МОДЕЛИ МАШИННОГО ОБУЧЕНИЯ КАК ЧАСТЬ ТЕОРЕТИЧЕСКИХ ОСНОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Муртазаева Умида Исакуловна,
Худойназарова Элмира,
Турсунова Сарвиноз,
Муродова Нозима
Самаркандский филиал Ташкентского университета информационных технологий имени Мухаммада ал-Хорезми
***e-mail**: murtazayeva1982@yandex.ru*

Аннотация. В статье рассматриваются модели машинного обучения (МО) как ключевой компонент теоретических основ искусственного интеллекта (ИИ). Показано, что математические принципы МО-статистическая аппроксимация, теория обобщающей способности (обобщающая ошибка, смещение–дисперсия, VC- размерность, PAC-границы), регуляризация, байесовский вывод, оптимизация-образуют фундамент для проектирования объяснимых и верифицируемых ИИ-систем.

Ключевые слова: машинное обучение, искусственный интеллект, теоретико-статистическая аппроксимация, обобщающая способность, байесовский вывод, ансамбли, глубокие нейросети, вероятностные модели.

Annotatsiya. Maqolada sun’iy intellekt (SI)ning nazariy asoslarining muhim tarkibiy qismi sifatida mashinaviy o‘qitish (MO) modellari ko‘rib chiqiladi. Unda MOning matematik tamoyillari - statistik aproksimatsiya, umumlashtirish qobiliyati nazariyasi (umumlashtirish xatosi, og‘ish–dispersiya muvozanati, VC o‘lchami, PAC-chegaralar), regulyarizatsiya, Bayescha xulosa chiqarish, optimallashtirish-tushuntiriladigan va verifikatsiya qilinadigan SI tizimlarini loyihalash uchun poydevor bo‘lishi ko‘rsatiladi.

Kalit so‘zlar: mashinaviy o‘qitish, sun’iy intellekt, teoretik-statistik yaqinlashtirish, umumlashtirish qobiliyati, Bayescha xulosa chiqarish, ansambllar, chuqur neyron tarmoqlar, ehtimollik modellar.

Abstract. The article examines machine learning (ML) models as a key component of the theoretical foundations of artificial intelligence (AI). It shows that the mathematical principles of ML-statistical approximation, the theory of generalization (generalization error, the bias–variance trade-off, VC dimension, PAC bounds), regularization, Bayesian inference, and optimization-form the foundation for designing explainable and verifiable AI systems.

Keywords: machine learning, artificial intelligence, statistical approximation theory, generalization ability, Bayesian inference, ensembles, deep neural networks, probabilistic models.

История искусственного интеллекта - это история диалога двух парадигм: символической (логики, правила, онтологии, планирование) и статистической (машинное обучение, вероятностные модели, оценивание). За последние десятилетия статистическая парадигма укрепила свою роль благодаря доступности больших данных и вычислений, однако теоретические основы по-прежнему определяют пределы и возможности моделей [6;14]. Когда мы говорим «модель машинного обучения», мы подразумеваем не только архитектуру и алгоритм обучения, но и совокупность предпосылок: гипотезы о распределении данных, гладкость искомой зависимости, ограничение сложности классов функций, априорные предположения (в байесовском смысле) и вычислительные допущения (оптимизационные схемы, стохастичность, приближённость) [10; 15].

Машинное обучение формализуется как задача минимизации риска: эмпирического (на обучающей выборке) и истинного (на распределении данных). Типовая постановка: выбрать функцию $f \in F$ из семейства гипотез, минимизирующую ожидаемую потерю $\mathrm{R(f)} = \mathrm{E}_{(\mathrm{x,y})\sim \mathrm{D}}[\ell(f(x), y)]$. На

практике мы минимизируем эмпирический риск $\hat{R}_n(f)$ и добавляем штраф за сложность (регуляризацию), ограничивая «ёмкость» класса F [9;10;12]. Теоретические границы обобщения (PAC-границы, оценки через VC-размерность или Rademacher-сложность) связывают разницу $R(f) - \hat{R}_n(f)$ с размером выборки и сложностью класса гипотез: чем богаче класс, тем выше риск переобучения без дополнительных ограничений [9;12;13].

Смещение–дисперсия. Обобщающая ошибка раскладывается на неизбежное смещение (bias) из-за неверной спецификации гипотезы, дисперсию (variance), вызванную вариативностью данных/обучения, и шум. Регуляризация (L2, L1, ранняя остановка, dropout как стохастическая регуляризация в сетях) уменьшает дисперсию ценой увеличения смещения - баланс выбирается по валидации [10;15].

Структурный риск (SRM). В основе таких методов, как опорные векторы, лежит принцип выбора гипотезы, минимизирующей верхнюю оценку истинного риска: эмпирический риск + штраф за сложность. Это обеспечивает контролируемую обобщающую способность [3].

Рис.1.Классы моделей

Рассмотрим классы моделей и их свойства:

Линейные и обобщённые линейные модели. Линейная регрессия, логистическая регрессия, GLM (с линком, например logit, log, identity)-

прозрачные модели с простой интерпретацией коэффициентов, выпуклой оптимизацией и хорошо изученными статистическими свойствами (несмещённость, состоятельность при корректной спецификации). Регуляризация L2 (ридж) и L1 (лассо) реализуют контроль сложности и отбор признаков [15].

Ядровые методы. Метод опорных векторов (SVM), ядровая логистическая регрессия, ядровые k-средних используют трюк ядра, позволяющий работать в высокоразмерных отображениях без явного перехода[10;12]. SVM с жёстким/мягким зазором оптимизирует выпуклую задачу, обладает хорошими теоретическими гарантиями обобщения через маржу. Минусы - чувствительность к выбору ядра/параметров и масштабу данных (квадратичная сложность по числу объектов в базовых реализациях).

Деревья решений и ансамбли. Деревья-дискретные, интерпретируемые модели с локальными правилами. Ансамбли-бустинг (AdaBoost, градиентный бустинг) и случайные леса-улучшают обобщение через агрегирование слабых моделей [7;10]. Бустинг часто даёт state-of-the-art на табличных данных; случайный лес устойчив к шуму и мало чувствителен к масштабированию признаков. Теоретически ансамбли снижают дисперсию (бэггинг) и/или смещение (бустинг) при должной регуляризации (ограничение глубины, темп обучения, число деревьев).

Вероятностные графические модели. Байесовские сети (направленные ациклические графы) и марковские случайные поля (ненаправленные) описывают факторизацию совместного распределения, позволяют явную причинно-следственную интерпретацию и реализацию выводов [5;11;15]. Инференс может быть точным (устранение переменных на деревьях кликов) или приближённым (MCMC, вариационные методы). Байесовская линейная регрессия и логистическая регрессия - частные случаи с априорами на параметрах.

Латентные и факторизационные модели. PCA, факторный анализ, NMF, SVD/ALS для коллаборативной фильтрации предоставляют компактные представления, подавляют шум и уменьшают размерность [10;15]. Теория охватывает спектральные свойства (сходимость, устойчивость), регуляризацию ранга/ядерной нормой.

Нейросетевые архитектуры. Полносвязные сети, сверточные (CNN), рекуррентные/трансформеры - универсальные аппроксиматоры с высокой выразительностью. Обучение - стохастический градиент (SGD) и его варианты (Adam, momentum), регуляризация - dropout, weight decay, data augmentation[4;14]. Современная теория обобщения сетей рассматривает нормы весов, плоскостность минимума, маржу, нейронную тюрему NTK, двойной спуск (double descent) - парадоксальную динамику ошибки при росте параметров сверх-порогов интерполяции.

Гибриды и нейро-символьные подходы. Интеграция с логикой и ограничениями (loss-ы соответствия правилам, дифференцируемые логические операторы, обучение «под ограничениями») соединяет обучаемые представления с верифицируемыми свойствами - критично для ИИ, работающего в правовых/медицинских доменах [14].

Таблица 1.

Сравнение семейств моделей машинного обучения

Семейство моделей	**Теоретическая база**	**Сложность обучения (тип.)**	**Обобщение и контроль сложности**	**Интерпрет и-руемость**	**Типичные области**
Линейные/GLM	Статистика, выпуклая оптимизация	Низкая–средняя	Регуляризация L1/L2, SRM	Высокая	Табличные данные, скоринг
Ядровые (SVM, ядровая логистика)	Маржа, ядровые методы	Средняя–высокая	Маржа, выбор ядра, C	Средняя	Классификация сложных границ

Деревья и ансамбли	Теория ансамблей, бэггинг/бустинг	Средняя	Ограничение глубины, темп, число деревьев	Средняя–высокая (деревья)	Табличные, ранжирование
Вероятностные графы	Байесовский вывод, факторизация	Средняя–высокая	Априоры, вариации/MCMC	Средняя–высокая	Диагностика, причинность
Латентные/факторизационные	Лин. алгебра, спектральная теория	Низкая–средняя	Ранг/ядерная норма	Средняя	Сжатие, рекомендации
Нейросети (DNN)	Функц. аппроксимация, SGD	Средняя–высокая	Weight decay, dropout, ранняя остановка	Низкая–средняя (post-hoc)	Изображения, речь, НЛП

Оптимизация и вычислительные аспекты. Многие модели сводятся к выпуклым задачам (ридж, лассо, логистическая регрессия, SVM), что гарантирует нахождение глобального минимума и известные скорости сходимости (градиентные/координатные методы, ускорения Нестерова). Нейросети приводят к невыпуклым задачам; тем не менее стохастическая оптимизация и явление «благоприятных ландшафтов» на практических архитектурах обеспечивают достижение хороших минимумов [10;14;15]. Важно учитывать **условное число** (conditioning), масштабы признаков и стратегии инициализации, нормализацию (Batch/Layer Norm), что влияет на сходимость и обобщение.

Регуляризация и управление сложностью. Регуляризационные техники реализуют априорные предпочтения: гладкость (L2), разреженность (L1), низкий ранг (ядерная норма), ограничение глубины деревьев и маржи (SVM), расширение данных (augmentation), случайное зануление (dropout). Байесовская регуляризация интерпретируется как априорные распределения на параметрах. Выбор силы регуляризации производится по валидации (k-fold,

отложенная выборка), с поправкой на множественные сравнения и утечки [9;10].

Объяснимость и проверяемость. Интерпретируемые модели (линейные, GLM, неглубокие деревья) дают локально понятные коэффициенты/правила. Для сложных моделей применяются post-hoc методы: локальные аппроксимации (LIME), атрибуции (SHAP, интегрированные градиенты), прототипы/критические примеры. Вероятностные модели дают причинные и контрфактические объяснения через структурные уравнения и байесовский вывод [10;15]. В критичных доменах требования нормативов требуют трассируемости, мониторинга дрейфа и периодической переоценки рисков.

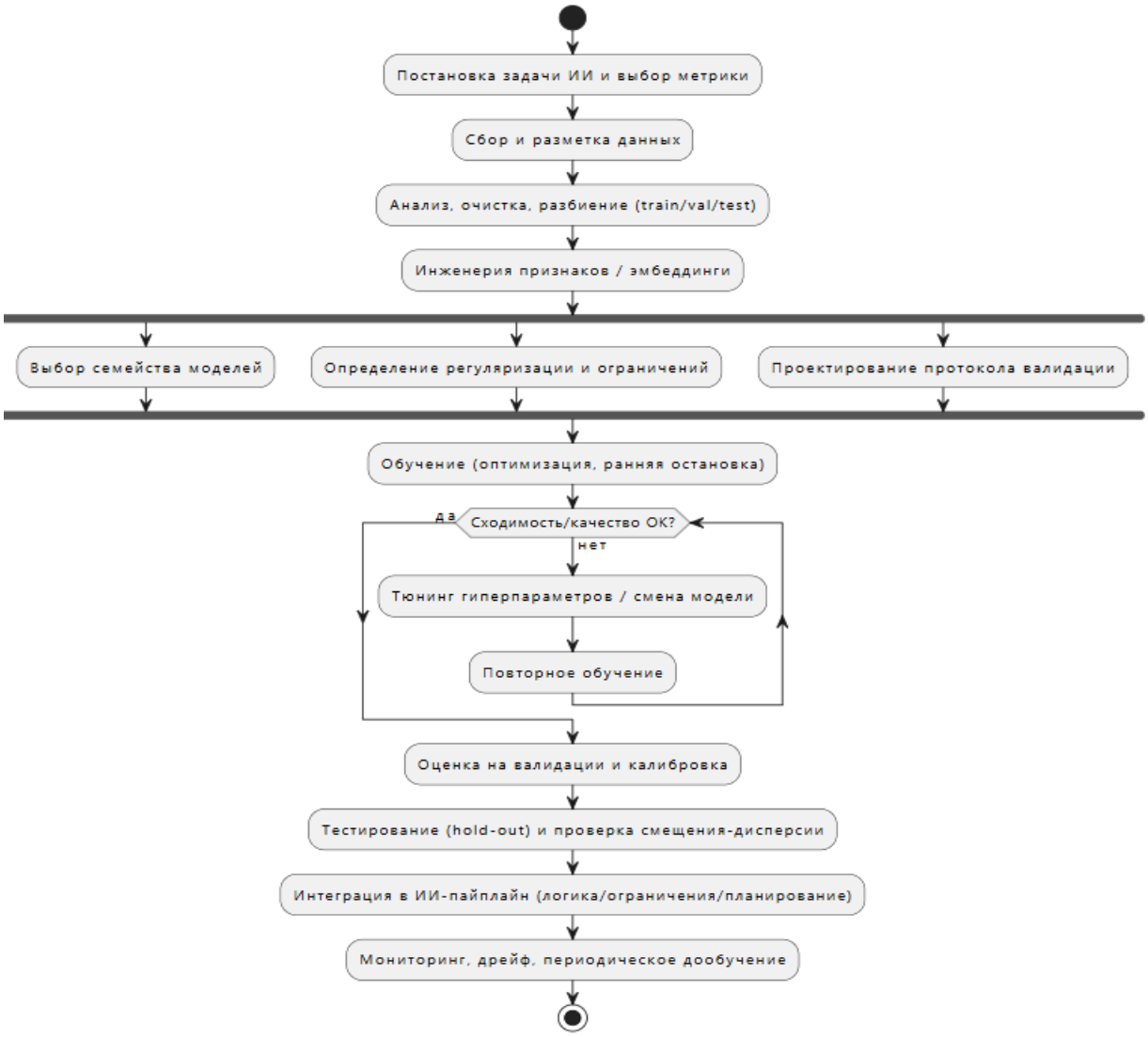

Рис.2. Конвейер машинного обучения в ИИ

Модели машинного обучения являются не приложением к искусственному интеллекту, а его **теоретическим ядром** в части статистического вывода и аппроксимации [10]. Именно здесь формализуются представления о сложности гипотез, механизм обобщения, компромисс смещения–дисперсии, роль регуляризации и априорных предположений. С другой стороны, сами модели - лишь часть общей архитектуры ИИ: их результаты должны соединяться с правилами, целями и ограничениями, обеспечивая объяснимость, проверяемость и безопасность.

Сопоставление семейств моделей показывает, что **нет универсального решения**: линейные модели выигрывают интерпретируемостью и скоростью, ядровые - геометрией маржи, деревья/ансамбли - устойчивостью на табличных данных, вероятностные графы - причинным анализом и явной неопределённостью, нейросети - выразительностью и переносом на «сырые» сигналы. Теоретические критерии (PAC-границы, VC-размерность, Rademacher-сложность, принципы байесовской регуляризации) помогают принимать осмысленные инженерные решения и прогнозировать поведение моделей за пределами обучающей выборки.

Будущее ИИ - за **синтезом** статистических и логических методов: нейро-символьными системами, формальными ограничениями поверх обучаемых представлений, безопасной оптимизацией и мониторингом жизненного цикла моделей. Такой подход позволит строить интеллектуальные системы, которые не только «точно предсказывают», но и «понимают, почему», - а значит, заслуживают доверия в критически важных сферах.

Список литературы:

1. Бишоп К. **Методы машинного обучения.** Пер. с англ. - М.: ДМК Пресс, 2020. - 752 с.

2. Вапник В. Н. **Восстановление зависимостей по эмпирическим данным.** - М.: Наука, 1979. - 416 с.

3. Вапник В. Н. **Статистическая теория обучения.** - М.: Машиностроение, 1999. - 528 с.

4. Гудфеллоу И., Бенджио И., Курвилл А. **Глубокое обучение.** Пер. с англ. - М.: ДМК Пресс, 2018. - 652 с.

5. Мёрфи К. П. **Машинное обучение. Вероятностная перспектива.** Пер. с англ. - М.: ДМК Пресс, 2023. - 1080 с.

6. Митчелл Т. **Машинное обучение.** Пер. с англ. - М.: Вильямс, 2006. - 432 с.

7. Рашка С., Мирджалили В. **Python и машинное обучение.** 3-е изд. - СПб.: Питер, 2020. - 768 с.

8. Хасти Т., Тибширани Р., Фридман Дж. **Статистическое обучение с применениями в R.** Пер. с англ. - М.: ДМК Пресс, 2017. - 768 с.

9. Шалев-Шварц Ш., Бен-Давид Ш. **Понимание машинного обучения: теория и алгоритмы.** Пер. с англ. - М.: ДМК Пресс, 2018. - 448 с.

10. Hastie T., Tibshirani R., Friedman J. **The Elements of Statistical Learning.** 2nd ed. - New York: Springer, 2009. - 745 p.

11. Murphy K. P. **Probabilistic Machine Learning: An Introduction.** - MIT Press, 2022. - 1104 p.

12. Vapnik V. N. **Statistical Learning Theory.** - New York: Wiley, 1998.- 736 p.

13. Valiant L. G. **A Theory of the Learnable.** // *Communications of the ACM*, 1984, Vol. 27, No. 11, pp. 1134–1142.

14. Goodfellow I., Bengio Y., Courville A. **Deep Learning.** - MIT Press, 2016. - 800 p.

15. Bishop C. M. **Pattern Recognition and Machine Learning.** - New York: Springer, 2006. - 738 p.