
УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«25» мая 2026 г.
Протокол №2

**Рабочая программа дисциплины (модуля)
«Машинное обучение. Продвинутый уровень»**

Направление подготовки: 02.03.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Прикладной искусственный интеллект

Квалификация (степень) выпускника: бакалавр

Форма обучения: очная

Срок освоения программы: 4 года

Год набора: 2026

**Москва
2026**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения	5
3. Тематический план	8
4. Содержание дисциплины (модуля)	8
5. Учебно-методическое обеспечение	10
6. Материально-техническое обеспечение	10
7. Методические и оценочные материалы	12

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Машинное обучение. Продвинутый уровень» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по специальности 02.03.01 «Математика и компьютерные науки», профиль «Прикладной искусственный интеллект», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Машинное обучение. Продвинутый уровень» позволяет студентам научиться разрабатывать модели, способные анализировать большие объемы данных и делать предсказания, что находит применение в различных отраслях. Кроме того, знание методов машинного обучения способствует автоматизации процессов и улучшению принятия решений.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 «Математика и компьютерные науки», профиль «Прикладной искусственный интеллект» и входит в Факультативные дисциплины.

Дисциплина (модуль) изучается на 2 курсе в 3 семестре, доступна к изучению после успешного освоения дисциплины (модуля): «Введение в статистику» или «Введение в искусственный интеллект».

Цель изучения дисциплины (модуля): заключается в формирование у студентов теоретических знаний и практических навыков по основам машинного обучения, овладение студентами инструментарием, моделями и методами машинного обучения, а также приобретение навыков исследователя данных (data scientist) и разработчика математических моделей, методов и алгоритмов анализа данных.

Задачи изучения дисциплины (модуля):

- освоение основ теории и алгоритмов машинного обучения;
- развитие умений работы с данными и построения моделей;
- формирование навыков программирования и критического мышления.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- теорию ансамблирования и мета-обучения: математические основы бэггинга и бустинга (AdaBoost, GBM, XGBoost, LightGBM, CatBoost), out-of-bag оценку, расчет важности признаков, а также архитектуру стекинга и блендинга;
- ядровые методы и байесовский подход: принципы работы SVM (прямая/двойственная задачи, алгоритм SMO), теорию Mercer-ядер, а также аппарат гауссовских процессов и вероятностных графических моделей (байесовские и марковские сети);
- алгоритмы со скрытыми переменными и Unsupervised ML: математическую логику EM-алгоритма (на примере GMM и HMM), основы факторизации матриц (PCA, SVD, NMF) и нелинейного снижения размерности (t-SNE, UMAP);
- анализ и прогнозирование временных рядов: математику классических статистических моделей (ARIMA, SARIMA, экспоненциальное сглаживание), понятия стационарности, автокорреляции (ACF/PACF) и специфику валидации данных с учетом времени;

уметь:

- строить высокоточные композиции моделей: настраивать современные ансамбли (XGBoost, LightGBM, CatBoost), эффективно обрабатывать категориальные признаки, подбирать мета-алгоритмы в стекинге и бороться с переобучением;

— моделировать вероятностные зависимости и скрытые структуры: оценивать неопределенность прогнозов с помощью гауссовских процессов, а также применять алгоритмы forward-backward и Витерби для анализа последовательностей в скрытых марковских моделях;

— выделять скрытые паттерны и аномалии: проводить комплексную кластеризацию высокой плотности (DBSCAN/HDBSCAN), визуализировать сложные многообразия данных, а также находить выбросы через статистические и ансамблевые подходы;

— работать с временными зависимостями: осуществлять проверку рядов на стационарность, генерировать продвинутое признаки (лаги, скользящие статистики) и реализовывать стратегии многошагового прогнозирования с помощью градиентного бустинга;

владеть навыками:

— промышленными фреймворками бустинга и ансамблирования: методами оптимизации, регуляризации и конфигурации современных библиотек (XGBoost, LightGBM, CatBoost) для работы в продакшене;

— методами точного вывода и оптимизации: алгоритмами Variable Elimination и Belief Propagation для вероятностных графических сетей, а также алгоритмами EM и SMO для ядровых и вероятностных моделей;

— инструментарием Unsupervised Learning и Anomaly Detection: алгоритмами плотностной кластеризации (DBSCAN, HDBSCAN), методами нелинейного погружения (t-SNE, UMAP) и детекторами аномалий (Isolation Forest, LOF, One-Class SVM);

— аппаратом валидации и анализа последовательностей: классическими статистическими пакетами для временных рядов (ARIMA/SARIMA) и специфическими техниками валидации моделей машинного обучения в условиях временных задержек (Time Series Split).

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Код и содержание компетенции	Индикатор компетенции и перечень планируемых результатов обучения по дисциплине (модулю)
УК-1. Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1. Знает методы поиска и анализа информации в области разработки, основные принципы критической оценки источников информации и их релевантности.
	УК-1.2. Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем.
	УК-1.3. Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации.
УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1. Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач.
	УК-2.2. Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения.
	УК-2.3. Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями.
ОПК-1. Способен консультировать и использовать фундаментальные знания в области математического анализа, комплексного и функционального анализа алгебры, аналитической геометрии, дифференциальной геометрии и топологии, дифференциальных уравнений, дискретной математики и математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов, теоретической механики в профессиональной деятельности	ОПК-1.1. Знает основные понятия, методы и результаты математического анализа, комплексного и функционального анализа, алгебры, аналитической и дифференциальной геометрии, топологии, дифференциальных уравнений, дискретной математики и математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов и теоретической механики.
	ОПК-1.2. Умеет применять фундаментальные математические знания и методы для анализа и решения профессиональных задач.
	ОПК-1.3. Имеет практический опыт выбора, разработки и реализации математических моделей и методов при решении задач профессиональной деятельности.

МО-1. Способен проектировать, разрабатывает и улучшать алгоритмы и модели машинного обучения	МО-1.1. Применяет классические методы и модели машинного обучения (МО) (базовый) МО-1.1 З-1. Знает алгоритмы обучения с учителем для регрессии и классификации. Линейные методы, ансамблевые методы, методы на опорных векторах. МО-1.1 З-1. Знает алгоритмы обучения без учителя. Методы кластеризации и поиска аномалий. МО-1.1 З-1. Знает основные ограничения классического МО и области применения. МО-1.1 У-1. Умеет применять стандартные библиотеки для обучения с учителем и без учителя. МО-1.1 У-2. Умеет использовать инструменты очистки данных и предварительной подготовки данных для статистического обучения (масштабирование, корреляционный анализ, заполнение пропусков в данных). МО-1.1 У-3. Умеет самостоятельно строить пайплайн очистки данных, обучения, подбора гиперпараметров и валидации моделей МО.
	МО-1.2. Применяет глубокие нейронные сети (средний) МО-1.2. З-1. Знает теоретические основы нейросетевых алгоритмов. МО-1.2. З-2. Знает методы обучения нейронных сетей с помощью градиентного спуска и обратного распространения ошибки. МО-1.2. З-3. Знает популярные архитектуры глубоких нейронных сетей и их применение: сверточные, рекуррентные, сети с памятью. МО-1.2. З-4. Знает механизм внимания в нейронных сетях и архитектуру трансформера. МО-1.2. У-1. Умеет применять стандартные библиотеки, реализующие основные архитектуры глубокого обучения. МО-1.2. У-2. Умеет разрабатывать нейронные сети с вниманием. МО-1.2. У-3. Умеет использовать предварительно обученные нейронные сети и перенос обучения.
	МО-1.3. Применяет автоматическое машинное обучение (средний) МО-1.3. З-1. Знает алгоритмы автоматизации работы с признаками. МО-1.3. З-1. Знает алгоритмы автоматической настройки гиперпараметров. МО-1.3. З-1. Знает алгоритмы автоматического поиска структуры модели (нейронных сетей и других графовых моделей). МО-1.3. У-1. Умеет обосновывать способы и варианты применения алгоритмов автоматического машинного обучения в задачах ИИ, включая их преобразование и адаптацию к специфике задачи МО-1.3. У-1. Умеет применять методы повышения устойчивости, надежности, безопасности алгоритмов автоматического машинного обучения для проверки разведочных гипотез и подготовки данных к применению современных методов ИИ

ППК-Р1. Способен разрабатывать и отлаживать программный код	ППК-Р1.1. Выполняет формализацию и алгоритмизацию поставленных задач для разработки программного кода (базовый) ППК-Р1.1. 3-1. Знает алгоритмы решения типичных задач, области и способы их применения. ППК-Р1.1. 3-2. Знает нотации и программное обеспечение для графического отображения алгоритмов. ППК-Р1.1. 3-3. Знает методы и приемы алгоритмизации поставленных задач. ППК-Р1.1. У-1. Умеет использовать методы и приемы формализации и алгоритмизации поставленных задач. ППК-Р1.1. У-2. Умеет применять алгоритмы решения типовых задач в соответствующих областях. ППК-Р1.2. Формализует задачу ИТ отрасли в язык естественнонаучных дисциплин (базовый) ППК-Р1.2. 3-1. Знает основные разделы математики, применяемые для анализа и моделирования непрерывных процессов и дискретных систем в прикладных задачах. ППК-Р1.2. У-1. Умеет выбирать адекватный математический аппарат для формализованного описания сущностей и отношений на основе ТЗ и бизнес-требований. ППК-Р1.2. В-1. Владеет навыками описания задач предметной области в виде формальных математических моделей, пригодных для последующей алгоритмизации и программной реализации.
	ППК-Р1.3. Осуществляет обоснованный выбор методов и алгоритмов для программной реализации формальной математической модели (базовый) ППК-Р1.3. 3-1. Знает основные классы методов программной реализации моделей и критерии выбора алгоритмов. ППК-Р1.3. У-1. Умеет проводить сравнительный анализ и обоснование выбора алгоритмов для программной реализации модели. ППК-Р1.3. В-1. Владеет методами адаптации методов и алгоритмов под специфику задачи. ППК-Р1.3. В-2. Владеет навыками оценки эффективности выбранных алгоритмов.

3. Тематический план

№п/ п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятель ная работа	
		Лекции	Практичес кие занятия			
1	Ансамблевые методы и градиентный бустинг		6		16	Домашнее задание Тест
2	Ядровые методы и метод опорных векторов		6		16	Домашнее задание Тест
3	Вероятностные графические модели и байесовский вывод		8		16	Домашнее задание Тест
4	Продвинутые методы обучения без учителя		6		16	Домашнее задание Тест
5	Анализ и прогнозирование временных рядов		4		18	Домашнее задание Тест
	Зачет			2		Проект
Итого:			30	2	82	
Объем дисциплины (модуля) (в ак. ч.)		114				
Объем дисциплины (модуля) (в зач. ед.)		3				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Ансамблевые методы и градиентный бустинг	Бэггинг и случайные леса: бутстрэп-агрегирование, out-of-bag-оценка, важность признаков. Бустинг: AdaBoost и градиентный бустинг (GBM). Современные реализации: XGBoost, LightGBM, CatBoost, работа с категориальными признаками. Стекинг и блендинг: мета-обучение, подбор ансамблей. Продвинутое обучение без учителя и борьба с переобучением в ансамблях.
2	Ядровые методы и метод опорных векторов	Теория ядер: Mercer-ядра, спрямляющее пространство, reproducing kernel Гильбертовы пространства. Примеры ядер для текста, графов, изображений. SVM-классификация: прямая и двойственная задача, мягкий зазор, алгоритм SMO. SVM-регрессия и одноклассовый SVM. Ядровые расширения: Ядерный PCA, Ядерная гребневая регрессия. Связь с гауссовскими процессами и регуляризацией.
3	Вероятностные графические модели и байесовский вывод	Байесовские методы в ML: сопряжённые априорные распределения, байесовская линейная регрессия, evidence approximation. Гауссовские процессы: регрессия и классификация, выбор ковариационной функции, оптимизация гиперпараметров, оценка неопределённости.

		Вероятностные графические модели: байесовские сети, марковские сети, d-разделимость, точный вывод. Скрытые переменные и ЕМ-алгоритм: смеси гауссовских распределений, скрытые марковские модели (HMM), алгоритм Витерби.
4	Продвинутые методы обучения без учителя	Факторизация матриц и понижение размерности: PCA, вероятностный PCA, NMF, SVD-приложения. Многообразия: t-SNE, UMAP. Кластеризация высокой плотности: DBSCAN, HDBSCAN, спектральная кластеризация. Метрики качества без учителя. Обнаружение аномалий и выбросов: LOF, одноклассовый SVM, статистические и ансамблевые подходы.
5	Анализ и прогнозирование временных рядов	Классические модели временных рядов: ARIMA, SARIMA, экспоненциальное сглаживание, проверка на стационарность, ACF/PACF. Машинное обучение на временных рядах: конструирование признаков (лаги, скользящие статистики), градиентный бустинг для прогноза, стратегии многошагового прогноза, валидация с учётом времени.

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебник / П. Флах. — 2-е изд. — Москва : ДМК Пресс, 2023. — 401 с. — ISBN 978-5-89818-300-4. — URL: <https://znanium.ru/catalog/product/2085038>.
2. Мэрфи, К. П. Вероятностное машинное обучение: введение : практическое руководство / К. П. Мэрфи ; пер. с англ. А. А. Слинкина. — Москва : ДМК Пресс, 2023. — 990 с. — ISBN 978-5-93700-119-1. — URL: <https://znanium.ru/catalog/product/2109489>.
3. Шолле, Ф. Глубокое обучение на Python : практическое руководство / Ф. Шолле. - Санкт-Петербург : Питер, 2023. - 576 с. - (Библиотека программиста). - ISBN 978-5-4461-1909-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2123389>.

Дополнительная литература:

1. Дьяконов, А.Г. Машинное обучение и анализ данных / А.Г. Дьяконов. — URL: https://github.com/Dyakonov/MLDM_BOOK/blob/main/README.md.
2. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2025. — 89 с. — (Высшее образование). — ISBN 978-5-534-20732-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/558662>.
3. Введение в статистическое обучение с примерами на языке R / Джеймс Г. [и др.]. — 2-е изд., испр. — Москва : ДМК Пресс, 2018. — 455 с. : ил. — Библиогр. : с. 450—455.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		

Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Машинное обучение. Продвинутый уровень» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекция, семинары и домашние задания, проект, тесты, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар – это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Проект – исследовательская работа по курсу и презентация результатов.

Для успешной подготовки к проекту: четко определите цели и задачи проекта, распределите роли и обязанности между участниками, а также установите сроки выполнения каждой части работы. Регулярно проводите встречи для обсуждения прогресса и решения возникающих вопросов.

Тест – особая форма проверки знаний. Проводится после освоения одной или нескольких тем и свидетельствует о качестве понимания основных понятий изучаемого

материала. Тестовые задания составлены к ключевым понятиям, основным разделам, важным терминологическим категориям изучаемой дисциплины.

Для подготовки к тесту необходимо знать терминологический аппарат дисциплины, понимать смысл научных категорий и уметь их использовать в профессиональной лексике. Владение понятийным аппаратом, включённым в тестовые задания, позволяет преподавателю быстро проверить уровень понимания студентами важных методологических категорий.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Машинное обучение. Продвинутый уровень»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *зачета*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
9	Отлично	
8	Отлично	

Десятибалльная оценка	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
7	Хорошо	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	
5	Удовлетворительно	Студент обладает базовыми знаниями по дисциплине, но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	
3	Не сдан	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	
1	Не сдан	

Дисциплина (модуль) «Машинное обучение. Продвинутый уровень» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	40%	15	Набор задач по темам недели
Тест	20%	15	Форма проверки знаний студентов через ответы на вопросы с выбором из нескольких вариантов ответов
Зачет	40%	1	Проект - исследовательская работа по дисциплине (модулю) и презентация результатов

Формула расчёта итоговой оценки по дисциплине (модулю) «Машинное обучение. Продвинутый уровень»: « $0,4 \times \text{среднее за домашние задания} + 0,2 \times \text{среднее за тесты} + 0,4 \times \text{Зачет}$ ».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные задания для теста

Тест 1.

1. В чем заключается основное отличие бэггинга (Bagging) от бустинга (Boosting) с точки зрения смещения (bias) и разброса (variance)?

- А) Бэггинг уменьшает смещение, бустинг уменьшает разброс.
- В) Бэггинг уменьшает разброс, бустинг уменьшает смещение.
- С) Оба метода уменьшают только смещение.

D) Оба метода уменьшают только разброс.

Ответ: В

2. Какая особенность реализации CatBoost позволяет ей эффективно работать с категориальными признаками без предварительного One-Hot Encoding?

A) Использование гистограммного обучения.

B) Ordered Boosting и обработка категорий на лету (target statistics).

C) Применение ядерных методов внутри деревьев.

D) Автоматическое удаление редких категорий.

Ответ: В

3. Что подается на вход мета-обучающему алгоритму (meta-learner) при использовании техники Стекинг (Stacking)?

A) Исходные признаки обучающей выборки.

B) Предсказания базовых моделей (out-of-fold predictions).

C) Градиенты функции потерь базовых моделей.

D) Случайный шум для регуляризации.

Ответ: В

4. Для чего используется Out-of-Bag (OOB) оценка в методе Случайного Леса?

A) Для ускорения обучения деревьев.

B) Для оценки качества модели без использования отдельной валидационной выборки.

C) Для отбора наиболее важных признаков.

D) Для балансировки классов в датасете.

Ответ: В

5. Какую роль играет регуляризация (L1/L2) в функции потерь XGBoost?

A) Увеличивает глубину деревьев.

B) Уменьшает скорость обучения.

C) Контролирует сложность модели и борется с переобучением.

D) Преобразует категориальные признаки в числовые.

Ответ: С.

Тест 2.

1. В чем заключается основная идея «ядерного трюка» (Kernel Trick)?

A) Сокращение размерности данных перед обучением.

B) Вычисление скалярных произведений в пространстве высокой размерности без явного построения этого пространства.

C) Удаление выбросов из обучающей выборки.

D) Преобразование нелинейной задачи в линейную путем удаления признаков.

Ответ: В

2. За что отвечает параметр C в SVM с мягким зазором (Soft-Margin SVM)?

A) За ширину разделяющей полосы.

B) За выбор ядра.

C) За компромисс между максимизацией зазора и минимизацией ошибок классификации.

D) За количество итераций алгоритма SMO.

Ответ: С

3. Какие объекты называются опорными векторами (Support Vectors) в методе SVM?

A) Все объекты обучающей выборки.

B) Объекты, лежащие ближе всего к разделяющей гиперплоскости или внутри зазора.

C) Объекты, которые модель классифицирует неверно.

D) Центроиды кластеров.

Ответ: В

4. В каком случае целесообразно использовать Ядерный PCA (Kernel PCA) вместо обычного линейного PCA?

- A) Когда данные имеют линейную структуру.
- B) Когда данных очень много и нужно ускорить вычисления.
- C) Когда данные лежат на нелинейном многообразии и линейное разделение невозможно.

D) Когда в данных много пропусков.

Ответ: C

5. Какое условие должно выполняться для функции, чтобы она могла быть использована как ядро (условие Мерсера)?

- A) Функция должна быть линейной.
- B) Матрица Грама, составленная из значений функции, должна быть положительно полуопределенной.
- C) Функция должна быть выпуклой.
- D) Функция должна возвращать только целые значения.

Ответ: B

Тест 3.

1. Какова основная цель EM-алгоритма (Expectation-Maximization)?

- A) Минимизация функции потерь градиентным спуском.
- B) Оценка параметров модели при наличии скрытых (латентных) переменных.
- C) Построение доверительных интервалов для прогнозов.
- D) Отбор признаков на основе взаимной информации.

Ответ: B

2. Какое ключевое преимущество Гауссовских процессов (Gaussian Processes) перед другими методами регрессии?

- A) Они работают быстрее всех.
- B) Они не требуют настройки гиперпараметров.
- C) Они дают оценку неопределенности (дисперсии) прогноза вместе с самим прогнозом.
- D) Они работают только с категориальными данными.

Ответ: C

3. Что позволяет определить свойство d-разделимости (d-separation) в байесовских сетях?

- A) Оптимальную структуру графа.
- B) Условную независимость между переменными при заданных наблюдениях.
- C) Вероятность скрытых состояний.
- D) Сходимость алгоритма вывода.

Ответ: B

4. Зачем используются сопряжённые априорные распределения (Conjugate Priors) в байесовском выводе?

- A) Чтобы сделать апостериорное распределение того же семейства, что и априорное, для аналитического решения.
- B) Чтобы уменьшить объем вычислений в 10 раз.
- C) Чтобы избежать переобучения модели.
- D) Чтобы работать с большими данными.

Ответ: A

5. Какую задачу решает алгоритм Витерби в контексте Скрытых Марковских Моделей (НММ)?

- A) Оценку параметров модели (обучение).
- B) Вычисление вероятности наблюдений.
- C) Поиск наиболее вероятной последовательности скрытых состояний.
- D) Кластеризацию временных рядов.

Ответ: C.

Примерное задание для проекта

Задание для проекта по теме: Линейные модели

Студентам предлагается провести анализ данных с использованием линейной регрессии, логистической регрессии и метода опорных векторов (SVM). Для этого необходимо выбрать набор данных из открытых источников (например, Kaggle, UCI Machine Learning Repository) и выполнить следующие шаги:

1. Выбор и описание данных:

- Выберите набор данных, который подходит для анализа с использованием линейной и логистической регрессии, а также SVM.
- Опишите данные: какие переменные присутствуют, каковы их типы, и какую задачу вы собираетесь решить (регрессия или классификация).

2. Предобработка данных:

- Проведите анализ данных на наличие пропусков и аномалий.
- Выполните необходимые преобразования, такие как нормализация, кодирование категориальных переменных и удаление выбросов.

3. Моделирование:

- Постройте модель линейной регрессии и оцените её качество, используя такие метрики, как MSE (среднеквадратичная ошибка) или R^2 .
- Постройте модель логистической регрессии и оцените её качество, используя такие метрики, как accuracy, precision, recall и F1-score.
- Постройте модель SVM и сравните её качество с предыдущими моделями.

4. Сравнительный анализ:

- Сравните результаты всех трех моделей. Обсудите, какая модель показала лучшие результаты и почему.
- Проанализируйте влияние различных факторов на качество моделей.

5. Выводы:

- Сделайте выводы о применимости различных моделей для решения поставленной задачи и предложите рекомендации по выбору модели в зависимости от специфики данных.

Критерии оформления проекта:

Структура:

- Введение (описание задачи и целей проекта).
- Описание данных (выбор набора данных и его характеристика).
- Предобработка данных (методы и результаты).
- Моделирование (описание каждой модели, результаты и метрики).
- Сравнительный анализ (сравнение моделей и выводы).
- Заключение (общие выводы и рекомендации).

Оформление:

- Проект должен быть оформлен в виде документа (например, PDF или Word).
- Используйте графики и таблицы для визуализации результатов.
- Все коды должны быть оформлены и комментированы (можно использовать Jupyter Notebook или аналогичные инструменты).

Объем:

Минимальный объем проекта — 10 страниц (без учета графиков и таблиц).

Сдача проекта:

- Проект необходимо представить в электронном виде до установленного срока (указать дату).
- Презентация проекта (10-15 минут) должна быть подготовлена для защиты: краткое изложение целей, методов, результатов и выводов.

— Оценка проекта будет проводиться на основе качества анализа, глубины понимания темы, ясности изложения и оригинальности подхода.

Примерные домашние задания

Домашнее задание 1.

Задание 1: Метрики и функционалы потерь

Найдите примеры функционалов потерь, используемых в задачах регрессии и классификации. Опишите их особенности и применения. Кроме того, объясните понятие средней гипотезы и ее связь с функционалами потерь.

Задание 2: Контроль качества и выбор модели

Опишите методы контроля качества моделей машинного обучения, такие как валидация и кросс-валидация. Explain понятие ошибки внутри и вне выборки и ошибки обобщения. Кроме того, объясните неравенство Хёфдинга и его применение в контроле качества моделей.

Задание 3: Метрические алгоритмы

Найдите примеры метрических алгоритмов, используемых в задачах классификации и регрессии. Опишите их особенности и применения. Кроме того, объясните понятие размерности Вапника-Червоненкиса и ее связь с метрическими алгоритмами.

Домашнее задание 2.

Задание 1: Линейная регрессия

Дан набор данных, содержащий информацию о цене на жилье и площади квартир в городе. Используя линейную регрессию, построить модель, которая позволяет предсказать цену на жилье на основе площади квартиры.

- Найдите коэффициенты линейной регрессии (β_0 , β_1) используя метод наименьших квадратов.
- Оцените качество модели с помощью коэффициента детерминации (R^2).
- Дайте интерпретацию результатов и объясните, как можно использовать эту модель для предсказания цены на жилье.

Задание 2: Логистическая регрессия и SVM

Дан набор данных, содержащий информацию о клиентах банка и их кредитной истории. Используя логистическую регрессию и SVM, построить модели, которые позволяют предсказать вероятность того, что клиент вернет кредит.

- Найдите коэффициенты логистической регрессии (β_0 , β_1) используя метод максимального правдоподобия.
- Построить модель SVM с линейным ядром и оценить ее качество с помощью коэффициента точности.
- Дайте интерпретацию результатов и объясните, как можно использовать эти модели для предсказания вероятности возврата кредита.

Задание 3: Линейные модели

Дан набор данных, содержащий информацию о влиянии различных факторов на результаты студентов в университете. Используя линейные модели, построить модель, которая позволяет предсказать результаты студентов на основе следующих факторов: возраст, пол, количество посещенных занятий и средний балл по предметам.

- Найдите коэффициенты линейной регрессии (β_0 , β_1 , ..., β_n) используя метод наименьших квадратов.
- Оцените качество модели с помощью коэффициента детерминации (R^2).

- Дайте интерпретацию результатов и объясните, как можно использовать эту модель для предсказания результатов студентов.

Домашнее задание 3.

Задача 1. Реализуйте алгоритмы построения дерева с критерием информационного выигрыша и критерием Джини и определению класса по мажоритарному классу в листе. Найдите оптимальную глубину дерева в обоих случаях (в отрезке 2-10).

Задача 2. Примените метод SVM (например, из библиотеки sklearn) для датасета blobs2. Визуализируйте результат (разбиение плоскости и опорные вектора) при разных вариантах ядер (линейное; полиномиальное степеней 2,3,5; RBF).

Задача 3. Найдите максимум функции с помощью алгоритма кросс-энтропийного поиска, изображая распределение на каждом шаге.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция	Индикатор
1	Назовите метод оценки качества ансамбля без выделения отдельной валидационной выборки	Out-of-bag оценка	УК-1	УК-1.2
2	Укажите источник для критического анализа реализаций градиентного бустинга	Рецензируемые статьи KDD NeurIPS JMLR	УК-1	УК-1.3
3	Определите подход к систематизации информации о ядровых методах из различных материалов	Классификация по типу ядра и задаче	УК-1	УК-1.2
4	Сформулируйте задачу поиска инструментов для сравнения XGBoost и CatBoost	Бенчмаркинг на тестовых датасетах	УК-1	УК-1.3
5	Выберите алгоритм оптимизации для решения двойственной задачи SVM	Алгоритм SMO	УК-2	УК-2.2
6	Укажите ограничение при выборе ядра для работы с графовыми данными	Вычислительная сложность ядровой матрицы	УК-2	УК-2.2
7	Назовите критерий выбора между мягким и жёстким зазором в SVM	Уровень шума в данных и допустимая ошибка	УК-2	УК-2.3
8	Определите круг задач для применения одноклассового SVM в соответствии с ограничениями	Обнаружение аномалий при отсутствии размеченных данных	УК-2	УК-2.3
9	Дополните утверждение: сопряжённое априорное распределение в байесовской регрессии обеспечивает	Аналитическую форму апостериорного распределения	ОПК-1	ОПК-1.2
10	Назовите функцию для оптимизации гиперпараметров в гауссовских процессах	Маргинальное правдоподобие	ОПК-1	ОПК-1.2

11	Укажите свойство графа для проверки условной независимости в байесовских сетях	D-разделимость	ОПК-1	ОПК-1.1
12	Определите критерий остановки ЕМ-алгоритма для смеси гауссовских распределений	Сходимость логарифмического правдоподобия	ОПК-1	ОПК-1.3
13	Выберите алгоритм кластеризации для обнаружения кластеров произвольной формы	DBSCAN или HDBSCAN	МО-1	МО-1.1
14	Назовите метод понижения размерности для визуализации многомерных данных с сохранением локальной структуры	t-SNE или UMAP	МО-1	МО-1.2
15	Укажите алгоритм поиска наиболее вероятной последовательности скрытых состояний в HMM	Алгоритм Витерби	МО-1	МО-1.3
16	Определите метод обнаружения аномалий на основе оценки локальной плотности точек	LOF или Local Outlier Factor	МО-1	МО-1.2
17	Назовите тест для проверки стационарности временного ряда перед построением ARIMA	Тест Дики-Фуллера или ADF тест	ППК-Р1	ППК-Р1.2
18	Укажите функции для анализа автокорреляции при подборе параметров ARIMA	ACF и PACF	ППК-Р1	ППК-Р1.2
19	Дополните требование к валидации временных рядов с учётом временной зависимости	Валидация с расширяющимся или скользящим окном	ППК-Р1	ППК-Р1.3
20	Назовите признаки для машинного обучения на временных рядах перед применением бустинга	Лаги и скользящие статистики	ППК-Р1	ППК-Р1.3