ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«24» июня 2025 г.
Протокол №2

Рабочая программа дисциплины (модуля)
«Введение в искусственный интеллект»

Направление подготовки: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Математика и искусственный интеллект*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: 2025

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0215B0C000D1B3FEB44A37EE631E0C27D4
Владелец: Ивашкевич Евгений Васильевич
Действителен: с 13.01.2026 до 13.04.2027

Электронный документ

Москва
2025

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Перечень планируемых результатов обучения** 5
3. **Тематический план** 5
4. **Содержание дисциплины (модуля)** 8
5. **Учебно-методическое обеспечение** 9
6. **Материально-техническое обеспечение** 9
7. **Методические и оценочные материалы** 11

1. Краткая характеристика дисциплины (модуля)

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

Изучение дисциплины (модуля) Введение в искусственный интеллект (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 Математика и искусственный интеллект и входит в обязательную часть Блока 1.

Дисциплина (модуль) изучается на 1 курсе во 2 семестре. Доступна после успешного освоения дисциплины (модуля) «Введение в статистику».

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:

— основную терминологию машинного обучения: модель, датасет, обучающая и тестовая выборки, метрика, функция потерь, оптимизация;

— доменные области машинного обучения: компьютерное зрение, задачи ранжирования, работа со звуком, обработка естественного языка, задачи на графах;

— задачи машинного обучения: сегментация, детекция, классификация, регрессия, машинный перевод, генерация изображений, распознавание речи;

— общее устройство архитектур ряда моделей: линейная регрессия, решающее дерево, ансамбль, нейросеть, генеративная нейросеть, трансформер;

— инструменты для решения бытовых задач с помощью искусственного интеллекта;

— вопросы этики, связанные с искусственным интеллектом;

уметь:

— запускать модель через внешние интерфейсы, без глубокого понимания процессов обучения, и базово анализировать результаты обучения;

— использовать инструменты искусственного интеллекта в повседневной жизни;

владеть:

— пониманием и применением базовой терминологии машинного обучения;

— ориентированием в ключевых доменных областях машинного обучения и знает примеры их применения;

— распознаванием различий основных типов задач машинного обучения и умением подобрать примеры для каждой;

— представлением об архитектуре распространённых моделей и их назначении;

— применением готовых инструментов искусственного интеллекта для решения бытовых и учебно-профессиональных задач;

— запуском и тестированием модели через внешние интерфейсы, интерпретацией базовых результатов;

— осознанием этических вопросов, связанных с разработкой и применением искусственного интеллекта.

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:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
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 концепции и теории в области математического анализа и смежных дисциплин; методы и подходы, используемые в различных областях математики
		ОПК-1.2.	Умеет применять математические методы для решения профессиональных задач
		ОПК-1.3.	Имеет практический опыт разработки и реализация математических моделей в профессиональной деятельности
ОПК-2.	Способен проводить под научным руководством исследование на основе существующих методов в конкретной области профессиональной деятельности	ОПК-2.1.	Знает основные методы и подходы к проведению исследований в области профессиональной деятельности; научные и этические стандарты, применяемые в исследовательской практике
		ОПК-2.2.	Умеет формулировать исследовательские вопросы и гипотезы на основе существующих знаний; анализировать и интерпретировать данные, полученные в ходе исследования
		ОПК-2.3.	Имеет практический опыт участия в исследовательских проектах под научным руководством; применения существующих методов для решения практических задач в рамках исследовательской деятельности в области профессиональной деятельности
ОПК-4.	Способен находить, анализировать, реализовывать программно и использовать на практике математические	ОПК-4.1.	Знает базовые основы современного математического аппарата, связанного с проектированием, разработкой,

	алгоритмы, в том числе с применением современных вычислительных систем		реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности.
		ОПК-4.2.	Умеет использовать этот математический аппарат в профессиональной деятельности.
		ОПК-4.3.	Имеет практический опыт применения современного математического аппарата, связанного с проектированием, разработкой, реализацией и оценкой качества программных продуктов и программных комплексов в различных областях человеческой деятельности.
ОПК-5.	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-5.1.	Знаете технологии, необходимые для прикладного программирования, включая современные функциональные языки программирования, а также основные принципы и понятия, применяемыми при использовании компьютерных сетей
		ОПК-5.2.	Умеет пользоваться технологиями прикладного программирования, включая среды высокоуровневого программирования
		ОПК-5.3.	Имеет практический опыт использования технологий прикладного программирования
ПК-3.	Способен применять методы математического и алгоритмического моделирования для решения как теоретических, так и практических задач в рамках профессиональной деятельности	ПК-3.1.	Знает основные методы математического и алгоритмического моделирования, а также их применение для решения теоретических и прикладных задач
		ПК-3.2.	Умеет применять методы математического и алгоритмического моделирования для анализа и решения различных задач в области математики и компьютерных наук
		ПК-3.3.	Имеет опыт использования методов математического и алгоритмического моделирования при решении теоретических и прикладных задач в профессиональной

			деятельности

3. Тематический план

№п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Контактная работа		Контроль	Самостоятельная работа	
		Лекции	Семинары			
1	Основы машинного обучения	6	6		20	Домашнее задание Коллоквиум
2	Обработка естественного языка	6	6		21	Домашнее задание
3	Компьютерное зрение	6	6		21	Домашнее задание Кейс
4	Обучение с подкреплением	6	6		21	Домашнее задание Кейс
5	Машинное обучение на практике	6	6		21	Домашнее задание Кейс
	Зачет с оценкой			6		
	Итого:	***30***	***30***	***6***	***124***	
Объем дисциплины (модуля) (в ак. ч.)		***190***				
Объем дисциплины (модуля) (в зач. ед.)		***5***				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Основы машинного обучения	Термины машинного обучения. Оптимизация. Классическое машинное обучение. Нейронные сети 1. Нейронные сети 2
2	Обработка естественного языка	Токенизация текста и статистические методы. Векторизация текста и Рекуррентные нейронные сети. Механизм внимания
3	Компьютерное зрение	Обработка изображений. Механизм свёртки. Свёрточные нейронные сети
4	Обучение с подкреплением	Марковский процесс принятия решений. Безградиентные методы. Q-learning
5	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. Постолит, А. В. Основы искусственного интеллекта в примерах на Python : самоучитель / А. В. Постолит. - 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» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекция, семинары и домашние задания, кейс, проект, 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 удается, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Участие в семинаре (аудиторная работа) – активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их.

При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

Кейс – практическая работа студентов над реальными или смоделированными задачами, что позволяет студенту применять теоретические знания на практике.

Студент самостоятельно разрабатывает стратегию решения поставленной задачи, что способствует развитию навыков критического мышления и самостоятельного принятия решений. Такой подход помогает подготовить будущих специалистов к реальным вызовам в их профессиональной деятельности.

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.

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»

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 с оценкой***, 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:

Десятибалльная оценка	Оценка за зачет	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
10	Зачтено	Отлично	Студент полностью владеет знаниями,

Десятибалльная оценка	Оценка за зачет	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
9	Зачтено	Отлично	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.
8	Зачтено	Отлично	
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
4	Зачтено	Удовлетворительно	

Десятибалльная оценка	Оценка за зачет	Пятибалльная оценка	Общая характеристика результата обучения по дисциплине (модулю)
			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.
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» оценивается следующим образом:

Активность	Вес	Количество	Описание
Домашние задания	40%	13	Набор задач по темам недели
Коллоквиум	15%	1	Устные ответы на вопросы, список которых известен студенту заранее
Кейс	30%	3	Практическая работа студентов над реальными или смоделированными задачами
Зачет с оценкой	15%	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»: «0,4 × среднее за домашние задания + 0,15×коллоквиум + 0,3×кейс + 0,15× зачет с оценкой».

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

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные задания по коллоквиуму

Формат проведения

Цель коллоквиума — проверить твои знания по пройденным материалам, навыки ориентироваться в темах и вести конструктивную беседу по программе курса. Коллоквиум будет проводиться в формате диалога, похожего на собеседование.

Все студенты жеребьёвкой разделятся на пары. В каждой паре сначала один студент отвечает на вопросы другого, а затем наоборот: второй студент отвечает на вопросы первого. Вопросы должны опираться на билет. Каждый из билетов будет содержать 2 базовых вопроса из списка и один кейс, в котором нужно представить себя в роли ML-специалиста известной компании. К каждому вопросу из билета (включая кейс) студент самостоятельно придумывает несколько дополнительных вопросов, чтобы развить диалог по теме. За ходом беседы будет наблюдать проверяющий, который может задавать уточняющие вопросы.

Список базовых вопросов

1. Чем различаются термины «Машинное обучение» и «Искусственный интеллект»?
2. Чем различаются теория вероятностей и математическая статистика?
3. Что такое EDA? Зачем он нужен?
4. Что такое таргет? Чем таргет отличается от других признаков? Что такое разметка датасета?
5. Чем обучение с учителем отличается от обучения без учителя?
6. Какие задачи машинного обучения ты знаешь? Чем они различаются?
7. Какие функции ошибки ты знаешь? В каких задачах они используются?
8. Что такое градиент функции?
9. Опиши алгоритм градиентного спуска.
10. Какие проблемы могут возникнуть при применении градиентного спуска?
11. Что такое переобучение?
12. Что такое гиперпараметр? Приведи 2–3 примера гиперпараметров.
13. Чем валидационная выборка отличается от тестовой?
14. Что такое бутстрэп?
15. Что такое скользящий контроль? Зачем он нужен?
16. Как увеличение обучающей выборки (и соответственное уменьшение тестовой и валидационной выборок) повлияет на процесс машинного обучения? А как повлияет уменьшение обучающей выборки?
17. Чем линейная регрессия отличается от логистической регрессии?
18. Какие модели называются метрическими? Приведи примеры метрических моделей.
19. Какие гиперпараметры существуют у решающего дерева? За что они отвечают?
20. Какие техники ансамблирования ты знаешь? В чём заключается их суть?
21. Какие свойства вероятности ты знаешь?
22. Какие события называются независимыми?
23. Что такое парадокс Монти Холла? Объясни, в чём состоит его решение.
24. Что такое распределение случайной величины? Чем дискретное распределение отличается от непрерывного? Приведи примеры распределений.
25. Что такое функция правдоподобия? Какая у неё формула? Где используется эта функция?
26. Что такое нейрон?
27. Что такое функция активации? Зачем она нужна?
28. Что такое многослойный перцептрон?
29. Что такое метод обратного распространения ошибки? Опиши его этапы.
30. Какие функции активации ты знаешь? В чём их недостатки и преимущества?

Пример билета

1. Чем линейная регрессия отличается от логистической регрессии?
   Дополнительные вопросы:
   - А какую задачу решает логистическая регрессия?
   - Какие значения может выдавать каждая из этих моделей в качестве ответа?

2. Какие функции ошибки ты знаешь? В каких задачах они используются?
   Дополнительные вопросы:
   - Является ли названная функция ошибки дифференцируемой? Почему это важно?
   - Какие функции ошибки для задачи многомерной регрессии ты знаешь?

3. Представь, что тебе поручили создание ML-алгоритмов и процессов обработки данных в Яндекс Еде. Подумай, как можно подойти к этой задаче?
   Дополнительные вопросы:
   - Как можно собирать данные?

- Как можно сформулировать задачу в терминах ML?
- Какие модели ты мог бы использовать?

ВАЖНО

Дополнительные вопросы не фиксированы. Они будут появляться в ходе диалога и могут отличаться от тех, что указаны выше.

Примерные домашние задания

Домашнее задание по теме «Табличные данные и визуализация»

Задача 1 [2 балла]

Необходимые данные хранятся в таблице. Загрузи их в ноутбук.

1. Воспользуйся командой gdown для загрузки данных. **[0,5 балла]**
2. Считай их с помощью Pandas. **[0,5 балла]**
3. Выведи датасет на экран. **[0,5 балла]**
4. Исключи переменную id или сделай её индексом. Можешь выбрать любой вариант. **[0,5 балла]**

Задача 2 [3 балла]

Проведи первичный анализ данных.

Для этого:

1. Построй описательную статистику (ты уже сталкивался с такой задачей на семинаре). **[0,5 балла]**
2. Напиши свои выводы. Что необычного ты заметил? **[0,5 балла]**
3. Выведи информацию по типам данных каждой колонки. **[0,5 балла]**
4. Сделай описательную статистику для нечисловых переменных. Тебе поможет документация. Посмотри внимательно, какой тип данных надо передать в аргументы функции описательной статистики. **[1 балл]**
5. Напиши свои выводы. Какие интересные закономерности в данных ты заметил? **[0,5 балла]**

   Что необычного ты заметил? Напиши выводы минимум по двум переменным.

   Какие закономерности ты заметил? Напиши выводы минимум по двум переменным.

Задача 3 [4 балла]

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

1. Нарисуй гистограмму распределения переменной Depression. **[0,5 балла]**
2. Нарисуй две гистограммы рядом: **[1,5 балла]**
   - Слева: гистограмма продолжительности сна.
   - Справа: гистограма 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 **[0,5 балла]**. У гистограмм из пункта 2 должны быть как отдельные подписи, так и общая. **[0,5 балла]**
4. Опиши результаты для каждого графика и получившиеся выводы. **[0,5 + 0,5 балла]**

   Проверь, что у каждого графика есть подписи осей, легенды и названия.

   Что ты заметил на графиках? Напиши минимум один вывод.

Задача 4 [1 балл и 2 бонусных балла]

Если ты сделал правильные выводы из задачи 3, то сегодня точно захочешь лечь пораньше. Теперь нужно проверить, получают ли студенты больше баллов, если вместо того, чтобы идти спать вовремя, продолжают решать домашнее задание.

Для этого тебе нужно:

1. Ознакомиться с документацией boxplot. **[0 баллов]**
2. Нарисовать диаграмы вида boxplot: **[1 балл + 1 бонусный балл]**
   - Это должен быть один график с множеством boxplot.
   - Для каждой группы сна нарисуй отдельный boxplot. **[0,5 балла]**
   - Группы сна должны быть упорядочены, как в задаче 3. **[0,5 балла]**
   - Также сделай разделение по переменной Depression. **[0,5 бонусных балла]**
   - Оси должны быть подписаны, выведено название и легенда. **[0,5 бонусных балла]**
3. Какие выводы можно сделать по графику? **[1 бонусный балл]**

Обязательно ознакомься с документацией, иначе график не построить.

Напиши, что можно увидеть на графике (минимум два вывода).

Домашнее задание по теме «Машинное обучение»

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. OHE-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 OHE и выбрать лучшую модель для задачи классификации.
Что нужно сделать:

1. Применить OHE 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 четвёртое 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:

$$Acc=1n\sum_{ni=1}I[y_i=y_i\hat{}],\text{где } n$$

— количество наблюдений; y_i — истинная метка класса для i-го наблюдения; $y_i\hat{}$ — 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 по теме «Оптимизация и табличные данные»

В этом домашнем задании тебе нужно:

- реализовать градиентный спуск для линейной регрессии с двумя параметрами для оптимизации: смещением и наклоном;
- разобраться, что такое шедулер, 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. Шедулер 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 шедулера. Почему бы нам не сделать большой learning rate вначале, а потом постепенно уменьшать его по мере обучения.

Шедулер (от англ. scheduler) — это несложная функция, которая по некоторому расписанию меняет learning rate для нашего алгоритма оптимизации. Иногда шедулер 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 шедулера. 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 градиентый спуск и описать, куда он сойдётся. **[0,5 балла]**
3. Реализовать косинусный шедулер. **[1 балл]**
4. Визуализировать, как работает шедулер — 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 по теме «Классические модели ML»

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 регресии 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 докумментацию VotingRegressor. **[0 баллов]**
2. Создать модель VotingRegressor, состоящую из трёх используемых ранее моделей **[1 балл]**, при этом:
   - включить линейную регерессию 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
1.	Как называется набор данных для обучения, в котором каждый объект описан некоторыми признаками и таргетом? ***Напиши ответ одни словом на русском языке***	Датасет / датасет / dataset / Dataset / Data set / data set / Data Set	ОПК-5
2.	Верно ли то, что градиентный спуск – это основной алгоритм оптимизации моделей, использующий частные производные функции потерь и минимизирующий её значение?	ДА/Да/да	ОПК-4
3.	_____ выборка - выборка данных, которая используется для подбора гиперпараметров. ***Напиши ответ одни словом***	Валидационная / валидационная	ОПК-4
4.	_____ выборка - выборка данных, которая используется для итоговой проверки качества модели. ***Напиши ответ одни словом***	Тестовая/тестовая/Тестовая выборка/тестовая выборка	ОПК-4
5.	Слой в нейронной сети, который является не входным, и не выходным, называется _____ слоем.	Скрытым/скрытым	ОПК-5
6.	Чему равно значение выражения: ReLU(3)?	3	ОПК-5
7.	Как называется небольшой набор объектов, который передается на вход нейросети для обучения? ***Напиши ответ одни словом на русском языке***	Батч / батч / batch / Batch	ОПК-5
8.	Верно ли то, что свёртка – это операция, при которой к картинке применяется фильтр, который поэлементно умножается на каждый участок того же размера?	Да/да	ОПК-5
9.	Как называется семейство моделей, которые содержат внутри себя множество других моделей? ***В ответе укажи одно слово во множественном числе***	Ансамбли/ансамбли	ОПК-5
10.	Как в задаче обработки текста называется единица текста?	Токен/токен	ОПК-5
11.	Назовите ключевой термин ML, обозначающий меру расхождения между предсказанной вероятностной моделью и реальными данными в вероятностном обучении.	Кульбак-Лейблер дивергенция	ОПК-1
12.	В табличных данных для классических моделей ML укажите статистическую метрику, оценивающую качество модели на основе доли правильных предсказаний.	Точность (accuracy)	ОПК-2
13.	В истории ИИ укажите период "зимы ИИ" (1970-е–1980-е), когда финансирование сократилось из-за	Вторая зима ИИ	ПК-3

	ограничений статистических моделей.		
14.	В задачах с изображениями укажите статистическую модель, используемую в CNN для вероятностной классификации пикселей (например, softmax).	Вероятностное распределение	ОПК-1
15.	При анализе задач с текстами в ИИ, какой статистический подход применяется для оценки семантической близости слов на основе вероятностей? a) Word2Vec с embedding b) TF-IDF для частот c) Графовые модели PageRank	b	ОПК-2
16.	В задачах с сигналами и речью укажите метрику, измеряющую статистическую ошибку предсказания фонем (word error rate).	Ошибка распознавания слов	ОПК-2
17.	Для задач с текстами укажите вероятностную модель, предсказывающую следующее слово на основе n-грамм.	N-gram модель	ПК-3
18.	В этике ИИ укажите статистическую проблему, когда модель показывает предвзятость из-за несбалансированных данных (bias).	Статистическая предвзятость	ОПК-1
19.	В этике ИИ назовите принцип, требующий статистической прозрачности моделей (explainable AI).	Объяснимость ИИ	ПК-3
20.	Как инструмент ИИ укажите платформу для этического тестирования моделей на предвзятость (fairness checks).	Hugging Face	ОПК-5