

Минобрнауки России

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)**



УТВЕРЖДАЮ
Заведующий кафедрой
Сирота Александр Анатольевич
Кафедра технологий обработки и защиты информации
23.04.2024

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.ДВ.05.01 Системы машинного обучения

1. Код и наименование направления подготовки/специальности:

09.03.02 Информационные системы и технологии

2. Профиль подготовки/специализация:

Инженерия информационных систем и технологий

3. Квалификация (степень) выпускника:

Бакалавриат

4. Форма обучения:

Очная

5. Кафедра, отвечающая за реализацию дисциплины:

Кафедра технологий обработки и защиты информации

6. Составители программы:

Отырба Ростислав Русланович, ассистент

7. Рекомендована:

протокол №5 от 03.05.2024

8. Учебный год:

2028-2029

9. Цели и задачи учебной дисциплины:

Целью данной учебной дисциплины является изучение теоретических основ машинного обучения и приобретение практических навыков разработки и применения современных методов и алгоритмов машинного обучения для решения реальных задач в различных областях.

Основные задачи дисциплины:

- формирование базовых знаний в области машинного обучения;
- изучение основных алгоритмов и инструментов машинного обучения и особенностей их применения при разработке автоматизированных алгоритмов обработки данных;
- овладение практическими навыками разработки, обучения и тестирования моделей машинного обучения.
- формирование навыков проведения научных исследований в области машинного обучения, анализа данных и разработки новых алгоритмов.
- развитие способности критически оценивать существующие методы и алгоритмы, а также выбирать и адаптировать их под конкретные задачи.

10. Место учебной дисциплины в структуре ООП:

Относится к вариативной части профессионального цикла дисциплин по выбору учебного плана.

Для успешного освоения дисциплины необходимы входные знания в области программирования, информатики, математического анализа, теории вероятности, математической статистики, технологий обработки информации.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников) и индикаторами их достижения:

Код и название компетенции	Код и название индикатора компетенции	Знания, умения, навыки
ПК-1 Способен проводить исследования на всех этапах жизненного цикла программных средств	ПК-1.2 Знает методы проведения экспериментов и наблюдений, обобщения и обработки информации	<p>Студент должен:</p> <p>Знать методы и технологии организации и проведения статистического компьютерного эксперимента в приложении к алгоритмам машинного обучения, основы функционирования и обучения основных методов машинного обучения (классические методы и методы глубокого обучения) для формирования умения обоснования выбора того или иного метода при проведении экспериментов, знает метрики оценки качества моделей машинного обучения.</p> <p>Уметь ставить цели исследования, формулировать гипотезы, выбирать методы исследования, собирать и анализировать данные, обоснованно выбирать алгоритмы машинного обучения для решения поставленных задач, настраивать их гиперпараметры, а также реализовывать и проектировать модели самостоятельно. Подготавливать и предварительно обрабатывать наборы данных и анализировать их с помощью статистических методов. Проводить обучение и тестирование модели. Осуществлять сравнительный анализ различных подходов. Анализировать результаты модельного эксперимента, сопоставляя получаемые и планируемые результаты.</p> <p>Владеть практическими навыками</p>

		<p>использования современных Python библиотек и математических пакетов для анализа, подготовки, обработки данных и проведения компьютерных экспериментов по оценке эффективности разработанной модели машинного обучения.</p>
<p>ПК-1 Способен проводить исследования на всех этапах жизненного цикла программных средств</p>	<p>ПК-1.4 Использует стандартное и оригинальное программное обеспечение и проводит компьютерный эксперимент, составляет его описание и формулирует выводы</p>	<p>Студент должен:</p> <p>Уметь разрабатывать системы машинного обучения с использованием различных сред и инструментов, применять библиотеки и фреймворки машинного обучения для решения задач, визуализировать данные и результаты экспериментов с помощью различных инструментов, грамотно оформлять отчеты о компьютерных экспериментах и презентовать результаты исследований.</p> <p>Владеть практическими навыками использования современных Python библиотек и математических пакетов для анализа, подготовки, обработки данных и проведения компьютерных экспериментов, а также навыками формирования результирующих отчетов с сравнительным анализом эффективности различных подходов и гиперпараметров моделей.</p>

12. Объем дисциплины в зачетных единицах/час:

3/108

Форма промежуточной аттестации:

Зачет с оценкой, Контрольная работа

13. Трудоемкость по видам учебной работы

Вид учебной работы	Семестр 8	Всего
Аудиторные занятия	50	50
Лекционные занятия	34	34
Практические занятия	0	0
Лабораторные занятия	16	16
Самостоятельная работа	22	22
Курсовая работа	0	0
Промежуточная аттестация	0	0
Часы на контроль	0	0
Всего	72	72

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
1	Введение в машинное обучение.	Лекции по разделу: 1. Искусственный интеллект и машинное обучение. 2. Обучение с учителем. Задача классификация и регрессия. Лабораторные работы по разделу: Нет	Размещены индивидуальные задания для выполнения лабораторных работ.
2	Задача регрессии. Линейная регрессия	Лекции по разделу 3. Задача регрессии. Линейная регрессия. Функция ошибки. Производная и градиентный спуск. Нелинейные виды регрессии. Стандартизация и нормализация данных. Лабораторные работы по разделу: 1. Задача регрессии. Построение и обучение регрессионной модели.	Размещены индивидуальные задания для выполнения лабораторных работ.

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
3	Задача классификации. Логистическая регрессия.	<p>Лекции по разделу</p> <p>4. Задача бинарной и многоклассовой классификации. Логистическая регрессия.</p> <p>5. Мультиномиальная регрессия.</p> <p>Лабораторные работы по разделу:</p> <p>2. Задача классификации. Построение и обучение логистической и мультиномиальной регрессии.</p>	Размещены индивидуальные задания для выполнения лабораторных работ.
4	Разведочный анализ данных.	<p>Лекции по разделу</p> <p>6. Базовые концепции и этапы разведочного анализа данных.</p> <p>Лабораторные работы по разделу:</p> <p>3. Разведочный анализ данных. Обучение и тестирование моделей с использованием Scikit-learn.</p>	Размещены индивидуальные задания для выполнения лабораторных работ.
5	Дерево решений.	<p>Лекции по разделу</p> <p>7. Архитектура дерева решений. Дерево решений в задаче классификации и регрессии. Техники предотвращения переобучения дерева</p> <p>Лабораторные работы по разделу:</p> <p>4. Построение дерева решений с нуля для задач классификации и регрессии.</p>	Размещены индивидуальные задания для выполнения лабораторных работ.

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
6	Ансамблевые методы.	<p>Лекции по разделу</p> <p>8. Одиночная модель vs Ансамбль моделей. Стратегии ансамблирования: Бэггинг, Бустинг, Стэкинг.</p> <p>Лабораторные работы по разделу:</p> <p>5. Реализация стратегии ансамблирования Бэггинг</p>	Размещены индивидуальные задания для выполнения лабораторных работ.
7	Метод k-ближайших соседей.	<p>Лекции по разделу</p> <p>9. Обычный метод k-ближайших соседей. Взвешенный метод k-ближайших соседей. Принцип работы метода в задачах классификации и регрессии.</p> <p>10. Проблема проклятия размерности признаков. Метод главных компонент. Принцип работы метода.</p> <p>Лабораторные работы по разделу:</p> <p>6. KNN в задаче классификации изображений</p>	
8	Нейрон и нейронные сети.	<p>Лекции по разделу.</p> <p>11. Математическая модель искусственного нейрона. Бинарная классификация. Решение логических операций «ИЛИ», «И» и «Исключающее ИЛИ» с помощью персептрона.</p> <p>12. Нейронные сети в задачах классификации и регрессии. Принцип построения нейронной сети MLP. Принцип работы нейронной сети MLP. Проблема переобучения.</p> <p>Лабораторные работы по разделу:</p> <p>7. Реализация нейронной сети с нуля для задач классификации и регрессии.</p>	

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела)	Лекционные занятия	Практические занятия	Лабораторные занятия	Самостоятельная работа	Всего
1	Введение в машинное обучение.	2	0	0	0	2
2	Задача регрессии. Линейная регрессия.	4	0	2	3	9
3	Задача классификации. Логистическая регрессия.	4	0	2	3	9
4	Разведочный анализ данных.	4	0	2	3	9
5	Дерево решений.	4	0	2	3	9
6	Ансамблевые методы.	6	0	2	3	11
7	Метод k-ближайших соседей.	4	0	2	3	9
8	Нейрон и нейронные сети.	6	0	4	4	14
		34	0	16	22	72

14. Методические указания для обучающихся по освоению дисциплины

(рекомендации обучающимся по освоению дисциплины: работа с конспектами лекций, презентационным материалом, выполнение практических заданий, тестов, заданий текущей аттестации и т.д.)

1. При изучении дисциплины рекомендуется использовать следующие средства: рекомендуемую основную и дополнительную литературу; методические указания и пособия; контрольные задания для закрепления теоретического материала; электронные версии учебников и методических указаний для выполнения лабораторно - практических работ (при необходимости материалы рассылаются по электронной почте).
2. Для максимального усвоения дисциплины рекомендуется проведение письменного опроса (тестирование, решение задач) студентов по материалам изученных тем и лабораторных работ. Подборка вопросов для тестирования осуществляется на основе изученного теоретического материала. Такой подход позволяет повысить мотивацию студентов при конспектировании лекционного материала.
3. При проведении лабораторных занятий обеспечивается максимальная степень соответствия с материалом лекционных занятий и осуществляется экспериментальная проверка методов, алгоритмов и технологий обработки информации, излагаемых в рамках лекций.
4. При использовании дистанционных образовательных технологий и электронного обучения выполнять все указания преподавателей, вовремя подключаться к online занятиям, ответственно подходить к заданиям для самостоятельной работы.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

№ п/п	Источник
1	Сирота, Александр Анатольевич. Методы и алгоритмы анализа данных и их моделирование в MATLAB : [учебное пособие] / А.А. Сирота .— Санкт-Петербург : БХВ- Петербург, 2016 .— 381 с. : ил. — Библиогр.: с. 371-374 .— Предм. указ.: с. 377-381 .— ISBN 978-5-9775-3778-0.
2	Практикум по курсу " Нейросетевые технологии обработки информации" [Электронный ресурс] : учебно-методическое пособие для вузов : [для магистров фак. компьютер. наук днев. формы обучения; для направления 09.04.02 - Информ. системы и технологии. / Воронеж. гос. ун-т ; сост.: А.А. Сирота, Е.Ю. Митрофанова , М.А. Дрюченко .— Электрон. текстовые дан. — Воронеж : Издательско-полиграфический центр Воронежского государственного университета, 2013

б) дополнительная литература:

№ п/п	Источник
1	Шолле Ф. Глубокое обучение на Python. / Ф. Шолле. -СПб.: Питер, 2023. — 576 с.
2	Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение / пер. с англ. А. А. Слинкина. — 2-е изд., испр. — М.: ДМК Пресс, 2018. — 652 с.
3	Посполит А. Основы искусственного интеллекта в примерах на Python. Самоучитель. / А. Посполит. -2-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2024. — 448 с.
4	Осинга Д. Глубокое обучение. Готовые решения / Д. Осинга. - М. : Диалектика, 2019. - 288 с.

№ п/п	Источник
5	Осовский С. Нейронные сети для обработки информации / С. Осовский. - М. : Горячая Линия - Телеком, 2017. - 448.
6	Элбон К. Машинное обучение с использованием Python. Сборник рецептов / К. Элбон. - БХВ-Петербург, 2019. - 384.
7	Анализ данных и процессов / А. Барсегян [и др.]. - СПб. : БХВ-ПИТЕР, 2009. - 512 с.
8	Алгазинов, Эдуарт Константинович. Анализ и компьютерное моделирование информационных процессов и систем : [учебное пособие для студ. вузов, обуч. по специальности 080801 "Приклад. информатика" и др. междисциплинар. специальностям] / Э.К. Алгазинов, А.А. Сирота ; под общ. ред. А.А. Сироты .— М. : Диалог-МИФИ, 2009 .— 416 с. : ил .— Библиогр. в конце разд. — ISBN 978-5-86404-233-5

в) информационные электронно-образовательные ресурсы:

№ п/п	Источник
1	Электронный каталог Научной библиотеки Воронежского государственного университета. – (http // www.lib.vsu.ru/).
2	Образовательный портал «Электронный университет ВГУ». – (https://edu.vsu.ru/)
3	ЭБС Лань – Лицензионный договор №3010-14/37-23 от 07.03.2023 (срок предоставления с 12.03.2023 по 11.03.2024)
4	ЭБС «Университетская библиотека online» – Контракт №3010-06/23-22 от 30.12.2022 (срок предоставления с 12.01.2023 по 11.01.2024)
5	ЭБС «Консультант студента» – Лицензионный договор №3010-06/22-22 от 30.12.2022 (с дополнительным соглашением №1 от 09.01.2023) (срок предоставления с 12.01.2023 по 11.01.2024)

16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1	Сирота, Александр Анатольевич. Методы и алгоритмы анализа данных и их моделирование в MATLAB : [учебное пособие] / А.А. Сирота .— Санкт-Петербург : БХВ-Петербург, 2016 .— 381 с. : ил. — Библиогр.: с. 371-374 .— Предм. указ.: с. 377-381 .— ISBN 978-5-9775-3778-0.
2	Практикум по курсу " Нейросетевые технологии обработки информации" [Электронный ресурс] : учебно-методическое пособие для вузов : [для магистров фак. компьютер. наук днев. формы обучения; для направления 09.04.02 - Информ. системы и технологии. / Воронеж. гос. ун-т ; сост.: А.А. Сирота, Е.Ю. Митрофанова , М.А. Дрюченко .— Электрон. текстовые дан. — Воронеж : Издательско-полиграфический центр Воронежского государственного университета, 2013

№ п/п	Источник
3	Практикум по курсу "Моделирование систем" [Электронный ресурс] : учебно-методическое пособие для вузов : [для студ. 4-5 курсов фак. компьютер. наук днев. и вечер. формы обучения; для направлений: 230200 - Информ. системы, 230400 - Информ. системы и технологии; специальности, 230201 - Информ. системы и технологии]. Ч. 1,2 / Воронеж. гос. ун-т ; сост.: А.А. Сирота, Е.Ю. Митрофанова , М.А. Дрюченко .— Электрон. текстовые дан. — Воронеж : Издательско-полиграфический центр Воронежского государственного университета, 2013

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение):

Для реализации учебного процесса используются:

1. Операционная система Windows 7, 8, 10, язык программирования Python, среда локальной разработки Pycharm или облачной разработки Google Colab и Kaggle Notebooks, библиотеки машинного обучения TensorFlow, Keras, PyTorch, scikit-learn, для обработки данных NumPy, Pandas, OpenCV, для визуализации данных: Matplotlib и Seaborn.
2. При проведении занятий в дистанционном режиме обучения используются технические и информационные ресурсы Образовательного портала "Электронный университет ВГУ (<https://edu.vsu.ru>), базирующегося на системе дистанционного обучения Moodle, развернутой в университете.

18. Материально-техническое обеспечение дисциплины:

1) 394018, г. Воронеж, площадь Университетская, д. 1, корпус 1а, ауд. 479

Учебная аудитория: специализированная мебель, компьютер преподавателя i5-8400-2,8ГГц, монитор с ЖК 19", мультимедийный проектор, экран

ПО: ОС Windows v.7, 8, 10, Набор утилит (архиваторы, файл-менеджеры), LibreOffice v.5-7, Foxit PDF Reader.

2) 394018, г. Воронеж, площадь Университетская, д. 1, корпус 1а, аудитория 292

Учебная аудитория: специализированная мебель, компьютер преподавателя Pentium-G3420-3,2ГГц, монитор с ЖК 17", мультимедийный проектор, экран. Система для видеоконференций Logitech ConferenceCam Group и ноутбук 15.6" FHD Lenovo V155-15API

ПО: ОС Windows v.7, 8, 10, Python ver 3.8, Набор утилит (архиваторы, файл-менеджеры), LibreOffice v.5-7, Foxit PDF Reader.

3) 394018, г. Воронеж, площадь Университетская, д. 1, корпус 1а, ауд. 380

Учебная аудитория: специализированная мебель, компьютер преподавателя i3-3240-3,4ГГц, монитор с ЖК 17", мультимедийный проектор, экран

ПО: ОС Windows v.7, 8, 10, Python ver 3.8, Набор утилит (архиваторы, файл-менеджеры), LibreOffice v.5-7, Foxit PDF Reader.

4) 394018, г. Воронеж, площадь Университетская, д. 1, корпус 1а, аудитория 290

Учебная аудитория: специализированная мебель, персональные компьютеры на базе i7-7800x-4ГГц, мониторы ЖК 27" (12 шт.), мультимедийный проектор, экран.

Лабораторное оборудование искусственного интеллекта: рабочие места – персональные компьютеры на базе i7-7800x-4ГГц, мониторы ЖК 27" (12 шт.); модули АО НПЦ «ЭЛВИС»: процессорный Салют-ЭЛ24ПМ2 (9 шт.), отладочный Салют-ЭЛ24ОМ1 (9 шт.), эмулятор MC-USB-JTAG (9 шт.).

Лабораторное оборудование электроники, электротехники и схмотехники: рабочие места – персональные компьютеры на базе i7-7800x-4ГГц, мониторы ЖК 27" (12 шт.); стенд для практических занятий по электрическим цепям (KL-100); стенд для изучения аналоговых электрических схем (KL-200); стенд для изучения цифровых схем (KL-300).

ПО: ОС Windows v.7, 8, 10, Python ver 3.8, Набор утилит (архиваторы, файл-менеджеры), LibreOffice v.5-7, Foxit PDF Reader.

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Разделы дисциплины (модули)	Код компетенции	Код индикатора	Оценочные средства для текущей аттестации
1	Разделы 1-8 Введение в машинное обучение. Задача регрессии. Линейная регрессия. Задача классификации. Логистическая регрессия. Разведочный анализ данных. Дерево решений. Ансамблевые методы. Метод k-ближайших соседей. Нейрон и нейронные сети.	ПК-1	ПК-1.2	Контрольная работа по соответствующим разделам и темам или устный опрос. Выполнение лабораторных работ 1-7.
2	Разделы 1-8 Введение в машинное обучение. Задача регрессии. Линейная регрессия. Задача классификации. Логистическая регрессия. Разведочный анализ данных. Дерево решений. Ансамблевые методы. Метод k-ближайших соседей. Нейрон и нейронные сети.	ПК-1	ПК-1.4	Контрольная работа по соответствующим разделам и темам или устный опрос. Выполнение лабораторных работ 1-7.

Промежуточная аттестация

Форма контроля - Зачет с оценкой, Контрольная работа

Оценочные средства для промежуточной аттестации

20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1 Текущий контроль успеваемости

Текущий контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

- Устный опрос на практических занятиях
- Контрольная работа по теоретической части курса
- Лабораторные работы

20.1.1 Примерный перечень применяемых оценочных средств

№ п/п	Наименование оценочного средства	Представление оценочного средства в фонде	Критерии оценки
1	Устный опрос на практических занятиях	Вопросы по темам/разделам дисциплины	Правильный ответ – зачтено, неправильный или принципиально неточный ответ - не зачтено
2	Контрольная работа по разделам дисциплины	Теоретические вопросы по темам/разделам дисциплины	Шкала оценивания соответствует таблице, приведенной ниже
3	Лабораторная работа	Содержит 7 лабораторных заданий, предусматривающих разработку, проведение сравнительных экспериментов и тестирование (на модельных и реальных данных) систем.	При успешном выполнении работ в течение семестра фиксируется возможность оценивания только теоретической части дисциплины в ходе промежуточной аттестации (зачета), в противном случае проверка задания по лабораторным работам выносится на зачет.

20.1.2 Пример задания для выполнения лабораторной работы

Лабораторная работа № 1

«Задача регрессии. Построение и обучение регрессионной модели»

Цель работы

Целью лабораторной работы является развитие навыков построения моделей регрессии для прогнозирования зависимостей между различными переменными, а также углубление знаний в методах минимизации ошибок регрессионной модели. Студенты должны овладеть навыками визуализации данных, реализации алгоритмов градиентного спуска с нуля и интерпретации результатов обученных моделей.

Форма контроля

Опрос в устной форме по исходному коду и результатам работы реализованной программы.

Пример варианта задания:

Задача: Прогнозирование уровня загрязнения воды в зависимости от количества промышленных выбросов

Вы исследуете влияние промышленных выбросов на уровень загрязнения водоемов. У вас есть данные о количестве промышленных выбросов (в тоннах) и уровне загрязнения воды (измеряемом в ppm – частях на миллион) для нескольких водоемов. Вашей задачей будет построить модель регрессии с квадратичной функцией, чтобы предсказать уровень загрязнения воды на основе количества промышленных выбросов.

Предлагаемый набор данных:

```
# Количество промышленных выбросов (в тоннах)
x = np.array([5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95])

# Уровень загрязнения воды (ppm)
y = np.array([118, 42, 76, 43, 73, 172, 106, 105, 183, 192, 268, 382, 427, 423, 478, 543, 657, 610, 726])
```

Требования:

1. Визуализируйте ваши данные на графике (используйте библиотеку matplotlib).
2. Найдите параметры регрессионной модели (с квадратичной функцией) путём минимизации среднеквадратичной ошибки (MSE) с помощью градиентного спуска, который вы должны реализовать с нуля. При этом, ход ваших вычислений частных производных по параметрам модели показать отдельно соответственно (например, в текстовом файле или в бумажном варианте).
3. Визуализируйте совместно график данных и полученную модель регрессии.
4. Визуализируйте график обучения, на котором было бы видно, как менялась ошибка с каждой итерацией градиентного спуска.
5. Имея полученную модель предскажите значение \hat{y} (уровень загрязнения воды) при $x = 100$ (количество промышленных выбросов). Посмотрите, насколько близким получилось ваше предсказание с реальным значением y . Реальное значение $y = 902$ при данном x .
6. Не забудьте к каждому графику сделать подписи к осям и название графика.

Примеры контрольных вопросов:

1. В чём состоит суть задачи регрессии? Опишите функцию линейной регрессии и её параметры.
2. Что такое функция ошибки? Какие функции ошибок используются для регрессионных моделей?
3. Объясните принцип работы градиентного спуска и его роль в обучении регрессионной модели.

Приведённые ниже задания рекомендуется использовать при проведении диагностических работ для оценки остаточных знаний по дисциплине.

Задания закрытого типа (1 балл)

1. Какие основные различия существуют между задачей классификации и задачей регрессии в контексте машинного обучения?
 - а) задача классификации направлена на прогнозирование непрерывных значений, тогда как задача регрессии оценивает категории;
 - б) задача регрессии ориентирована на предсказание вероятности классов, в то время как задача классификации предсказывает конкретные числовые значения;
 - в) задача классификации разделяет данные на классы, тогда как задача регрессии предсказывает непрерывные значения;
 - г) в задаче регрессии используются только линейные модели, в то время как в задаче классификации допускается использование любых моделей.
2. Как выглядит математическая модель искусственного нейрона?
 - а) $y = \omega x + b$;
 - б) $y = \frac{1}{1 + e^{-(\omega x + b)}}$;
 - в) $y = e^{\omega x + b}$;
 - г) $y = f(\omega x + b)$.
3. Какая из перечисленных стратегий ансамблирования использует бутстрап выборки для создания подмножеств обучающих данных?
 - а) бэггинг;
 - б) бустинг;
 - в) стэкинг;
 - г) кластеризация.
4. Что представляет собой метод k-ближайших соседей в задаче классификации?
 - а) метод, основанный на нахождении среднего значения ближайших соседей.;
 - б) метод, который предсказывает категорию объекта на основе категорий его ближайших соседей;
 - в) метод, преобразующий данные для уменьшения размерности;
 - г) метод, который осуществляет разделение данных на схожие группы без использования правильных ответов.
5. Как работает метод главных компонент (PCA)?
 - а) уменьшает размерность данных, сохраняя при этом максимально возможное количество их информации;

- б) увеличивает количество признаков для лучшей адаптации модели;
- в) использует случайные компоненты для построения модели;
- г) игнорирует все признаки, кроме главных.

6. Зачем используется функция ошибки в процессе обучения модели машинного обучения?

- а) сглаживание градиента для улучшения стабильности оптимизации;
- б) максимизация точности модели на обучающем наборе данных;
- в) минимизация разницы между предсказанными и реальными значениями для нахождения оптимальных параметров модели;
- г) определение оптимальных параметров модели для максимальной сложности.

7. Что означает термин «градиентный спуск» в контексте оптимизации функции ошибки в машинном обучении?

- а) метод оптимизации, направленный на максимизацию функции ошибки для повышения обобщающей способности модели;
- б) алгоритм обучения, который использует градиенты для выбора случайных подмножеств данных и повышения стабильности модели;
- в) техника увеличения размерности данных с использованием градиентов для более точного предсказания значений целевой переменной;
- г) метод поиска локального минимума функции ошибки, который использует производные для определения направления наискорейшего убывания функции.

Задания открытого типа (2 балла)

1. Назовите регрессионный метод, позволяющий предсказывать категории с более чем двумя классами. Используется в случаях, когда результат не принадлежит только двум категориям.
2. Назовите подход к машинному обучению, где модели обучаются на данных без явных ответов. Включает в себя задачи кластеризации, визуализации и обнаружения аномалий.

Задания с развёрнутым ответом (0-3 баллов)

1. Опишите основной принцип работы k-ближайших соседей и взвешенного метода k-ближайших соседей в задачах классификации и регрессии. Какие метрики используются для определения ближайших соседей?

Критерии оценивания	Шкала оценок
Обучающийся приводит развернутое и корректное описание принципа работы методов k-ближайших соседей и взвешенного k-ближайших соседей, а также их применения в задачах классификации и регрессии. Приводит примеры метрик для определения ближайших соседей.	3 балла
Обучающийся приводит достаточно развернутое описание принципа работы методов k-ближайших соседей и взвешенного k-ближайших соседей, а также их применения в задачах классификации и регрессии. Приводит примеры метрик для определения ближайших соседей. В описании допускаются незначительные неточности.	2 балла
Представлено частичное или недостаточно развернутое описание принципа работы методов k-ближайших соседей и взвешенного k-ближайших соседей, а также их применения в задачах классификации и регрессии. В описании могут содержаться отдельные неточности. Отсутствуют примеры метрик для определения ближайших соседей.	1 балл
Представлено неполное или содержащее грубые ошибки описание принципа работы методов k-ближайших соседей и взвешенного k-ближайших соседей. Отсутствуют примеры метрик для определения ближайших соседей.	0 баллов

2. Объясните метод градиентного спуска. Опишите последовательность шагов этого процесса и

объясните, как каждый из этих шагов способствует обновлению весов модели машинного обучения для минимизации ошибки.

Критерии оценивания	Шкала оценок
Обучающийся дает развернутое и точное описание метода градиентного спуска, включая последовательность шагов и его роль в обновлении весов.	3 балла
Обучающийся дает достаточно развернутое описание метода градиентного спуска с некоторыми неточностями или упущениями в деталях.	2 балла
Представлено частичное или недостаточно развернутое описание метода градиентного спуска. Отсутствуют конкретные шаги или объяснения их роли в обновлении весов.	1 балл
Представлено неполное или содержащее грубые ошибки описание метода градиентного спуска.	0 баллов

20.2 Промежуточная аттестация

Промежуточная аттестация может включать в себя проверку теоретических вопросов, а также, при необходимости (в случае невыполнения в течение семестра), проверку выполнения установленного перечня лабораторных заданий, позволяющих оценить уровень полученных знаний и/или практическое (ие) задание(я), позволяющее (ие) оценить степень сформированности умений и навыков.

Для оценки теоретических знаний используется перечень контрольно-измерительных материалов. Каждый контрольно-измерительный материал для проведения промежуточной аттестации включает два задания - вопросов для контроля знаний, умений и владений в рамках оценки уровня сформированности компетенции. При оценивании используется количественная шкала. Критерии оценивания приведены в таблице, приведенной ниже.

Для оценивания результатов обучения на зачете используются следующие содержательные показатели (формулируется с учетом конкретных требований дисциплины):

1. знание теоретических основ учебного материала, основных определений, понятий и используемой терминологии;
2. умение связывать теорию с практикой, умение выявлять и анализировать основные закономерности, полученные, в том числе, в ходе выполнения практических заданий;
3. умение обосновывать свои суждения и профессиональную позицию по излагаемому вопросу;
4. владение навыками программирования в рамках выполняемых практических заданий;
5. владение навыками проведения компьютерного эксперимента, тестирования алгоритмов обработки информации.

Различные комбинации перечисленных показателей определяют критерии оценивания результатов обучения (сформированности компетенций) на государственном экзамене:

- высокий (углубленный) уровень сформированности компетенций;
- повышенный (продвинутый) уровень сформированности компетенций;
- пороговый (базовый) уровень сформированности компетенций.

Для оценивания результатов обучения на зачете используется – зачтено, не зачтено по результатам тестирования.

Соотношение показателей, критериев и шкалы оценивания результатов обучения на государственном экзамене представлено в следующей таблице.

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Обучающийся демонстрирует полное соответствие знаний, умений, навыков по приведенным критериям свободно оперирует понятийным аппаратом и приобретенными знаниями, умениями, применяет их при решении практических задач. Успешно выполнены лабораторные работы в соответствии с установленным перечнем.	Повышенный уровень	Отлично
Ответ на контрольно-измерительный материал не полностью соответствует одному из перечисленных выше показателей, но обучающийся дает правильные ответы на дополнительные вопросы. При этом обучающийся демонстрирует соответствие знаний, умений, навыков приведенным в таблицах показателям, но допускает незначительные ошибки, неточности, испытывает затруднения при решении практических задач. Успешно выполнены лабораторные работы в соответствии с установленным перечнем.	Базовый уровень	Хорошо
Обучающийся демонстрирует неполное соответствие знаний, умений, навыков приведенным в таблицах показателям, допускает значительные ошибки при решении практических задач. При этом ответ на контрольно-измерительный материал не соответствует любым двум из перечисленных показателей, обучающийся дает неполные ответы на дополнительные вопросы. Успешно выполнены лабораторные работы в соответствии с установленным перечнем.	Пороговый уровень	Удовлетворительно
Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Ответ на контрольно-измерительный материал не соответствует любым трем из перечисленных показателей. Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки. Не выполнены лабораторные работы в соответствии с установленным перечнем.	–	Неудовлетворительно

Примерный перечень вопросов к зачету

№	Содержание
1	Искусственный интеллект и машинное обучение. Типы задач машинного обучения. Задача классификации и регрессии.
2	Задача регрессии. Линейная регрессия.
3	Функция ошибки. Производная и градиентный спуск.
4	Стандартизация и нормализация данных. Причины необходимости в нормализации данных.
5	Задача бинарной классификации. Логистическая регрессия.
6	Задача многоклассовой классификации. Мультиномиальная регрессия.
7	Базовые концепции и этапы разведочного анализа данных.
8	Архитектура дерева решений. Техники предотвращения переобучения деревьев.
9	Одиночная модель и ансамбль моделей. Стратегия ансамблирования: Бэггинг.
10	Одиночная модель и ансамбль моделей. Стратегия ансамблирования: Бустинг.
11	Одиночная модель и ансамбль моделей. Стратегия ансамблирования: Стэкинг.
12	Обычный метод k-ближайших соседей. Взвешенный метод k-ближайших соседей. Принцип работы метода в задачах классификации и регрессии.
13	Проблема проклятия размерности признаков. Метод главных компонент. Принцип работы метода.
14	Математическая модель искусственного нейрона. Решение логических операций «ИЛИ», «И» и «Исключающее ИЛИ» с помощью персептрона.
15	Типичная архитектура нейронной сети в задачах классификации и регрессии.
16	Принцип обучения нейронной сети MLP. Проблема переобучения.
17	Метрики оценки моделей классификации: точность, полнота, F1-мера.
18	Метрики оценки моделей регрессии: среднеквадратичная ошибка, средняя абсолютная ошибка, R-квадрат.

Пример контрольно-измерительного материала

УТВЕРЖДАЮ

Заведующий кафедрой технологий обработки и защиты информации

_____ А.А. Сирота

_____.2024

Направление подготовки / специальность 09.03.02 Информационные системы и технологии

Дисциплина Б1.В.ДВ.05.01 Системы машинного обучения

Форма обучения Очное

Вид контроля Дифференцированный зачет

Вид аттестации Промежуточная

Контрольно-измерительный материал № 1

1. Искусственный интеллект и машинное обучение. Типы задач машинного обучения. Задача классификации и регрессии.
2. Обычный и взвешенный метод k-ближайших соседей. Взвешенный метод k-ближайших соседей. Принцип работы метода в задачах классификации и регрессии.

Преподаватель _____ Отырба Р. Р.