
УТВЕРЖДЕНА

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

**Рабочая программа дисциплины (модуля)
«Введение в искусственный интеллект. 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	4
3. Тематический план	7
4. Содержание дисциплины (модуля)	7
5. Учебно-методическое обеспечение	8
6. Материально-техническое обеспечение	8
7. Методические и оценочные материалы	10

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Введение в искусственный интеллект. Продвинутый уровень» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению 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) «Введение в искусственный интеллект. Продвинутый уровень» (далее – ИИ) активно внедряется в различные сферы жизни, включая медицину, финансы и транспорт, что открывает новые возможности для повышения эффективности и качества услуг. Кроме того, понимание принципов работы ИИ способствует подготовке специалистов, способных разрабатывать эти технологии и решать возникающие этические и социальные вопросы.

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 1 или 2 курсе в 1, 2, 3 или 4 семестре.

Цель изучения дисциплины (модуля): заключается в освоении методов и алгоритмов, позволяющих системам имитировать интеллектуальное поведение человека для решения сложных задач.

Задачи изучения дисциплины (модуля):

- изучение основных доменов ИИ, задач машинного обучения, архитектур моделей и инструментов;
- практическое выполнение задач, таких как запуск моделей через интерфейсы, анализ результатов и интеграция ИИ в повседневную жизнь;
- развитие терминологии, способности различать типы задач и моделей, а также этической осведомленности (например, обсуждение предвзятости, приватности и ответственности в ИИ-приложениях).

В результате освоения дисциплины (модуля), обучающийся должен:

знать:

- различия между классическими моделями машинного обучения;
- различия в архитектурах различных ансамблей;
- математические обоснования машинного обучения: вероятностная интерпретация, bias-variance decomposition;
- продвинутые методы обработки изображений;

уметь:

- вручную реализовывать основные механизмы машинного обучения: цикл обучения, полносвязный перцептрон, среда;
- уметь пользоваться фреймворками для работы с машинным обучением на базовом уровне: scikit-learn, gymnasium;

владеть:

- сравнения классических моделей машинного обучения и выбора подходящей модели под задачу, структуру данных и ограничения;
- подбора архитектуры ансамбля для улучшения качества модели.

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-Р1.2. 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.
--	--

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	Основы машинного обучения		14		14	Домашнее задание
2	Основы нейронных сетей		2		14	Домашнее задание
3	Обработка естественного языка		4		14	Домашнее задание
4	Компьютерное зрение		4		14	Домашнее задание Контест
5	Обучение с подкреплением		4		14	Домашнее задание Контест
6	Машинное обучение на практике		2		12	Домашнее задание Контест
	Зачет			2		
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	Основы машинного обучения	Термины Продвинутые методы поиска оптимального решения Продвинутые методы оптимизации Продвинутые методы обучения с учителем Продвинутые методы обучения без учителя Особенности ансамблей Закрепление и проверка материала
2	Основы нейронных сетей	Нейрон
3	Обработка естественного языка	Токенизация текста Особенности рекуррентных нейронных сетей
4	Компьютерное зрение	Продвинутая обработка изображений Применение свёрток для решения задач компьютерного зрения
5	Обучение с подкреплением	Математические основы обучения с подкреплением Архитектуры нейронных сетей для решения задач обучения с подкреплением
6	Машинное обучение на практике	Машинное обучение на практике

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. Барретт, С. Ф. Arduino: искусственный интеллект и машинное обучение : практическое руководство / С. Ф. Барретт ; с англ. Ю. В. Ревича. – Москва : ДМК Пресс, 2024. - 244 с. – ISBN 978-5-93700-276-1. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2205065>.

2. Дейтел, П. Python: Искусственный интеллект, большие данные и облачные вычисления : практическое руководство / П. Дейтел, Х. Дейтел. - Санкт-Петербург : Питер, 2020. - 864 с. - (Серия «Для профессионалов»). - ISBN 978-5-4461-1432-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1733685>.

3. Постоит, А. В. Основы искусственного интеллекта в примерах на Python : самоучитель / А. В. Постоит. - Санкт-Петербург : БХВ-Петербург, 2021. - 448 с. - (Самоучитель). - ISBN 978-5-9775-6765-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2123395>.

Дополнительная литература:

1. Воронов, М. В. Системы искусственного интеллекта : учебник и практикум для вузов / М. В. Воронов, В. И. Пименов, И. А. Небаев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 268 с. — (Высшее образование). — ISBN 978-5-534-17032-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/567794>

2. Новиков, Ф. А. Символический искусственный интеллект: математические основы представления знаний : учебник для вузов / Ф. А. Новиков. — Москва : Издательство Юрайт, 2025. — 278 с. — (Высшее образование). — ISBN 978-5-534-00734-3. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/561410>.

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;
- механические калькуляторы;
- 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) «Введение в искусственный интеллект. Продвинутый уровень» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекция, семинары и домашние задания, контесты, 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.

Практические занятия — активная форма обучения, в рамках которой студенты применяют теоретические знания на практике, выполняя задачи, эксперименты или упражнения для закрепления навыков и умений. Они способствуют развитию практических компетенций, стимулируют самостоятельность и позволяют интегрировать знания из различных дисциплин в реальные сценарии решения проблем.

Для успешной подготовки к практическому занятию заранее изучите теоретический материал по теме, спланируйте этапы работы (подготовка, выполнение задач, обсуждение и рефлексия) и распределите время с учётом дедлайнов. Активно участвуйте в групповой работе, применяйте критическое мышление, а после занятия проанализируйте результаты и используйте полученную обратную связь для дальнейшего улучшения навыков.

Контекст – интерактивная платформа с заданиями разного уровня сложности и автоматической проверкой результатов.

Контекст позволяет оперативно оценивать усвоение материала и выявлять пробелы в знаниях через тесты и практические задачи. Такой формат способствует регулярной самопроверке и повышает мотивацию к изучению дисциплины.

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.

Бонусные баллы — это оценки, которые студенты могут получить за выполнение дополнительных заданий.

Формат бонусных баллов позволяет студентам улучшить общую оценку по курсу и стимулирует углубленное изучение материала.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Введение в искусственный интеллект. 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	Оценка за зачет	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
10	Зачтено	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
9	Зачтено	Отлично	
8	Зачтено	Отлично	

Десятибалльная оценка	Оценка за зачет	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
			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.
7	Зачтено	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	Зачтено	Хорошо	
5	Зачтено	Удовлетворительно	Студент обладает базовыми знаниями по дисциплине (модулю), 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
4	Зачтено	Удовлетворительно	

Десятибалльная оценка	Оценка за зачет	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
			минимальным набором методов исследования.
3	Не зачтено	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 зачтено	Не сдан	
1	Не зачтено	Не сдан	

Дисциплина (модуль) «Введение в искусственный интеллект. Продвинутый уровень» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	50%	11	Набор задач по темам недели
Контекст	30%	3	Интерактивные задания разного уровня сложности с автоматической проверкой результатов
Зачет	20%	1	Письменная или устная работа над заданием, направленным на проверку полученных знаний и навыков по курсу

Формула расчёта итоговой оценки по дисциплине (модулю) «Введение в искусственный интеллект. Продвинутый уровень»: $\langle 0,5 \times \text{среднее за домашние задания} + 0,3 \times \text{контекст} + 0,2 \times \text{зачет} \rangle$.

В рамках изучения дисциплины (модуля) возможно получение бонусных баллов.

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные задания для контекста

Контекст 1

Необходимые данные хранятся в таблице. Загрузи их в ноутбук.

1. Воспользуйся командой `gdown` для загрузки данных.
2. Считай их с помощью `Pandas`.
3. Выведи датасет на экран.
4. Исключи переменную `id` или сделай её индексом. Можешь выбрать любой вариант.

Контекст 2

Проведи первичный анализ данных.

Для этого:

1. Построй описательную статистику (ты уже сталкивался с такой задачей на семинаре).
2. Напиши свои выводы. Что необычного ты заметил?
3. Выведи информацию по типам данных каждой колонки.
4. Сделай описательную статистику для нечисловых переменных. Тебе поможет документация. Посмотри внимательно, какой тип данных надо передать в аргументы функции описательной статистики.
5. Напиши свои выводы. Какие интересные закономерности в данных ты заметил? Что необычного ты заметил? Напиши выводы минимум по двум переменным. Какие закономерности ты заметил? Напиши выводы минимум по двум переменным.

Контеcт 3

Построй графики, которые проверят базовые навыки владения Pandas и библиотеками визуализации.

1. Нарисуй гистограмму распределения переменной Depression.
2. Нарисуй две гистограммы рядом:
 - Слева: гистограмма продолжительности сна.
 - Справа: гистограмма продолжительности сна, где каждая категория должна отображаться с учётом переменной Depression.
 - Для каждой гистограммы порядок колонок должен быть упорядочен от наименьшей продолжительности сна до максимальной. Категорию Others помести в начало или в конец.
3. Каждый график должен быть подписан, легенда выведена. У гистограмм из пункта 2 должны быть как отдельные подписи, так и общая.
4. Опиши результаты для каждого графика и получившиеся выводы. Проверь, что у каждого графика есть подписи осей, легенды и названия.
Что ты заметил на графиках? Напиши минимум один вывод.

Примерные домашние задания

Домашнее задание 1.

Сегодня ты разберёшь задачу классификации для датасета Predict Pet Adoption Status Dataset. По характеристикам животного надо предсказать, заберут ли его из приюта.

Изучи переменные в данных:

PetType — тип животного (например, птица, кролик, собака).

Breed — порода.

AgeMonths — возраст в месяцах.

Color — цвет.

Size — размер.

WeightKg — вес.

Vaccinated — есть ли прививка (0 — нет, 1 — да).

HealthCondition — здоровое или больное животное (0 — здоровое, 1 — больное).

TimeInShelterDays — количество дней в приюте.

AdoptionFee — сколько стоит забрать животное из приюта.

PreviousOwner — был ли у животного уже хозяин (0 — нет, 1 — да).

AdoptionLikelihood — найдут ли хозяина животному (0 — низкая вероятность, 1 — высокая вероятность). Это целевая переменная.

Задача 1. Разделение датасета на train, val и test [4 балла]

Сначала необходимо разбить выборку на обучающую, валидационную и тестовую. Для этого нужно выбрать переменную, по которой будут сбалансированно разделяться данные. Чтобы решить эту задачу было проще, выбирай только категориальные переменные.

Что нужно сделать:

1. Визуализировать разделение на train и val по всем категориальным переменным. [1 балл]
2. Предложить разделение по некоторой категориальной переменной так, чтобы распределения по другим переменным были похожи между валидационной и обучающей выборкой. Обосновать своё решение, ориентируясь на визуализацию. [2 балла]
Важно. Помни, что целевая переменная тоже категориальная. Обязательно рассмотри ее в процессе анализа.
3. Сделать разбиение на train, val и test по выбранной переменной. [1 балл]

Как будет выглядеть твоя визуализация (вся сложная часть кода уже написана):

- Большое полотно с множеством гистограмм.
- Каждая строка полотна соответствует разбиению по одной из категориальных переменных.
- По столбцам каждой строки полотна изображены гистограммы переменных, по которым разбиение не делали. Изображать переменную, по которой делали разбиение, не имеет смысла, так как она будет максимально сбалансированной.
- Все оси подписаны, названия графиков и легенды выведены.
- `# Эти библиотеки тебе пригодятся`
- `import seaborn as sns`
-
- `import matplotlib.pyplot as plt`
- `from sklearn.model_selection import train_test_split`

Предложи разбиение, которое считаешь наиболее сбалансированным. Обоснуй, почему ты так считаешь.

Задача 2. ONE-обработка категорий и выбор алгоритма для обучения [3 балла]

После разделения данных на train, val и test нужно обучить модель машинного обучения. В этой задаче тебе нужно преобразовать категории к ONE и выбрать лучшую модель для задачи классификации.

Что нужно сделать:

1. Применить ONE из sklearn [1 балл]:
 - обучить его train-подвыборке;
 - применить результаты к val и test.
 2. Выбрать 4 различные модели для классификации из библиотеки sklearn (рекомендуется взять LogisticRegression, чтобы проще было решать четвертое задание на бонусные баллы). [0 баллов]
- Примечание.** Нормально, что ты до сих пор не понимаешь, что это за модели. С ними мы будем знакомиться в следующих занятиях. Сейчас важно показать, как ты будешь выстраивать пайплайн обучения, для этого сама модель и качество не важны.
3. Обучить каждую, получить предсказания для train- и val-подвыборок. [1 балл]
 4. Посчитать метрику [Accuracy](#) для train и val. [0,5 балла]
 5. На основе валидации выбрать модель для дальнейшей работы, написать свои мысли по поводу переобучения. [0,5 балла]

Метрика Accuracy имеет вид:

$$Acc = \frac{1}{n} \sum_{i=1}^n I[y_i = \hat{y}_i], \text{ где } n$$

— количество наблюдений; y_i — истинная метка класса для i -

го наблюдения; $\hat{y}_i$ — предсказанная метка класса для i -го наблюдения.

Задача 3. Подбор параметров для модели и проверка результатов на test [3 балла]

Теперь нужно подобрать параметры выбранной модели, как мы это делали на семинаре.

Что нужно сделать:

1. Посмотреть документацию выбранной модели и выбрать два параметра для перебора. [0 баллов]
- Примечание.** Тебе не надо знать, что они значат. Попробуй разные варианты разных параметров.
2. Запустить по ним цикл в цикле и выбрать лучшее сочетание. Лучшую модель, её параметры и результаты метрики для train и val проще хранить в отдельных переменных. [1 балл]
 3. Получить предсказание для test-подвыборки и оценить его качество по метрике. [1 балл]

4. Описать полученные результаты. [1 балл]

Важно. Так как теперь нужно следить за метрикой, а не за функцией потерь, не стоит бездумно копировать код с семинара. Для корректной работы надо его изменить.

Домашнее задание 2.

В этом домашнем задании тебе нужно:

- реализовать градиентный спуск для линейной регрессии с двумя параметрами для оптимизации: смещением и наклоном;
- разобраться, что такое шедюлер, и реализовать его для алгоритма градиентного спуска;
- заняться оптимизацией более сложной функции, чем на семинаре.

Задача 1. Добавление параметра сдвига к линейной регрессии [4 балла]

На семинаре была рассмотрена линейная регрессия для случая, когда есть всего один обучаемый параметр наклона. Сейчас требуется усложнить модель и добавить в неё параметр сдвига.

Тебе нужно:

1. Написать формулу вычисления градиента для параметра сдвига и наклона.
Примечание. Это можно сделать в формате LaTeX. Все необходимые части кода уже были в семинаре. Если трудно использовать LaTeX, напиши своё решение на бумаге и прикрепи его к ноутбуку. [1 балл]
2. Реализовать функцию для вычисления градиента для параметров сдвига и наклона для MSE. [1 балл]
3. Реализовать градиентный спуск для двух параметров сразу. [1,5 балла]
4. Сделать пошаговую визуализацию процесса оптимизации. [0,5 балла]

Чтобы решить задачу, заполни пропуски в формулах и коде.

Задача 2. Шедюлер и оптимизация сложной функции [5 баллов]

Мы уже пробовали разный темп обучения (англ. learning rate). Если он был очень большой, то оптимизатор быстро начинал сходиться в сторону минимума, но потом где-то застревал, не доходя до него. Если мы брали маленький learning rate, то алгоритм сходился медленно с самого старта. В этом случае мы не ждали, чтобы не тратить время на семинаре. Ждать пришлось бы долго, но результат был бы не хуже, чем для варианта с большим learning rate.

Отсюда следует идея шедюлера. Почему бы нам не сделать большой learning rate вначале, а потом постепенно уменьшать его по мере обучения.

Шедюлер (от англ. scheduler) — это несложная функция, которая по некоторому расписанию меняет learning rate для нашего алгоритма оптимизации. Иногда шедюлер называют «планировщик».

В этой задаче тебе нужно оптимизировать одну сложную функцию с множеством локальных минимумов и одним глобальным. Обходить минимумы ты будешь с помощью косинусного шедюлера. То есть скорость обучения будет то падать, то расти. Почему так?

Вспомним семинар: при большом learning rate значение нашей функции потерь росло. То есть в данной задаче при попадании в локальный минимум у оптимизатора будет шанс выбраться из него и исследовать другие минимумы.

Тебе нужно:

1. Визуализировать функцию, реализовать её вычисление и подсчитать градиенты. [1 балл]
2. Запустить из случайной точки градиентный спуск и описать, куда он сойдётся. [0,5 балла]
3. Реализовать косинусный шедюлер. [1 балл]
4. Визуализировать, как работает шедюлер — подать в него последовательно аргументы и сделать список из lr, который затем надо нарисовать на графике. [0,5 балла]

5. Запустить оптимизатор с косинусным шедулером. В результате оптимизатор должен обойти хотя бы несколько локальных минимумов (здорово, если он свалится в глобальный по итогу, но это необязательно). **[1,5 балла]**
6. Нарисовать маршрут оптимизатора с шедулером и написать свои выводы. **[0,5 балла]**

Задача 3. Перезапуск оптимизатора [1 балл + 2 бонусных балла]

Ещё один способ поиска глобального минимума среди локальных — это перезапуск оптимизатора с разных точек функции потерь.

Изначально мы могли перебирать точку `init` по сетке, а потом просто выбрать лучшую траекторию, но в реальной ситуации это было бы долго, ведь пространство для перебора точек по сетке крайне велико.

Чтобы как-то сократить пространство для перебора, мы один раз применили шедулер, который помог нам исследовать пространство нашей функции.

Тебе нужно:

1. Извлечь 20 точек с минимальным значением функции потерь. **[1 балл]**
2. Из каждой точки запустить обычный градиентный спуск без шедулера (это нужно, чтобы не вылететь из оптимума). **[1 балл]**
3. Нарисовать одну лучшую траекторию. **[1 балл]**

Домашнее задание 3.

Задача 1. Обучение разных моделей регрессии [6 баллов]

Тебе нужно обучить ML-модели и объяснить, почему получились такие предсказания. Не забывай, что модели обучены на подвыборке `train`, поэтому особое внимание удели интерпретации работы алгоритмов на `test`-подвыборке.

Используй модели:

1. Линейная регрессия.
2. SVM.
3. kNN.
4. Дерево решений.
5. Случайный лес.
6. Бустинг над деревьями.

Для каждой модели нужно выполнить следующие пункты:

1. Обучить модель регрессии на `train`-подвыборке. **[0,25 балла]**
2. Визуализировать регрессию и объяснить, почему получился такой результат. **[0,25 балла]**
3. Для каждого алгоритма один раз изменить параметры или, например, добавить полиномы, чтобы улучшить результат решения задачи. **[0,25 балла]**
4. Описать, что получилось. Если качество решения не удаётся улучшить простыми методами, объясни, с чем это связано. **[0,25 балла]**

Важно. Для линейной регрессии пункт 1 уже выполнен выше как пример, поэтому баллы за этот пункт не начисляются. Начинай сразу с пункта 2. За построение всех моделей, визуализацию и доработку ты можешь получить $0,75 + 1 + 1 + 1 + 1 + 1 = 5,75$ балла.

В конце напиши выводы о преимуществах и недостатках каждой модели. **[0,25 балла]**

Задача 2. Ансамблирование [4 балла]

Чтобы частично решить эту проблему прямо сейчас, можно построить ансамбль моделей. Для этого нужно взять несколько моделей и усреднить их предсказания. Можно взять не среднее, а с некоторыми весами, которые подбираются ровно так же, как и гиперпараметры моделей.

Тебе нужно:

1. Посмотреть документацию [VotingRegressor](#). **[0 баллов]**

2. Создать модель VotingRegressor, состоящую из трёх используемых ранее моделей [1 балл], при этом:
 - включить линейную регрессию с кубическим полиномом;
 - включить деревянную модель или kNN;
 - добавить третью модель на своё усмотрение.
3. Описать, что получилось. [1 балл]
4. Изменить весовую схему в VotingRegressor в пользу линейной регрессии. [1 балл]
5. Описать, что получилось. [1 балл]

Примечание. Если включать некачественно работающую модель в ансамбль (деревянные модели или kNN), то совокупный результат будет хуже, чем для линейной регрессии с полиномом третьей степени. Но смысл этой задачи — показать, как простыми методами можно делать ансамбль моделей и что часто ансамбль позволяет улучшить решение.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция	Индикатор
1	Назовите критерий достоверности источников при изучении алгоритмов градиентного бустинга	Рецензируемые конференции KDD NeurIPS ICML	УК-1	УК-1.2
2	Укажите инструмент для анализа техник токенизации текста из различных материалов	Библиотеки Hugging Face Transformers или NLTK	УК-1	УК-1.3
3	Определите подход к системному анализу архитектур CNN для задачи классификации	Сравнение метрик ассигасы и количества параметров FLOPs	УК-1	УК-1.2
4	Сформулируйте задачу поиска инструментов для отслеживания экспериментов машинного обучения	Использование MLflow или Weights and Biases	УК-1	УК-1.3
5	Выберите метод оптимизации при ограничениях на выпуклость функции потерь	Градиентный спуск или стохастический градиентный спуск	УК-2	УК-2.2
6	Укажите ограничение при выборе алгоритма обучения с подкреплением для непрерывного пространства действий	Необходимость использования Policy Gradient методов	УК-2	УК-2.2
7	Назовите критерий целесообразности использования ансамблей моделей в продакшене	Соотношение прироста точности и времени инференса	УК-2	УК-2.3
8	Определите круг задач для квантования модели в соответствии с ограничениями по памяти	Развёртывание на мобильных или периферийных устройствах	УК-2	УК-2.3
9	Дополните утверждение: метод обратного распространения ошибки вычисляет градиент по отношению к	Весам и смещениям нейронной сети	ОПК-1	ОПК-1.2
10	Назовите принцип обновления весов в методе градиентного спуска	Движение в направлении антиградиента функции потерь	ОПК-1	ОПК-1.2

11	Укажите метрику для оценки качества свёрточной нейронной сети на несбалансированных данных	F1 мера или ROC AUC	ОПК-1	ОПК-1.1
12	Определите критерий сходимости политики агента в обучении с подкреплением	Стабильность функции ценности или награды на эпизод	ОПК-1	ОПК-1.3
13	Выберите алгоритм обучения с учителем для задачи бинарной классификации	Логистическая регрессия или метод опорных векторов	МО-1	МО-1.1
14	Назовите архитектуру рекуррентной сети для обработки длинных последовательностей текста	LSTM или GRU	МО-1	МО-1.2
15	Укажите показатель для улучшения разнообразия базовых моделей в ансамбле	Снижение корреляции ошибок между моделями	МО-1	МО-1.1
16	Определите метод использования предварительно обученных сетей в компьютерном зрении	Transfer learning или дообучение классификатора	МО-1	МО-1.2
17	Назовите структуру данных для представления изображения в свёрточной сети	Тензор размерности высота ширина каналы	ППК-Р1	ППК-Р1.2
18	Укажите объект для формализации задачи векторного представления слов	Матрица эмбедингов или словарь токенов	ППК-Р1	ППК-Р1.2
19	Дополните требование к реализации среды для обучения агента с подкреплением	Интерфейс gym step и reset методы	ППК-Р1	ППК-Р1.2
20	Назовите библиотеку Python для автоматической настройки гиперпараметров модели	Optuna или Hyperopt	ППК-Р1	ППК-Р1.3