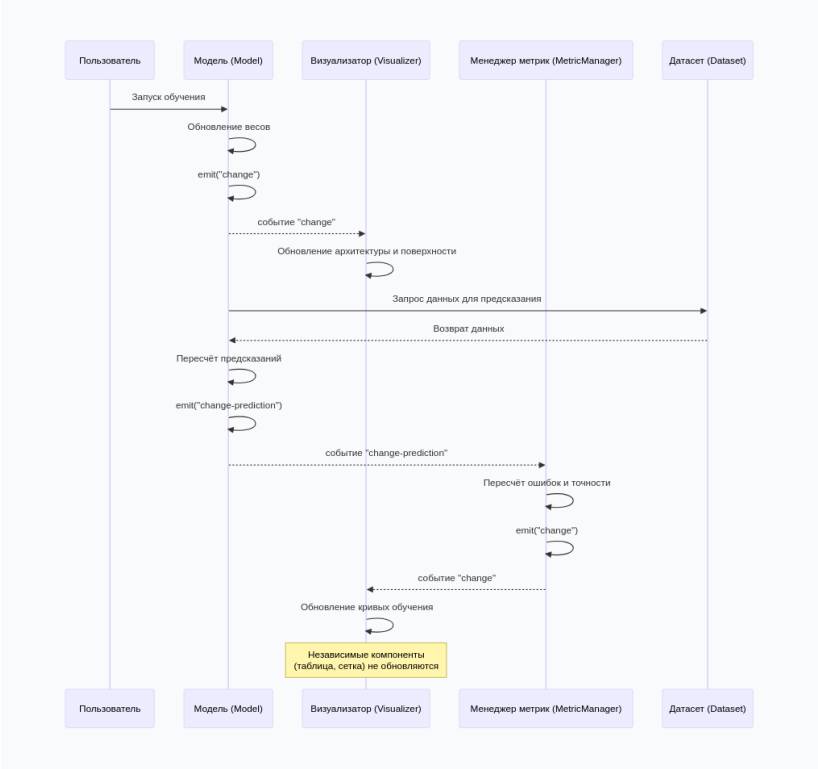


Рис. 7 — Последовательность событий после шага обучения

Раздел 5.3 посвящен реализации вычислительного ядра. Реализованы полносвязный слой с различными функциями активации (ReLU,

LeakyReLU, Abs, линейная), модель нейронной сети с поддержкой динамического изменения архитектуры, функции потерь (MSE, MAE, Huber, LogCosh) и оптимизаторы градиентного спуска (SGD, Momentum, Adam, Adamax, Adadelta, Adagrad, RMSprop) с L1/L2 регуляризацией.

Для минимизации накладных расходов на выделение памяти применены типизированные массивы (`Float64Array`) и преаллокация буферов с учётом максимального размера пакета. Ключевой оптимизацией вычислений в однопоточной среде JavaScript является разворачивание циклов для операций матричного умножения. Это позволяет эффективно использовать параллелизм на уровне инструкций (ILP) современного CPU: независимые инструкции заполняют конвейер процессора, исключая простои, связанные с зависимостями «чтение после записи» (рис. 8).

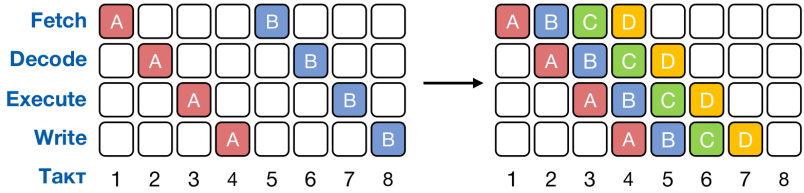


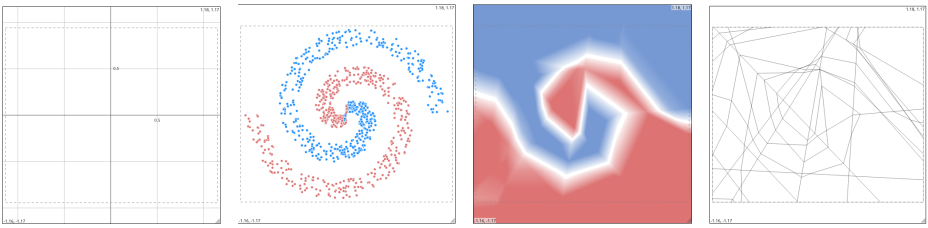
Рис. 8 — Эффект параллелизма на уровне инструкций (ILP)

Оптимизированная версия прямого распространения с разворачиванием на 4 итерации обеспечивает существенный прирост производительности. Коэффициент разворачивания определён эмпирически как оптимальный для целевых архитектур.

Верификация корректности проведена посредством модульного тестирования и сравнительного анализа с эталонной реализацией PyTorch. Для всех функций потерь, оптимизаторов и полносвязных сетей различной глубины максимальная разность результатов не превышала 10^{-15}, что подтверждает численную эквивалентность реализованного ядра.

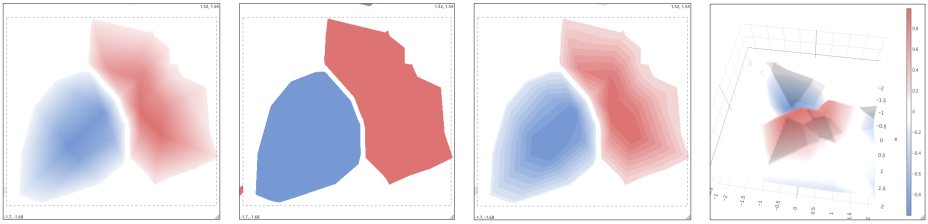
Раздел 5.4 описывает подсистему визуализации. Реализован принцип многослойного рендеринга: слой сетки, слой данных (отрисовка точек через SVG с цветовой кодировкой классов), слой модели (визуализация выхода нейросети через HTML5 Canvas), слой ячеек (границы разбиения компакта) — рис. 9. Все слои синхронизированы через общий объект ViewBox, обеспечивающий преобразование координат.

Для визуализации многомерных данных ($d > 2$) реализован механизм проекций на выбранные координаты с фиксацией остальных измерений. Визуализация выхода модели поддерживает линейный режим, дискретный режим (2, 4, 10 уровней) и 3D-поверхность (рис. 10). Визуализация архитектуры отображает веса связей (цветом и толщиной линии) и состояние нейронов (рис. 11).



а) Слой сетки б) Слой данных в) Слой модели г) Слой ячеек

Рис. 9 — Визуализация слоёв отрисовки



а) Линейный режим б) Дискретный режим (2 уровня) в) Дискретный режим (10 уровней) г) 3D режим

Рис. 10 — Визуализация выхода модели

Раздел 5.5 содержит анализ производительности. Сравнительное тестирование показало, что применение разворачивания циклов обеспечивает ускорение вычислений в 1.1–2.8 раза в зависимости от типа операции, при этом наибольший выигрыш достигнут на наиболее вычислительно интенсивных операциях обратного распространения.

Система полностью кроссплатформенна и функционирует в любом современном браузере под управлением Windows, Linux, macOS, а также на мобильных платформах. Эмпирически установленные границы комфортной работы составляют до 200 нейронов суммарно, до 10^6 точек в 20-мерном пространстве при размере пакета до 128.

Раздел 5.6 демонстрирует примеры использования системы для реализации разработанных методов:

– бинарная классификация двух спиралей с визуализацией разделяющей поверхности (рис. 12);
– унарная классификация с порогом доверия β и зоной отказа (рис. 13);
– генерация синтетических данных на основе обученной унарной модели с контролем статистических характеристик (рис. 14);

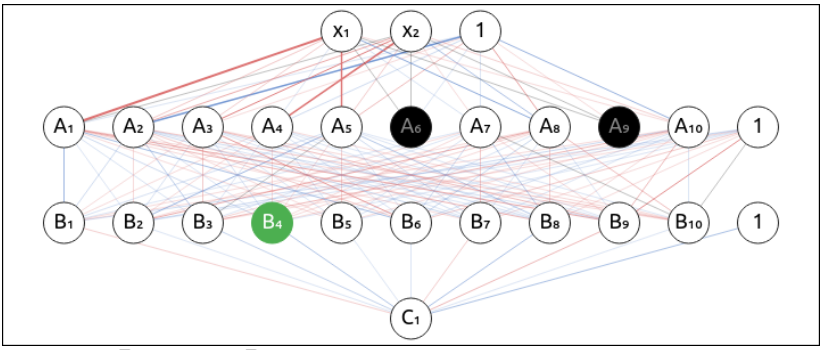


Рис. 11 — Визуализация архитектуры нейросети

– построение объясняющего дерева eXBTree с анализом содержимого ячеек и геометрии пространства (рис. 15).

Рис. 12 — Пример выполнения бинарной классификации

В разделе 5.7 сформулированы выводы. Разработанная интеллектуальная система полностью удовлетворяет всем сформулированным требованиям. Она реализует полный цикл исследования доверенных классификаторов: от создания данных и обучения модели до интерпретации решений и генерации синтетических выборок. Обеспечена автономность, кроссплатформенность, численная корректность (эквивалентность PyTorch в пределах машинной точности) и интерактивная визуализация в реальном времени. Применённая оптимизация разворачивания циклов позволила достичь ускорения вычислений до 2.8 раз в однопоточной среде

Рис. 13 — Пример выполнения унарной классификации

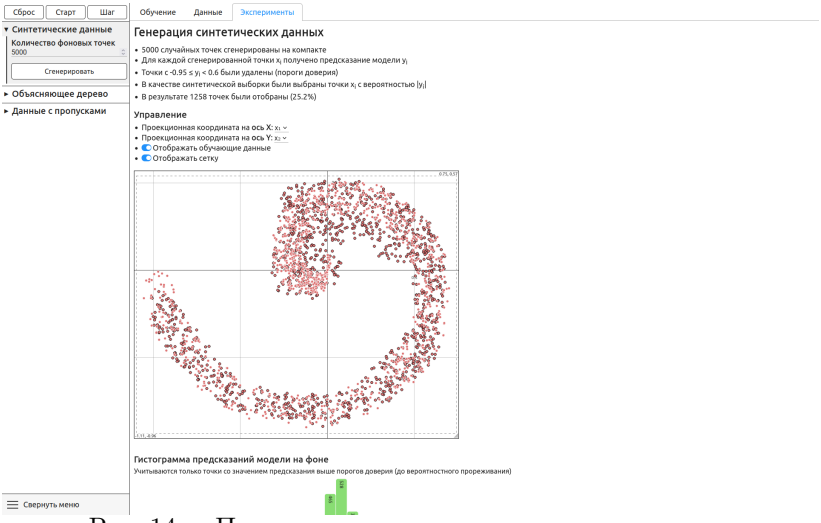


Рис. 14 — Пример построения синтетических данных

JavaScript. Система опубликована как открытый программный продукт на платформе GitHub и служит инструментом для экспериментального подтверждения теоретических результатов диссертации, а также для интерактивного анализа и отладки доверенных персептронных моделей.

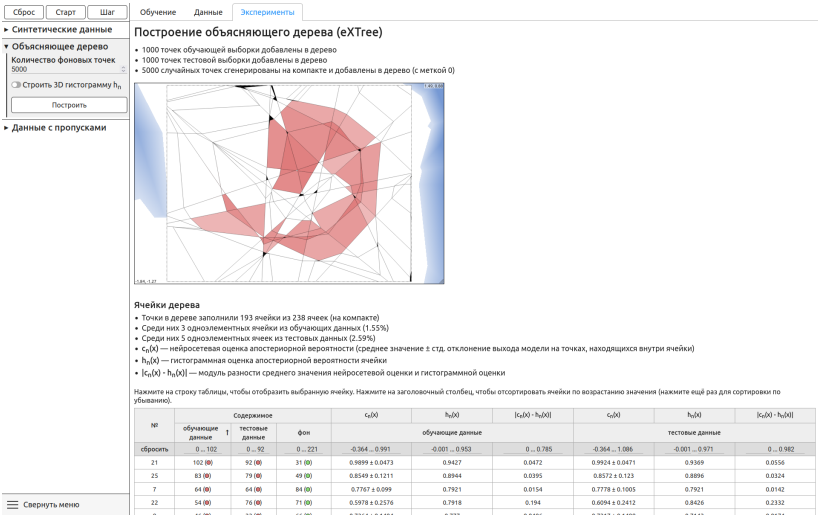


Рис. 15 — Пример работы с объясняющим деревом eXBTree

В **заключении** приведены основные результаты работы, которые заключаются в следующем:

В представленной работе разработаны теоретические основы и методы построения доверенных классификаторов на основе многослойного персептрона для данных малой размерности, обеспечивающих способность к отказу от классификации вне носителя распределения, устойчивость к дисбалансу классов и интерпретируемость принимаемых решений, что позволило сократить разрыв между эмпирическими нейросетевыми методами и формальным аппаратом математической статистики.

Основные результаты работы заключаются в следующем:

1. Разработана теоретическая база непараметрического оценивания в условиях дисбаланса классов и малой размерности. Сформулированы и доказаны теоремы, обосновывающие асимптотическую связь между нейросетевой и гистограммной оценками апостериорной вероятности. Данный результат обеспечивает формальное статистическое обоснование для предложенного подхода.
2. Разработан метод построения статистически обоснованного объяснимого байесовского классификатора на основе многослойного персептрона и дерева решений. Метод обеспечивает оценку апостериорных вероятностей с теоретическими гарантиями и формирование интерпретируемых правил классификации.
3. Разработан метод построения унарного классификатора, устойчивого к дисбалансу классов и позволяющего генерировать синтетические данные, сохраняющие геометрические и статистические свойства исходного распределения.

4. Создана интеллектуальная система машинного обучения, реализующая предложенные методы и обеспечивающая решение задач классификации данных малой размерности в условиях дисбаланса классов и высокой неопределённости вне носителя распределения.

Результаты экспериментального исследования подтвердили устойчивость классификатора к дисбалансу классов, корректность работы в условиях выхода за носитель распределения и адекватность генерируемых синтетических данных.

В работе предложен формально обоснованный подход к созданию доверенных классификаторов, сочетающий выразительность нейросетевых моделей со строгостью статистических методов.

Публикации автора по теме диссертации

1. *Perminov, A.* SLAP—Simple Linear Attack against Perceptron (SLAP) [Текст] / A. Perminov // Programming and Computer Software. — 2025. — Т. 51, № 6. — С. 446—452.
2. Extrapolation of the Bayesian classifier with an unknown support of the two-class mixture distribution [Текст] / K. S. Lukianov [и др.] // Uspekhi Matematicheskikh Nauk. — 2024. — Т. 79, № 6. — С. 57—82.
3. *Eliseev, N. A.* Convergence of a multilayer perceptron to histogram Bayesian regression [Текст] / N. A. Eliseev, A. I. Perminov, D. Y. Turdakov // Uspekhi Matematicheskikh Nauk. — 2025. — Т. 80, № 6. — С. 45—72.
4. *Perminov, A.* CONSISTENT METHOD FOR SYNTHETIC TABULAR DATA OBTAINING USING A MULTILAYER PERCEPTRON [Текст] / A. Perminov, A. Kovalenko, D. Turdakov // Journal of Mathematical Sciences. — 2026. — С. 1—12.
5. *Belyaeva, O. V.* Synthetic data usage for document segmentation models fine-tuning [Текст] / O. V. Belyaeva, A. I. Perminov, I. S. Kozlov // Proceedings of the Institute for System Programming of the RAS (Proceedings of ISP RAS). — 2020. — Т. 32, № 4. — С. 189—202.
6. *Perminov, A. I.* Method for training perceptron on tabular data with missing values [Текст] / A. I. Perminov, A. P. Kovalenko, D. Y. Turdakov // Proceedings of the Institute for System Programming of the RAS. — 2025. — Т. 37, № 62. — С. 93—106.

7. *Свидетельство о гос. регистрации программы для ЭВМ.* DenseNetworkVisualizer: программное обеспечение для геометрической и вероятностной интерпретации и визуализации многослойного персептрона [Текст] / А. И. Перминов [и др.] ; Ф. государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова Российской академии наук. — № 2023689161 ; заявл. 26.12.2023 ; опубл. 26.12.2023, 2023689161 (Рос. Федерация).

8. *Свидетельство о гос. регистрации программы для ЭВМ.* Программа реализации атаки уклонения в отношении модели обнаружения вторжений [Текст] / М. И. Булгакова [и др.] ; Ф. государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова Российской академии наук. — № 2022682843 ; заявл. 28.11.2022 ; опубл. 28.11.2022, 2022682843 (Рос. Федерация).

9. *Свидетельство о гос. регистрации программы для ЭВМ.* Программа защиты от атаки уклонения в системе обнаружения вторжений [Текст] / М. И. Булгакова [и др.] ; Ф. государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова Российской академии наук. — № 2022685576 ; заявл. 26.12.2022 ; опубл. 26.12.2022, 2022685576 (Рос. Федерация).

10. *Свидетельство о гос. регистрации программы для ЭВМ.* Программное обеспечение для выявления и устранения предвзятости моделей машинного обучения [Текст] / И. С. Алексеевская [и др.] ; Ф. государственное бюджетное учреждение науки Институт системного программирования им. В.П. Иванникова Российской академии наук. — № 2024692147 ; заявл. 12.12.2024 ; опубл. 12.12.2024, 2024692147 (Рос. Федерация).

11. *Коваленко, А. П.* Подход к решению «проблемы экстраполяции» нейросетевого классификатора [Текст] / А. П. Коваленко, А. И. Перминов // Материалы 32-й научно-технической конференции «Методы и технические средства обеспечения безопасности информации». — 2023.

12. *Коваленко, А. П.* Доверять... или не доверять? Лемма об экстраполяции байесовского классификатора [Текст] / А. П. Коваленко, А. И. Перминов, П. А. Яськов // Материалы 33-й научно-технической конференции «Методы и технические средства обеспечения безопасности информации». — 2024.

13. *Коваленко, А. П.* Метод унарной классификации [Текст] / А. П. Коваленко, А. И. Перминов // Материалы 34-й научно-технической конференции «Методы и технические средства обеспечения безопасности информации». — 2025.

Перминов Андрей Игоревич

Доверенный байесовский классификатор для данных малой размерности на основе многослойного персептрона

Автореф. дис. на соискание учёной степени канд. физ.-мат. наук

Подписано в печать ___.___.____. Заказ № _____
Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.
Типография _____