

УДК: 519.8

Научный руководитель: Албогачиева Мадина Магометовна

Физико-математический факультет

ФГБОУ ВО «Ингушский государственный университет»

Янарсанова Тамара А-Р., Чаниева Фатима Б.

студентки 3-го курса специальности, «Математика»,

физико-математического факультета

**ЗАДАЧИ СТАТИСТИЧЕСКОЙ ТЕОРИИ И МАШИННОГО
ОБУЧЕНИЯ: МЕТОДОЛОГИЧЕСКИЕ ОСНОВАНИЯ,
СОВРЕМЕННЫЕ ПРИМЕНЕНИЯ И ИНТЕРПРЕТАЦИЯ НА
МАТЕРИАЛЕ ХУДОЖЕСТВЕННЫХ ОБРАЗОВ**

***Аннотация:** Статья посвящена анализу задач статистической теории и машинного обучения как взаимосвязанных направлений современной науки о данных. Рассматривается вопрос о том, в какой мере машинное обучение продолжает традиции математической статистики и в чём проявляется специфика его алгоритмического подхода. Показано, что статистическая теория формирует фундамент для работы с неопределённостью, вероятностными моделями, выборочными данными и оценкой ошибок, тогда как машинное обучение развивает эти идеи в условиях больших массивов информации, высокой размерности признакового пространства и необходимости автоматизированного принятия решений. В статье особое внимание уделено проблемам классификации, регрессии, кластеризации, переобучения, обобщающей способности модели и интерпретируемости результатов. Научная значимость исследования состоит в систематизации ключевых задач статистической теории и машинного обучения в едином методологическом поле, а также в выявлении их значения для медицины, экономики, гуманитарных исследований и цифровых технологий. Для расширения междисциплинарного контекста приведены примеры из художественной литературы, позволяющие показать, что вероятностное мышление, анализ признаков и вывод на основе неполной информации имеют не только техническую, но и культурную природу. В работе предложены*

также три рисунка, иллюстрирующие соотношение статистики и машинного обучения.

Ключевые слова: *статистическая теория, машинное обучение, анализ данных, классификация, регрессия, переобучение, интерпретируемость.*

Scientific adviser: Albogachieva Madina Magometovna

Faculty of Physics and Mathematics

FGBOU VO "Ingush State University"

Yanarsanova Tamara A-R., Chanieva Fatima B.

3rd year students of the specialty "Mathematics",

Faculty of Physics and Mathematics

Tasks of Statistical Theory and Machine Learning: Methodological Foundations, Modern Applications and Interpretation through Literary Images

Abstract: *The article analyzes the tasks of statistical theory and machine learning as interconnected areas within contemporary data science. It examines the extent to which machine learning continues the traditions of mathematical statistics and identifies the specific features of its algorithmic approach. It is shown that statistical theory provides the foundation for dealing with uncertainty, probabilistic models, sample data, and error estimation, while machine learning develops these ideas under conditions of large-scale data, high-dimensional feature spaces, and the need for automated decision-making. Special attention is paid to the problems of classification, regression, clustering, overfitting, model generalization, and interpretability of results. The scientific significance of the study lies in the systematization of the key tasks of statistical theory and machine learning within a unified methodological framework, as well as in revealing their importance for medicine, economics, the humanities, and digital technologies. To broaden the interdisciplinary perspective, the article includes examples from fiction, demonstrating that probabilistic thinking, feature analysis, and inference based on incomplete information have not only a technical but also a cultural dimension. The paper also proposes three figures illustrating the relationship between statistics and machine learning and the internal logic of model construction and evaluation.*

Keywords: *statistical theory, machine learning, data analysis, classification, regression, overfitting, interpretability.*

Введение

Современное общество всё в большей степени организовано вокруг данных. Научные исследования, управление сложными системами, экономическое прогнозирование, медицинская диагностика и цифровые сервисы опираются на методы, позволяющие извлекать закономерности из наблюдений, отделять значимую информацию от шума и строить обоснованные предсказания. В этом контексте статистическая теория и машинное обучение оказываются не просто смежными дисциплинами, а двумя формами единого исследовательского поля, сосредоточенного на анализе неопределённости и получении знания на основе эмпирического материала.

Статистическая теория исторически сложилась как наука о закономерностях массовых явлений и методах вывода по выборочным данным. Машинное обучение возникло значительно позже, но опиралось на многие идеи статистики, переработав их в условиях вычислительной автоматизации, высокой размерности данных и необходимости адаптивного построения моделей. Несмотря на распространённое мнение о противопоставлении статистики и машинного обучения, в современной научной литературе всё чаще подчёркивается их глубокая взаимосвязь: статистика обеспечивает строгость вывода и язык неопределённости, а машинное обучение расширяет инструментальные возможности анализа данных [1], [2].

Актуальность темы определяется тем, что в условиях цифровой трансформации практически каждая прикладная область сталкивается с задачей выбора, интерпретации и оценки моделей. При этом научный интерес представляет не только техническая сторона вопроса, но и методологическая: каким образом статистическая теория и машинное обучение соотносятся как способы познания, где проходит граница между объяснением и предсказанием и почему вопрос об ошибке является центральным для обеих областей.

Цель исследования

Целью настоящей статьи является выявление и осмысление основных задач статистической теории и машинного обучения, установление их методологического единства и различий, а также раскрытие научной значимости данных направлений для современной исследовательской и прикладной практики. Дополнительная цель заключается в том, чтобы

показать на примерах из художественной литературы, что логика статистического вывода и машинного обучения соотносится с более широкими культурными моделями понимания мира, основанными на интерпретации признаков, вероятностной оценке и принятии решения при неполной информации.

Материал и методы исследования

Материалом исследования послужили труды отечественных и зарубежных учёных по математической статистике, теории вероятностей, статистическому обучению, анализу данных и машинному обучению, а также художественные произведения, в которых присутствуют модели рассуждения, сходные с логикой предсказания и вероятностного вывода. В качестве теоретической базы использовались работы В. Н. Вапника, Т. Хасты, Р. Тибширани, Дж. Фридмана, К. Бишопа, И. Гудфеллоу, К. Мёрфи и других исследователей, посвящённые статистической теории обучения, распознаванию образов, вероятностным моделям и обобщающей способности алгоритмов [3], [4].

В работе применялись сравнительно-аналитический и интерпретативный методы. Сравнительно-аналитический метод позволил сопоставить задачи статистической теории и машинного обучения в историческом, содержательном и прикладном аспектах. Интерпретативный метод использовался для включения в анализ примеров из художественной литературы, где можно обнаружить структуры, аналогичные задачам классификации, прогнозирования и оценки признаков. Кроме того, был применён метод концептуального синтеза, позволивший объединить математико-статистическую и гуманитарную перспективы.

Следует подчеркнуть, что статья не ставит целью исчерпывающее описание всех алгоритмов и теорем. В центре внимания находятся фундаментальные задачи, определяющие логику развития обеих областей, и их значение для современной науки.

Основная часть

1. Статистическая теория как фундамент вывода в условиях неопределённости

Статистическая теория формируется вокруг базовой эпистемологической проблемы: как по конечной выборке судить о более широкой реальности, которая не дана исследователю непосредственно. Иначе говоря, статистика занимается переходом от наблюдаемого к ненаблюдаемому. В этом переходе всегда присутствует неопределённость, поэтому центральной задачей

становится не просто вычисление величин, а оценка надёжности получаемых выводов.

Классические задачи статистической теории — оценивание параметров, проверка статистических гипотез, построение доверительных интервалов, исследование структуры распределений, анализ зависимостей между переменными — объединяются общей логикой: любая модель должна быть не только удобной, но и обоснованной. Если по выборке строится некоторое представление о генеральной совокупности, то необходимо понимать меру ошибки, вероятность отклонения и границы применимости такого представления. Именно поэтому статистика не сводится к набору вычислительных процедур; она выступает как теория доказуемого вывода при неполноте данных.

Важнейшая задача статистической теории состоит в различении закономерности и случайности. На практике это означает необходимость отделить устойчивый сигнал от случайных флуктуаций, которые возникают в любой эмпирической информации. Если исследователь игнорирует эту проблему, он рискует принять случайное совпадение за существенную связь. Данная опасность особенно возрастает в условиях большого числа признаков, когда вероятность ложных зависимостей существенно увеличивается. Следовательно, статистическая теория играет роль методологического фильтра, препятствующего произвольной интерпретации данных.

2. Машинное обучение как развитие статистического мышления

Машинное обучение часто описывается как область, разрабатывающая алгоритмы, способные обучаться по данным. Однако за этим определением стоит более глубокая идея: алгоритм должен построить правило, которое будет работать не только на известной выборке, но и на новых объектах. Это требование делает машинное обучение прямым наследником статистической теории, поскольку в обоих случаях ключевым вопросом является обобщение.

Разница состоит в том, что машинное обучение возникло в среде интенсивных вычислений, больших данных и прикладных задач, где важно не только объяснить закономерность, но и использовать её в автоматизированной форме. Если классическая статистика в большей степени ориентировалась на интерпретацию параметров и проверку гипотез, то машинное обучение исторически сместило акцент на предсказательную эффективность. Но это смещение не означает разрыва с традицией статистики. Напротив, многие алгоритмы машинного обучения — от линейной и логистической регрессии до

байесовских методов и методов опорных векторов — строятся на статистических принципах [5].

Центральные задачи машинного обучения — классификация, регрессия, кластеризация, сокращение размерности, выявление аномалий — представляют собой различные формы поиска структуры в данных. Классификация направлена на отнесение объекта к одному из заранее заданных классов; регрессия описывает зависимость непрерывной величины от набора признаков; кластеризация ищет внутреннее разделение данных без предварительной разметки; снижение размерности позволяет выделить наиболее информативные компоненты сложного пространства. Во всех этих случаях решается единая проблема: как по ограниченному эмпирическому материалу построить модель, которая не будет механически повторять обучающие данные, а выявит скрытую закономерность.

3. Проблема обобщающей способности и переобучения

Одной из важнейших общих задач статистической теории и машинного обучения является обеспечение обобщающей способности модели. Модель считается полезной лишь тогда, когда она способна давать удовлетворительные результаты за пределами того набора данных, на котором была построена. Если же модель слишком точно воспроизводит детали обучающей выборки, включая случайный шум, возникает переобучение. Такая модель демонстрирует высокое качество на уже известных данных, но резко теряет эффективность на новых наблюдениях.

Проблема переобучения показывает, что точность на обучающей выборке не может служить единственным критерием научной состоятельности модели. В статистике эта идея давно была выражена в требованиях к состоятельности оценок, в анализе остатков и в проверке адекватности моделей. В машинном обучении она принимает форму разделения данных на обучающие, валидационные и тестовые множества, использования перекрёстной проверки, регуляризации и контроля сложности модели. Тем самым машинное обучение фактически развивает статистическую интуицию в алгоритмической форме [2].

Данная проблема имеет не только техническое, но и философское значение. Она показывает, что знание не тождественно запоминанию. Модель, которая «выучила» прошлое слишком буквально, не обязательно поняла структуру явления. В этом смысле переобучение можно сравнить с ошибкой читателя художественного текста, который принимает второстепенную деталь за ключевой символ и строит на ней чрезмерно сложную интерпретацию.

Подлинное понимание требует умения различать существенное и случайное — именно это и составляет ядро статистического и машинного обучения.

4. Интерпретируемость и объяснимость моделей

Одной из центральных задач современной науки о данных становится интерпретируемость результатов. В ранних этапах развития машинного обучения считалось, что высокая точность прогноза может компенсировать недостаток объяснимости. Однако по мере внедрения алгоритмов в медицину, судебную практику, банковскую сферу и государственное управление стало очевидно, что одного лишь корректного ответа недостаточно. Необходимо понимать, на основании каких признаков принято решение, насколько оно устойчиво и не содержит ли скрытой дискриминации.

Статистическая теория традиционно придавала интерпретации первостепенное значение. Коэффициенты модели, доверительные интервалы, уровни значимости и характеристики ошибки позволяли исследователю не только получить результат, но и встроить его в причинное или функциональное объяснение. В машинном обучении, особенно в глубоких нейронных сетях, эта прозрачность часто утрачивается. Поэтому одной из ключевых задач становится разработка методов объяснимого искусственного интеллекта, которые помогают вскрыть вклад признаков, локальные механизмы решения и области неустойчивости модели.

Научная значимость этой проблемы чрезвычайно велика. Если статистика и машинное обучение претендуют на участие в принятии социально значимых решений, они должны не только обеспечивать эффективность, но и допускать критику, проверку и воспроизводимость. Иначе модель превращается в технологически сложный, но *epistemologically weak* инструмент, которому доверяют без достаточных оснований.

5. Примеры из художественной литературы как форма интерпретации статистического мышления

В рассказах и романах Артура Конан Дойла о Шерлоке Холмсе герой действует как исследователь, работающий с признаками и гипотезами. Он собирает наблюдения, отсеивает нерелевантные детали, сопоставляет возможные версии и выбирает ту, которая лучше объясняет совокупность данных. По существу Холмс решает задачу классификации и вероятностного вывода: он определяет, к какому классу причин относится наблюдаемое событие. Его метод напоминает аналитический процесс в статистике и машинном обучении,

где каждая деталь выступает как признак, а каждая версия — как гипотеза о скрытом механизме.

В романе Л. Н. Толстого «Анна Каренина» Константин Левин пытается понять закономерности сельскохозяйственного труда и поведения людей. Его размышления нередко терпят неудачу именно потому, что он стремится применить упрощённую схему к более сложной реальности. В терминах машинного обучения это можно рассматривать как ситуацию недостаточной модели, которая не улавливает структуру данных. Вместе с тем его постоянная коррекция взглядов по мере накопления наблюдений напоминает итеративное обновление модели на новой информации.

Художественные примеры демонстрируют, что любая модель мира строится на ограниченном наборе наблюдений и всегда связана с риском ошибки. Они также напоминают о том, что анализ данных не является чисто машинным процессом: он включён в более широкий контекст понимания, интерпретации и ответственности.

Обсуждение результатов исследования

Проведённый анализ позволяет сделать вывод о том, что статистическая теория и машинное обучение не образуют конкурирующих систем знания. Напротив, они находятся в отношениях преемственности и взаимного дополнения. Статистическая теория задаёт концептуальные основания анализа данных: она формирует представления о вероятности, ошибке, достоверности и параметрической интерпретации. Машинное обучение, в свою очередь, усиливает эти основания за счёт алгоритмической гибкости, способности работать с большим числом признаков и адаптации к сложным структурам данных.

Особенно значимым является то обстоятельство, что ключевые задачи обеих областей совпадают на фундаментальном уровне. И статистика, и машинное обучение стремятся обнаружить устойчивую закономерность, оценить степень неопределённости, предотвратить ложный вывод, обеспечить переносимость результата на новые данные и, по возможности, объяснить механизм принятого решения. Различия между ними связаны главным образом с исследовательскими акцентами и инструментальными средствами, а не с принципиально разной природой научного поиска.

Использование литературных примеров позволило показать, что проблемы классификации, выбора признаков, неопределённости и ошибки имеют

глубокий культурный смысл. Это особенно важно для гуманитарного осмысления цифровой эпохи, в которой алгоритмы всё активнее участвуют в интерпретации мира. Научная ценность такого подхода заключается в возможности соединить строгую методологию с более широким пониманием познавательной деятельности человека.

Заключение

Исследование показало, что задачи статистической теории и машинного обучения образуют целостное и методологически связанное пространство современной науки о данных. Статистическая теория обеспечивает основу для анализа неопределённости, проверки гипотез, оценки параметров и контроля ошибок. Машинное обучение развивает эти идеи в направлении автоматизированного извлечения закономерностей, прогнозирования и адаптации моделей к многомерным и быстро меняющимся данным.

К числу наиболее значимых задач относятся построение моделей, различение сигнала и шума, обеспечение обобщающей способности, предотвращение переобучения, достижение интерпретируемости и согласование объяснения с предсказанием. Эти задачи имеют не только прикладное, но и фундаментальное значение, поскольку они определяют способы получения знания в условиях неполноты информации.

Научная значимость темы состоит в том, что статистическая теория и машинное обучение становятся универсальным языком современной исследовательской культуры. Они применяются в естественных, технических, социальных и гуманитарных науках, формируя новые стандарты доказательности, воспроизводимости и аналитической ответственности. Обращение к художественной литературе позволило дополнительно показать, что вероятностное мышление и вывод по признакам укоренены в более широком человеческом опыте понимания мира. Следовательно, дальнейшее развитие статистической теории и машинного обучения должно быть связано не только с ростом вычислительной мощности, но и с углублением интерпретируемости, методологической строгости и гуманитарной рефлексии.

Список литературы

Вапник, В. Н. Статистическая теория обучения / В. Н. Вапник. — М. : Наука, 1999. — 736 с.

Хасты, Т. Основы статистического обучения: интеллектуальный анализ данных, логический вывод и прогнозирование / Т. Хасты, Р. Тибширани, Дж. Фридман ; пер. с англ. — М. : Вильямс, 2016. — 768 с.

Бишоп, К. М. Распознавание образов и машинное обучение / К. М. Бишоп ; пер. с англ. — М. : Диалектика, 2020. — 738 с.

Гудфеллоу, И. Глубокое обучение / И. Гудфеллоу, Й. Бенджио, А. Курвилль ; пер. с англ. — М. : ДМК Пресс, 2018. — 652 с.

Murphy, K. P. Machine Learning: A Probabilistic Perspective / K. P. Murphy. — Cambridge : MIT Press, 2012. — 1104 p.

Кендалл, М. Дж. Теория статистических выводов и связей / М. Дж. Кендалл, А. Стьюарт ; пер. с англ. — М. : Наука, 1973. — 899 с.

Айвазян, С. А. Прикладная статистика и основы эконометрики : учеб. для вузов / С. А. Айвазян, В. С. Мхитарян. — М. : ЮНИТИ-ДАНА, 2001. — 1022 с.

Дойл, А. К. Записки о Шерлоке Холмсе / А. К. Дойл. — СПб. : Азбука, 2022. — 416 с.

Толстой, Л. Н. Анна Каренина / Л. Н. Толстой. — М. : Эксмо, 2021. — 864 с.

Достоевский, Ф. М. Преступление и наказание / Ф. М. Достоевский. — М. : АСТ, 2023. — 672 с.