

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«КАЛИНИНГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ»

В. Е. Еремичева, М. А. Романов

АНАЛИЗ ДАННЫХ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Учебно-методическое пособие по изучению дисциплины
для студентов бакалавриата по направлению подготовки
15.03.04 Автоматизация технологических процессов и производств

Калининград
Издательство ФГБОУ ВО «КГТУ»
2025

УДК 519.6

Рецензент

Кандидат педагогических наук, доцент кафедры прикладной математики и информационных технологий ФГБОУ ВО «Калининградский государственный технический университет» и.о. зав. кафедрой прикладной математики и информационных технологий Ирина Геннадиевна Булан.

Еремичева, В.Е., Романов, М.А.

Анализ данных и искусственный интеллект: учебно-методическое пособие по изучению дисциплины для студентов бакалавриата, обучающихся по направлению подготовки 15.03.04 Автоматизация технологических процессов и производств / В. Е. Еремичева, М. А. Романов. – Калининград: Изд-во ФГБОУ ВО «КГТУ», 2025. – 70 с.

В учебно-методическом пособии приведены цели и задачи изучения дисциплины, тематическое содержание дисциплины. Представлены методические указания по подготовке к практическим занятиям и самоподготовке. Даны рекомендации по подготовке к промежуточной аттестации, приведены критерии оценивания текущей работы студентов. Пособие подготовлено в соответствии с требованиями утвержденной рабочей программы цифрового модуля по дисциплине «Анализ данных и искусственный интеллект» для направления подготовки 15.03.04 Автоматизация технологических процессов и производств.

Табл. 4, список лит. – 12 наименований

Учебно-методическое пособие рассмотрено и одобрено в качестве локального электронного методического материала на заседании кафедры прикладной математики и информационных технологий института цифровых технологий ФГБОУ ВО «Калининградский государственный технический университет» 11.12.2025 г., протокол № 8

Учебно-методическое пособие по изучению дисциплины рекомендовано к использованию в учебном процессе в качестве локального электронного методического материала методической комиссией ИЦТ ФГБОУ ВО «КГТУ» от 23.12.2025 г., протокол № 8

УДК 519.6

© Федеральное государственное бюджетное образовательное учреждение высшего образования «Калининградский государственный технический университет», 2025г.

© Еремичева В. Е., Романов М. А., 2025 г.

ОГЛАВЛЕНИЕ

1 ВВЕДЕНИЕ	5
1.1 Цели и задачи изучения дисциплины.....	5
1.2 Результаты обучения.....	6
1.3 Место дисциплины в структуре основной профессиональной образовательной программы	6
2 ТЕМАТИЧЕСКИЙ ПЛАН.....	8
3 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ИЗУЧЕНИЮ ДИСЦИПЛИНЫ ...	8
Раздел I. Основные понятия анализа данных, машинного обучения и искусственного интеллекта.	8
Раздел II. Сбор, хранение и подготовка информации.	12
Раздел III. Регрессионный анализ.....	15
Раздел IV. Классификация.....	17
Раздел V. Кластеризация.	19
Раздел VI. Искусственные нейронные сети.....	24
Раздел VII. Рекуррентные и сверточные нейронные сети.	29
Раздел VIII. Применение искусственного интеллекта и нейронных сетей для решения профессиональных задач.	33
4 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ПОДГОТОВКЕ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ	36
Практическое занятие №1. Сбор и хранение информации.....	37
Практическое занятие №2. Обработка информации.	38
Практическое занятие №3. Визуализация данных.....	39
Практическое занятие №4. Линейная регрессия.....	41
Практическое занятие №5. Логистическая регрессия.	42
Практическое занятие №6. Бинарная классификация.	43
Практическое занятие №7. Линейный дискриминантный анализ.	44
Практическое занятие №8. Метод опорных векторов.....	46
Практическое занятие №9. Деревья решений.	47
Практическое занятие №10. Метод k-соседей.	48
Практическое занятие №11. Метод k-ближайших соседей.	49
Практическое занятие №12. DBSCAN.....	50
Практическое занятие №13. Нейронные сети. Перцептрон.	51
Практическое занятие №14. Рекуррентные нейронные сети.....	53
Практическое занятие №15. Сверточные нейронные сети.	54
Практическое занятие №16. Применение нейронных сетей для решения широкого спектра задач.....	55
5 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО САМОСТОЯТЕЛЬНОЙ ПОДГОТОВКЕ	57
6 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ВЫПОЛНЕНИЮ КОНТРОЛЬНОЙ РАБОТЫ (ДЛЯ ЗАОЧНОЙ ФОРМЫ ОБУЧЕНИЯ).....	59

7 ТЕКУЩАЯ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ДИСЦИПЛИНЕ.....	61
Формы текущего контроля	61
Требования к аттестации по дисциплине	61
Критерии оценивания результатов освоения дисциплины	62
8 БИБЛИОГРАФИЧЕСКИЙ СПИСОК.....	65
Основная литература.....	65
Дополнительная литература.....	65
Электронные ресурсы для изучения дисциплины	67
ПРИЛОЖЕНИЕ	68

1 ВВЕДЕНИЕ

Данное учебно-методическое пособие предназначено для студентов направления подготовки 15.03.04 Автоматизация технологических процессов и производств очной и заочной форм обучения, изучающих дисциплину «Анализ данных и искусственный интеллект» и соответствует требованиям, предъявляемым к формируемым компетенциям, знаниям, умениям и навыкам в области применения математического аппарата в профессиональной области, определенными Федеральными государственными образовательными стандартами высшего образования.

1.1 Цели и задачи изучения дисциплины

Дисциплина «Анализ данных и искусственный интеллект» является важной частью современного образования в области информационных технологий, математики, статистики и смежных наук. Изучение этой дисциплины преследует несколько ключевых целей:

- формирование базовых знаний в областях математики, статистики, теории вероятностей, машинного обучения, анализа данных и искусственного интеллекта;
- развитие аналитического мышления;
- практическое применение навыков машинного анализа для решения широкого спектра задач;
- актуализация знаний в области машинного обучения, анализа данных и искусственного интеллекта.

Задачами дисциплины является формирование отдельных элементов общепрофессиональных компетенций, среди них способность:

- понимать основные понятия и концепции анализа данных, таких как регрессия, классификация, кластеризация, работа с временными рядами и другие;
- освоить методы предобработки данных, включая очистку, нормализацию, агрегирование и визуализацию данных;
- овладеть основными алгоритмами машинного обучения;
- применять инструменты и программные средства, использовать специализированные платформы и облачные сервисы для обработки больших объемов данных;
- решать прикладные задачи в различных предметных областях-
- уметь проводить эксперименты с данными, оценивать качество моделей и выбирать наиболее подходящие подходы для конкретных задач.

Полученные знания по данной дисциплине используются при изучении большинства специальных дисциплин.

1.2 Результаты обучения

В результате освоения дисциплины «Анализ данных и искусственный интеллект» обучающийся должен быть способен:

- понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности;
- применять теоретические знания и практические навыки решения задач профессиональной деятельности с применением анализа данных и искусственного интеллекта.

В результате освоения дисциплины студент должен:

знать:

- обзор основных методов анализа и оценки информации, полученной в том числе с помощью цифровых средств;

уметь:

- критически оценивать надежность источников информации в условиях неопределенности и избытка/недостатка информации для решения поставленных задач, в том числе в цифровой среде;
- выявлять и анализировать проблемную ситуацию, выделяя ее структурные составляющие и связи между ними;
- применять методы и алгоритмы машинного обучения для решения профессиональных задач;

владеть:

- аналитическими и системными навыками, способностью к поиску информации.

1.3 Место дисциплины в структуре основной профессиональной образовательной программы

Дисциплина «Анализ данных и искусственный интеллект» относится к цифровому модулю основной профессиональной образовательной программы высшего образования по направлению подготовки 15.03.04 Автоматизация технологических процессов и производств.

При изучении дисциплины используются знания, умения и навыки довузовской подготовки по информатике (знание основных принципов работы ПК и базовые умения работы с пакетами программ системного и офисного назначения) и математики, а также знания, получаемые студентами при освоении дисциплины «Высшая математика»

Дисциплина «Анализ данных и искусственный интеллект» изучается на втором курсе обучения и носит фундаментальный характер. Знания, полученные при освоении данной дисциплины, будут использоваться как в дальнейшей учебной, так и в будущей профессиональной деятельности.

Основными видами аудиторных учебных занятий по дисциплине являются лекции и практические занятия. Формирование знаний обучающихся обеспечивается проведением лекционных занятий.

В ходе изучения дисциплины при проведении практических занятий предусматривается обязательное использование аудиторий, оборудованных персональными компьютерами (ПК). В целях применения эффективных методик обучения в тематическом плане дисциплины на выполнение заданий одного практического занятия отведено 2 академических часа.

В пособии представлены методические материалы дисциплины, включающие тематический план занятий, методические указания по ее изучению, контрольные вопросы для самопроверки и рекомендуемую литературу. Изучать разделы следует последовательно, так, как они приведены в методических указаниях. Изложены методические рекомендации по подготовке к практическим занятиям и самоподготовке, по выполнению контрольной работы для заочной формы обучения. Приведены формы текущего контроля и итоговой аттестации, критерии оценивания. В приложении представлен типовой вариант заданий контрольной работы.

Рекомендуемую литературу можно найти в электронной библиотечной системе университета, в списке литературы указаны ссылки на конкретные издания, размещенные в ЭБС «Университетская библиотека онлайн».

В случае возникновения сложностей при изучении дисциплины студенту необходимо обязательно обращаться к преподавателю в дни консультаций.

Помимо данного пособия, студентам следует использовать материалы, размещенные в соответствующем данной дисциплине разделе ЭИОС университета, в которые более оперативно вносятся изменения для адаптации материалов дисциплины под конкретную группу.

2 ТЕМАТИЧЕСКИЙ ПЛАН

Тематический план лекционных занятий (ЛЗ) представлен в таблице 1.

Таблица 1 – Тематический план ЛЗ

№ п/п	Раздел дисциплины
I	Основные понятия анализа данных, машинного обучения и искусственного интеллекта
II	Сбор, хранение и подготовка информации
III	Регрессионный анализ
IV	Классификация
V	Кластеризация
VI	Искусственные нейронные сети
VII	Рекуррентные и сверточные нейронные сети
VIII	Применение искусственного интеллекта и нейронных сетей для решения профессиональных задач

3 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ИЗУЧЕНИЮ ДИСЦИПЛИНЫ

При подготовке к лекционным занятиям студенту необходимо:

- повторить ранее изученный материал;
- регулярно конспектировать лекционный материал (рекомендуется вести систематизированный, логически связанный и ясный конспект);
- конспект должен быть кратким, но содержать основные положения и ключевые идеи.

После занятия рекомендуется:

- повторить изученный материал;
- дополнить записи на основе прочтения рекомендованной литературы или изучения методических материалов.

Использование рекомендованной литературы и пособий помогает глубже понять изучаемую тему.

В структуру дисциплины входит восемь разделов и двенадцать тем.

Раздел I. Основные понятия анализа данных, машинного обучения и искусственного интеллекта

Тема 1. Введение в анализ данных и машинное обучение

Перечень изучаемых вопросов

1. Определение данных.
2. Типы данных (числовые, категориальные, временные ряды)
3. Сбор данных
4. Предварительная обработка данных
5. Исследование данных
6. Моделирование
7. Интерпретация результатов
8. Основные задачи машинного обучения (классификация, регрессия, кластеризация)
9. Типы моделей машинного обучения
10. Методы оценки качества модели

Методические указания

Данная тема является вводной в курсе дисциплины «Анализ данных и искусственный интеллект», а также других дисциплин информационно-технологического цикла. Основные вопросы, которые освещает данная тема, следующие.

Рассматриваются определение данных, над которыми можно проводить анализ, основные типы данных: числовые, категориальные, временные ряды. Важно понимать, какие именно данные будут использоваться и как они связаны с поставленной задачей машинного обучения. Уделяется внимание основным принципам сбора информации: качество, полнота и актуальность. Процесс получения необходимых данных может включать различные источники: базы данных, веб-сайты, опросы, датчики и другие. Перед использованием данных для анализа необходимо провести их очистку и преобразование: удаление пропущенных значений, нормализация и стандартизация данных, приведение категориальных переменных к числовому виду. Рассматриваются этапы исследования данных и моделирования, на этих этапах проводится анализ собранных данных для выявления закономерностей, корреляций и аномалий. Используются методы визуализации, статистический анализ и другие инструменты. Создаются математические модели, которые будут использовать обучающие данные для предсказания новых значений. Рассматриваются основные задачи и алгоритмы машинного обучения (классификация, регрессия, кластеризация) и методы интерпретации работы моделей. Уделяется внимание методам оценки качества модели.

Контрольные вопросы

1. Что включает в себя понятие "определение данных"?
2. Какие типы данных рассматриваются в рамках этой темы?
3. Почему важно понимать связь между данными и поставленной задачей машинного обучения?
4. Какие принципы следует учитывать при сборе данных?
5. Из каких источников могут собираться данные?
6. Какие шаги включают в себя процесс очистки и преобразования данных перед анализом?
7. Зачем нужно удалять пропущенные значения и нормализовать данные?
8. Какие методы используются на этапе исследования данных?
9. Для чего применяются методы визуализации и статистического анализа?
10. Что такое математическая модель в контексте машинного обучения?
11. Какие основные задачи решаются методами машинного обучения?
12. Чем отличаются задачи классификации, регрессии и кластеризации?
13. Какие алгоритмы используются для решения задач машинного обучения?
14. Как интерпретируются результаты работы моделей?
15. Какие методы оценки качества модели существуют?
16. Почему важно оценивать качество модели?
17. Каковы критерии выбора методов оценки качества модели?
18. Какие аспекты работы с данными наиболее важны для успешного применения машинного обучения?

Рекомендуемая литература по теме: [1, гл. 1, § 1.1, 1.2], [12, гл. 4].

Тема 2. Искусственный интеллект

Перечень изучаемых вопросов

1. Первые шаги ИИ. Современные достижения.
2. Подходы к созданию ИИ.
3. Применение ИИ в различных областях.
4. Этические вопросы использования ИИ.
5. Безопасность ИИ-систем.

Методические указания

Исследуются истоки и эволюция искусственного интеллекта (ИИ), начиная с ранних экспериментов по созданию машин, выполняющих интеллектуальные задачи. Основное внимание уделяется первым исследованиям, таким как тест Тьюринга, направленным на оценку разумности машин. Затем рассматриваются

практические применения ИИ, приведшие к развитию машинного обучения, глубоких нейронных сетей и других современных технологий. Анализируются два основных подхода к разработке ИИ: символический, основанный на правилах и логике, и нейросетевой, имитирующий биологические нейроны. Проводится сравнение их преимуществ и недостатков, а также обсуждается потенциал гибридных методов. Исследуются ключевые элементы каждого подхода и их вклад в результаты. Рассматриваются приложения ИИ в медицине, финансах, транспорте, образовании и искусстве, где подчеркиваются возможности ИИ для повышения эффективности и качества процессов. Приводятся примеры успешного использования ИИ в диагностике болезней, управлении финансовыми рисками, автоматизации транспорта, улучшении образования и создании искусства. Особое внимание уделяется этическим вопросам, связанным с ИИ, таким как ответственность за действия ИИ-систем, защита личных данных, влияние автоматизации на занятость и социальные последствия. Обсуждается необходимость разработки и соблюдения этических норм при создании и эксплуатации ИИ. Также анализируются угрозы и уязвимости ИИ-систем, включая атаки через данные, внедрение вредоносного ПО и изменения в поведении системы. Предлагаются меры по защите ИИ, включающие тестирование, сертификацию, защиту данных и регулярные обновления. Подчеркивается важность непрерывного мониторинга и контроля состояния ИИ-систем для минимизации рисков.

Контрольные вопросы

1. Что такое искусственный интеллект (ИИ)?
2. Какие ранние эксперименты проводились по созданию машин, выполняющих интеллектуальные задачи?
3. В чем заключается суть теста Тьюринга?
4. Каковы были первые практические применения ИИ?
5. Что привело к развитию машинного обучения и глубоких нейронных сетей?
6. Какие два основных подхода существуют в разработке ИИ?
7. В чем заключаются преимущества и недостатки символического подхода к ИИ?
8. В чем заключаются преимущества и недостатки нейросетевого подхода к ИИ?
9. Какой потенциал у гибридных методов в области ИИ?
10. Какие ключевые элементы характерны для символического подхода к ИИ?

11. Какие ключевые элементы характерны для нейросетевого подхода к ИИ?
12. В каких областях применяются технологии ИИ?
13. Как ИИ используется в медицине?
14. Как ИИ применяется в финансовом секторе?
15. Какую роль играет ИИ в транспорте?
16. Как ИИ улучшает образование?
17. Как ИИ влияет на создание искусства?
18. Какие этические вопросы связаны с использованием ИИ?
19. Какие меры предлагаются для защиты ИИ-систем от угроз?
20. Почему важно непрерывно мониторить состояние ИИ-систем?

Рекомендуемая литература по теме: [2, гл. 3], [5, гл. 6.5].

Раздел II. Сбор, хранение и подготовка информации

Тема 3. Сбор, хранение и подготовка информации

Перечень изучаемых вопросов

1. Источники данных.
2. Способы сбора данных.
3. Правовые аспекты сбора данных.
4. Реляционные базы данных.
5. Облачное хранение данных.
6. Очистка данных.
7. Преобразование данных.
8. Кодирование категориальных данных.
9. Мониторинг данных.
10. Контроль качества данных.
11. Визуализация данных.
12. Создание интерактивных дашбордов.

Методические указания

Рассматриваются различные источники данных, служащие основой для последующего анализа. Среди них выделяются базы данных, веб-сайты, опросы, датчики и прочие ресурсы. Особенное внимание обращается на то, какие формы и структуры данных предоставляются разными источниками, и как эти характеристики влияют на процессы дальнейшей обработки и анализа. Описываются методики и технологии, применяемые для извлечения данных из

разнообразных источников. В частности, такие автоматизированных способы, как парсинг веб-страниц и взаимодействие с API, а также традиционные подходы, такие как проведение опросов и анкетирование. Проводится анализ преимуществ и ограничений каждого из этих способов, а также оценивается их воздействие на качество получаемых данных. Изучаются законодательные нормы и требования, регулирующие сбор данных. Акцентируется внимание на важности соблюдения правовых норм для избежания юридических последствий и сохранения доверия со стороны пользователей. Раскрывается суть реляционных баз данных, их внутренняя организация и принципы функционирования. Детально объясняется, как информация хранится в таблицах, связанных друг с другом посредством отношений, и как SQL-запросы помогают извлекать нужные сведения. Оцениваются преимущества реляционных баз данных для хранения структурированных данных и их значимость в процессе анализа. Анализируются особенности облачных хранилищ данных, их сильные и слабые стороны. Большое значение придается вопросам безопасности и доступности данных в облаке, а также преимуществам, связанным с масштабируемостью и экономией ресурсов. Подробно освещаются процедуры очистки данных, являющиеся необходимыми подготовительными шагами перед проведением анализа. Рассматриваются методы удаления недостающих значений, исправления ошибок и устранения дублирующихся записей. Подчеркивается важность качественного очищения данных для повышения точности и достоверности дальнейших анализов. Исследуются методы преобразования данных для приведения их к требуемому формату и структуре. Описываются такие операции, как агрегирование, фильтрация, сортировка и объединение данных. Указывается на необходимость учета особенностей исходных данных и целей анализа при выборе конкретных методов преобразования. Объясняются методы кодирования категориальных данных. Анализируются различия между различными методами кодирования и их влияние на результаты анализа. Обращается внимание на правильный выбор метода кодирования в зависимости от характеристик данных и используемых алгоритмов машинного обучения. Изучаются методы и инструменты для мониторинга данных в режиме реального времени. Описываются подходы к отслеживанию изменений в данных, выявлению аномалий и оперативному реагированию на происходящие события. Отмечается важность мониторинга для поддержания актуальности и высокого качества данных. Анализируются методы и подходы к контролю качества данных. Рассматриваются критерии оценки качества данных, такие как полнота, точность, согласованность и актуальность. Делается акцент на значении регулярных проверок качества данных для обеспечения достоверности и надежности аналитических выводов.

Описываются методы и инструменты для визуализации данных. Рассматриваются различные типы графиков, диаграмм и карт, а также их использование для наглядного и понятного представления данных. Подчеркивается важность визуализации для быстрого восприятия и интерпретации информации. Обсуждаются подходы к созданию интерактивных дашбордов для отображения данных. Описываются популярные инструменты и их функциональные возможности для построения динамических отчетов. Выделение делается на важности интерактивности для эффективного взаимодействия с пользователями и оперативного анализа данных.

Контрольные вопросы

1. Какие источники данных служат основой для анализа?
2. Какие формы и структуры данных предоставляют разные источники?
3. Как влияют характеристики данных на дальнейшую обработку и анализ?
4. Какие методики используются для извлечения данных из различных источников?
5. В чём заключаются преимущества и ограничения парсинга веб-страниц?
6. Каким образом соблюдение правовых норм влияет на доверие пользователей?
7. Что представляет собой структура реляционной базы данных?
8. Для чего используются SQL-запросы в реляционных базах данных?
9. Какие преимущества имеют реляционные базы данных для хранения структурированных данных?
10. Какие особенности присущи облачным хранилищам данных?
11. Какие шаги включают процедуры очистки данных?
12. Зачем необходимо удалять недостающие значения и устранять дубликаты?
13. Какие методы используются для преобразования данных?
14. В чём заключается процесс кодирования категориальных данных?
15. Какие методы мониторинга данных используются в реальном времени?
16. Какие критерии оцениваются при контроле качества данных?
17. Какие виды визуализаций данных наиболее эффективны?
18. Какие инструменты используются для создания интерактивных дашбордов?
19. Почему важна регулярная проверка качества данных?
20. Какая связь существует между особенностями исходных данных и выбором методов их преобразования?

Рекомендуемая литература по теме: [1, гл. 2, 4], [6, гл. 1].

Раздел III. Регрессионный анализ

Тема 4. Регрессионный анализ. Основные понятия. Линейная модель парной регрессии. Нелинейные модели парной регрессии

Перечень изучаемых вопросов

1. Определение регрессионного анализа.
2. Задачи регрессионного анализа.
3. Оценка влияния факторов на зависимую переменную.
4. Парная регрессия.
5. Множественная регрессия.
6. Уравнение линейной регрессии.
7. Оценка параметров модели.
8. Коэффициент детерминации.
9. Полиномиальная регрессия.
10. Экспоненциальная регрессия.
11. Логистическая регрессия.
12. Степенная регрессия.
13. Практическое применение регрессионного анализа.
14. Интерпретация результатов.

Методические указания

Рассматриваются основные аспекты регрессионного анализа. Регрессионный анализ – это статистический метод, предназначенный для изучения взаимосвязей между различными переменными. Он помогает выявить, насколько изменение одной переменной влияет на другую, а также построить модель, которая описывает эту зависимость. В контексте машинного обучения регрессионный анализ часто применяется для решения задач прогнозирования и оценки влияния различных факторов на результат. Изучаются основные задачи регрессионного анализа: прогнозирование будущих значений зависимой переменной на основе наблюдаемых данных, понимание того, как изменения в независимых переменных влияют на зависимую переменную, оценка значимости отдельных факторов и выявление наиболее важных из них, построение математических моделей, которые могут быть использованы для принятия решений. Рассматривается оценка степени влияния каждой независимой переменной на зависимую. Это достигается путем вычисления коэффициентов регрессии, которые показывают, насколько сильно изменяется

значение зависимой переменной при изменении соответствующей независимой переменной на единицу. Также используются методы проверки гипотез для определения статистической значимости каждого фактора. Изучаются основные виды регрессии. Парная регрессия исследует связь между двумя переменными: одной независимой и одной зависимой. Множественная регрессия расширяет возможности парной регрессии, позволяя учитывать несколько независимых переменных одновременно. Уравнение линейной регрессии описывает прямую линию, которая наилучшим образом аппроксимирует зависимость между переменными. Полиномиальная регрессия используется, когда зависимость между переменными нелинейная, но может быть приближена полиномом. Экспоненциальная регрессия применяется, когда зависимость между переменными носит экспоненциальный характер. Логистическая регрессия используется для прогнозирования вероятности события, когда зависимая переменная принимает два возможных значения (например, "да/нет"). Степенная регрессия предполагает использование степенного закона для описания зависимости между переменными. Приводятся примеры практического применения регрессионного анализа. Регрессионный анализ находит широкое применение в различных областях, включая экономику, медицину, инженерию и социологию.

Контрольные вопросы

1. Что такое регрессионный анализ?
2. Какие основные задачи решаются с помощью регрессионного анализа?
3. Как оценивается степень влияния каждой независимой переменной на зависимую?
4. Что такое коэффициент регрессии?
5. Какие методы проверки гипотез используются для определения значимости факторов?
6. Что такое парная регрессия?
7. В чем разница между парной и множественной регрессией?
8. Что описывает уравнение линейной регрессии?
9. Когда применяется полиномиальная регрессия?
10. В каких случаях используется экспоненциальная регрессия?
11. Для чего предназначена логистическая регрессия?
12. Каково назначение степенной регрессии?
13. В каких областях применяется регрессионный анализ?
14. Как регрессионный анализ используется в экономике?
15. Как регрессионный анализ используется в медицине?
16. Как регрессионный анализ используется в инженерии?

17. Как регрессионный анализ используется в социологии?
18. Какие проблемы могут возникнуть при использовании регрессионного анализа?
19. Какие меры принимаются для улучшения точности регрессионных моделей?
20. Как выбрать подходящий вид регрессии для конкретной задачи?

Рекомендуемая литература по теме: [1, гл. 1, § 4.4], [12, гл. 5].

Раздел IV. Классификация

Тема 5. Определение классификации. Применение классификации. Основные методы классификации

Перечень изучаемых вопросов

1. Классификация как задача машинного обучения. Определение задачи классификации.
2. Отличие классификации от регрессии и кластеризации.
3. Бинарная классификация.
4. Многоклассовая классификация.
5. Целевое применение классификации.
6. Применение классификации: медицина, маркетинг, финансы, компьютерное зрение.
7. Классификация методом логистической регрессии.
8. Классификация методом дерева решений.
9. Классификация методом ближайших соседей.
10. Классификация методом опорных векторов.
11. Наивный байесовский классификатор.
12. Оценка качества классификации.
13. Кросс-валидация.

Методические указания

Рассматривается место классификации в контексте машинного обучения. Классификация представляет собой задачу, направленную на распределение объектов или событий по заранее определенным классам. Модель машинного обучения обучается на размеченных данных, где каждому экземпляру соответствует конкретный класс. По завершении обучения модель способна присваивать новый экземпляр соответствующему классу на основе накопленных знаний. Проводится сравнительный анализ классификации с другими методами машинного обучения: кластеризацией и регрессией. Рассматриваются два

основных типа классификации: бинарная и многоклассовая. Бинарная классификация – это разновидность задачи классификации, в которой имеется всего два возможных класса. Примером может служить определение, является ли электронное письмо спамом или нет. Бинарная классификация часто применяется в тех случаях, когда требуется простое деление на две категории, и она относительно проста в реализации и интерпретации. Многоклассовая классификация – это задача, в которой количество классов больше двух. Например, при распознавании рукописных цифр задача заключается в определении одной из десяти возможных цифр (от 0 до 9). Многоклассовая классификация сложнее бинарной, поскольку требует учета большего числа факторов и часто нуждается в более сложных моделях. Изучаются сферы применения классификации с конкретными примерами. Классификация находит множество применений в различных сферах. В медицине она может использоваться для диагностики заболеваний, в маркетинге – для сегментации клиентской базы, в финансах – для оценки кредитных рисков. В области компьютерного зрения классификация используется для распознавания объектов на изображениях, например, для определения наличия лица на фотографии или распознавания дорожных знаков. Во всех этих случаях ключевым моментом является способность модели точно отнести объект к правильному классу, что позволяет принимать обоснованные решения. Изучаются основные методы классификации. Логистическая регрессия – это популярный метод для решения задач бинарной классификации. Он основан на применении сигмоидной функции, которая преобразует линейную комбинацию признаков в вероятность принадлежности к одному из двух классов. Дерево решений – это структура, представляющая собой последовательность вопросов и ответов, ведущих к конечному решению. Каждый узел дерева представляет собой проверку значения определенного признака, а листья соответствуют классам. Метод k-ближайших соседей (kNN) основан на принципе, согласно которому объекты, находящиеся близко друг к другу в пространстве признаков, вероятно, относятся к одному и тому же классу. Для нового объекта определяются ближайшие соседи из обучающей выборки, и на основе их классов принимается решение о классе нового объекта. Метод опорных векторов (SVM) ищет гиперплоскость, которая наилучшим образом разделяет объекты двух классов в пространстве признаков. SVM особенно эффективен для задач с высоким уровнем шума и сложными границами разделения классов. Наивный байесовский классификатор основан на теореме Байеса и предположении о независимости признаков. Несмотря на свою простоту, этот метод может показывать хорошие результаты в ряде задач, особенно там, где признаки действительно независимы или слабо зависимы друг от друга. Приводятся метрики качества классификации (точность, полнота,

мера и AUC-ROC). Описывается методика кросс-валидации для оценки обобщающей способности модели.

Контрольные вопросы

1. Что такое классификация в контексте машинного обучения?
2. Чем отличается бинарная классификация от многоклассовой?
3. В каких сферах применяется классификация?
4. Что такое логистическая регрессия и как она используется в задаче бинарной классификации?
5. Как работает дерево решений в задаче классификации?
6. В чем принцип работы метода k-ближайших соседей (kNN)?
7. Как метод опорных векторов (SVM) решает задачу классификации?
8. Что такое наивный байесовский классификатор и в каких условиях он эффективен?
9. Какие метрики используются для оценки качества классификации?
10. Что такое точность в контексте классификации?
11. Что такое полнота в контексте классификации?
12. Что такое F-мера и как она рассчитывается?
13. Что такое AUC-ROC и как эта метрика интерпретируется?
14. Что такое кросс-валидация и зачем она нужна?
15. Какие данные называются размеченными в контексте машинного обучения?
16. Как модель машинного обучения обучается на размеченных данных?
17. Как определяется класс нового экземпляра после завершения обучения модели?
18. В чем отличие классификации от кластеризации и регрессии?
19. Какие факторы учитываются при выборе метода классификации?
20. Какие сложности возникают при решении задачи многоклассовой классификации?

Рекомендуемая литература по теме: [1, гл. 4.3], [5, гл. 6].

Раздел V. Кластеризация

Тема 6. Основные определения кластеризации. Области применения.

Цели и задачи

Перечень изучаемых вопросов

1. Определение кластеризации.
2. Плотностные кластеры.

3. Иерархические кластеры

4. Связные кластеры.

5. Области применения кластеризации (маркетинг, биология и медицина, информационный поиск, географические информационные системы).

6. Разбиение данных на группы. Обнаружение скрытых структур в данных. Упрощение анализа больших объемов данных.

7. Поиск закономерностей. Выявление общих черт у объектов внутри одного кластера. Определение различий между разными кластерами.

8. Примеры задачи кластеризации

9. Оценка качества кластеризации

Методические указания

Рассматриваются основные понятия, относящиеся к кластеризации. Кластеризация – это метод машинного обучения, который предназначен для группировки объектов или событий в однородные группы, называемые кластерами. Основной целью кластеризации является обнаружение скрытых структур в данных, при этом объекты внутри одного кластера должны быть максимально похожими друг на друга, а между кластерами – отличаться. В отличие от классификации, кластеризация не требует предварительной разметки данных, поэтому она относится к задачам неконтролируемого обучения. Изучаются основные виды кластеров. Плотностные методы кластеризации основаны на концепции плотности данных. Они предполагают, что объекты, принадлежащие одному кластеру, образуют плотные области в пространстве признаков, отделенные от других кластеров областями меньшей плотности. Иерархическая кластеризация строит дерево кластеров, называемое дендрограммой. Существует два типа иерархической кластеризации: агломеративная (объединяющая) и дивизивная (разделяющая). Агломеративная кластеризация начинается с отдельных объектов, постепенно объединяя их в кластеры, пока все объекты не окажутся в одном большом кластере. Дивизивная кластеризация, наоборот, начинает с одного большого кластера и последовательно делит его на меньшие подгруппы. Связные методы кластеризации основываются на понятии расстояния между объектами. Они стремятся объединить объекты в кластеры таким образом, чтобы объекты внутри одного кластера находились на небольшом расстоянии друг от друга, а объекты из разных кластеров – на большем. Изучаются основные области применения алгоритмов кластеризации. Кластеризация находит применение в различных областях: маркетинг, биология, медицина, информационный поиск, географические информационные системы (ГИС). Рассматриваются вопросы разбиения данных на группы, обнаружение скрытых структур в данных,

упрощение анализа больших объемов данных. Кластеризация позволяет разбить большие объемы данных на управляемые группы, что облегчает дальнейший анализ. Обнаружение скрытых структур помогает выявить закономерности, которые могут быть незаметны при поверхностном рассмотрении данных. Это способствует упрощению анализа, делая его более эффективным и целенаправленным. Изучаются вопросы поиска закономерностей и выявление общих черт у объектов внутри одного кластера, а также определение различий между разными кластерами. Кластеризация помогает обнаружить общие черты у объектов внутри одного кластера, что может дать ценные инсайты для дальнейшего анализа. Одновременно с этим, она позволяет выявить различия между разными кластерами, что полезно для дифференциации объектов и создания стратегий на основе этих различий. Приводятся примеры задач кластеризации. Уделяется внимание оценке качества кластеризации.

Контрольные вопросы

1. Что такое кластеризация?
2. Какова цель кластеризации?
3. В чем основное отличие кластеризации от классификации?
4. Что представляют собой плотностные методы кластеризации?
5. В чем заключается принцип работы иерархической кластеризации?
6. Какие типы иерархической кластеризации существуют?
7. Как работают связные методы кластеризации?
8. В каких областях применяется кластеризация?
9. Как кластеризация помогает упростить анализ больших объемов данных?
10. Как кластеризация помогает выявить скрытые структуры в данных?
11. Как кластеризация используется в маркетинге?
12. Как кластеризация используется в биологии?
13. Как кластеризация используется в медицине?
14. Как кластеризация используется в информационных поисковых системах?
15. Как кластеризация используется в географических информационных системах (ГИС)?
16. Как кластеризация помогает найти общие черты у объектов внутри одного кластера?
17. Как кластеризация помогает определить различия между разными кластерами?
18. Какие примеры задач кластеризации приводятся в тексте?
19. Как оценивается качество кластеризации?

20. Какие метрики используются для оценки качества кластеризации?

Рекомендуемая литература по теме: [2, гл. 2], [5, гл. 6].

Тема 7. Алгоритмы кластеризации. Иерархические алгоритмы. Нечеткие алгоритмы. Искусственный интеллект. Сравнительная характеристика алгоритмов

Перечень изучаемых вопросов

1. Общие сведения об алгоритмах кластеризации.
2. Классификация алгоритмов кластеризации.
3. Основные этапы кластеризации.
4. Иерархические алгоритмы.
5. Нечеткие алгоритмы.
6. Искусственный интеллект в кластеризации.
7. Сравнительная характеристика алгоритмов.
8. Перспективы развития кластеризации.

Методические указания

Приводятся общие сведения об алгоритмах кластеризации. Алгоритмы кластеризации представляют собой методы машинного обучения, предназначенные для автоматической группировки объектов или событий в однородные группы, именуемые кластерами. Ключевая особенность кластеризации заключается в отсутствии необходимости предварительной разметки данных, что отличает её от контролируемого обучения, такого как классификация. Алгоритмы кластеризации направлены на обнаружение естественных структур в данных, основанных на схожести объектов по различным признакам. Разбирается классификация алгоритмов классификации. Существует несколько типов алгоритмов кластеризации, которые можно разделить на три основные категории: иерархические алгоритмы, нечёткие алгоритмы, связные алгоритмы. Рассматриваются основные этапы кластеризации. Процесс кластеризации обычно включает следующие шаги: подготовка данных, выбор алгоритма, настройка параметров, выполнение кластеризации, анализ результатов. Изучаются иерархические алгоритмы. Иерархические алгоритмы кластеризации формируют структуру кластеров в виде дерева, называемого дендрограммой. Они бывают двух видов: агломеративные, дивизивные. Изучаются нечеткие алгоритмы. Нечёткая кластеризация позволяет объектам принадлежать сразу нескольким кластерам с

разной степенью вероятности. Приводятся перспективы развития кластеризации. Развитие кластеризации идёт в направлении интеграции с искусственным интеллектом и глубокого обучения, что позволит автоматизировать выбор алгоритмов и повысить качество кластеризации. Будущее кластеризации связано с обработкой больших и разнородных данных, таких как мультимедийные данные и сенсорные измерения.

Контрольные вопросы

1. Какие основные этапы процесса кластеризации вы можете назвать?
2. Что такое дендрограмма в контексте иерархических алгоритмов кластеризации?
3. В чем отличие агломеративных и дивизивных иерархических алгоритмов?
4. Что означает нечеткая кластеризация и как она работает?
5. Какие перспективы развития кластеризации вы видите?
6. Каким образом интеграция кластеризации с искусственным интеллектом и глубоким обучением может улучшить качество кластеризации?
7. Какие типы данных будут преимущественно обрабатываться алгоритмами кластеризации в будущем?
8. Какие параметры нужно настроить перед выполнением кластеризации?
9. Как проводится анализ результатов кластеризации?
10. Какие преимущества имеет иерархическая кластеризация перед другими видами?
11. Какие трудности могут возникать при выполнении кластеризации?
12. Какие критерии используются для выбора подходящего алгоритма кластеризации?
13. Как можно оценить качество полученных кластеров?
14. Какие области применения алгоритмов кластеризации вы знаете?
15. Какие данные требуют предварительной подготовки перед началом кластеризации?
16. Какие дополнительные методы могут применяться совместно с кластеризацией для улучшения результатов?
17. Какие современные направления исследований в области кластеризации вы можете назвать?

Рекомендуемая литература по теме: [2, гл. 2], [5, гл. 6].

Раздел VI. Искусственные нейронные сети

Тема 8. Введение в искусственные нейронные сети. Основы архитектуры нейронных сетей. Обучение нейронных сетей

Перечень изучаемых вопросов

1. История развития нейронных сетей.
2. Основные компоненты нейронных сетей. Нейроны.
3. Весовые коэффициенты.
4. Активационные функции.
5. Типы нейронных сетей.
6. Слои нейронной сети.
7. Функции потерь.
8. Регуляризация.
9. Оптимизационные методы.
10. Обратное распространение ошибки

Методические указания

Изучается история развития нейронных сетей. Она берет начало в середине XX века, когда исследователи начали разрабатывать идеи, вдохновленные биологическими нейронами мозга. Первые теоретические основы искусственных нейронов были заложены Уорреном Мак-Каллоком и Уолтером Питтсом в 1943 г., когда они предложили модель искусственного нейрона, который мог выполнять простые логические операции. Позже Фрэнк Розенблатт разработал перцептрон – первый алгоритм, который смог обучаться на основе входных данных. Однако настоящий прорыв произошел в 1980-х годах, когда были разработаны многослойные нейронные сети и обратное распространение ошибки, что позволило решать более сложные задачи. В последние десятилетия развитие нейронных сетей ускорилось благодаря росту вычислительных мощностей и появлению больших наборов данных, таких как ImageNet, которые позволили тренировать глубокие нейронные сети, способные решать задачи распознавания изображений, обработки естественного языка и генерации текста. Изучаются основные компоненты нейронных сетей (нейроны). Нейронные сети состоят из нескольких компонентов, основными из которых являются нейроны. Нейроны – это базовые строительные блоки нейронной сети, они принимают входные сигналы, взвешивают их с помощью весовых коэффициентов и передают результат активации через активационную функцию. Входные сигналы поступают на нейрон, умножаются на соответствующие веса, суммируются и проходят через активационную функцию, которая определяет, передавать ли сигнал дальше. Разбираются весовые коэффициенты. Весовые

коэффициенты определяют силу связи между нейроном и его входами. В процессе обучения сеть настраивает эти веса, чтобы минимизировать ошибку на выходе. Каждое соединение между нейронами имеет свой вес, который обновляется в ходе обучения. При каждом проходе данных через сеть ошибка на выходе сравнивается с ожидаемым результатом, и на основе разницы между ними происходит обновление весов. Рассматриваются активационные функции. Активационные функции решают, стоит ли передавать сигнал далее или нет. Они играют важную роль в работе нейронной сети, определяя, сколько сигнала должно пройти через нейрон. Популярные активационные функции включают ReLU, sigmoid и tanh. Каждая функция выполняет нелинейное преобразование входного сигнала, чтобы улучшить способность сети к обучению сложным функциям. Разбираются основные типы нейронных сетей. Существуют различные типы нейронных сетей, разработанные для решения специфических задач. Конволюционные нейронные сети (CNN): применяются для обработки изображений, видео и аудиосигналов. Рекуррентные нейронные сети (RNN): работают с последовательностями данных, такими как тексты или временные ряды. Генеративные состязательные сети (GAN): могут создавать реалистичные изображения или тексты. Каждая архитектура нейронной сети предназначена для решения определенной задачи, будь то классификация, регрессия или генерация данных. Приводится описание слоев нейронной сети. Нейронные сети состоят из слоев, которые выполняют различные виды преобразований данных. Входной слой получает исходные данные, промежуточные слои проводят обработку данных, а выходной слой выдает окончательный результат. Количество слоев и их сложность варьируются в зависимости от сложности задачи, стоящей перед сетью. Разбирается функция потерь. Функции потерь измеряют разницу между выходными данными и целевыми значениями. Они служат индикатором того, насколько хорошо сеть решает задачу. Например, средняя абсолютная ошибка измеряет среднюю разницу между предсказаниями и истинными значениями, а кросс-энтропия оценивает эффективность классификации. Рассматриваются оптимизационные методы. Оптимизационные методы, такие как градиентный спуск, позволяют изменять веса в сети таким образом, чтобы минимизировать ошибку на выходе. Градиентный спуск вычисляет производные функций потерь по отношению к весовым коэффициентам и использует эту информацию для коррекции весов. Другие методы, такие как Adam и RMSProp, адаптивно настраивают шаг градиента, чтобы ускорить обучение и стабилизировать процесс.

Контрольные вопросы

1. Что такое перцептрон, и кто его разработал?
2. Когда произошел значительный прорыв в развитии нейронных сетей?
3. Какие факторы способствовали ускоренному развитию нейронных сетей в последние десятилетия?
4. Из каких основных компонентов состоят нейронные сети?
5. Какова роль нейронов в нейронной сети?
6. Что такое весовые коэффициенты и какую роль они играют в обучении нейронной сети?
7. Какие активационные функции используются в нейронных сетях и как они работают?
8. Какие основные типы нейронных сетей вы знаете?
9. Как рекуррентные нейронные сети (RNN) обрабатывают последовательности данных?
10. Из каких слоев состоит нейронная сеть и какая у них роль?
11. Что такое функция потерь и как она используется в обучении нейронной сети?
12. Какие оптимизационные методы применяются для настройки весов в нейронной сети?
13. Как градиентный спуск помогает корректировать веса в нейронной сети?
14. Какие задачи могут решать нейронные сети?
15. Какие наборы данных использовались для тренировки глубоких нейронных сетей?
16. Какие достижения в области нейронных сетей произошли в последнее время?
17. Какие перспективы развития нейронных сетей вы видите в будущем?

Рекомендуемая литература по теме: [2, гл. 3], [7, гл. 3–6].

Тема 9. Глубокое обучение. Применение нейронных сетей. Этические аспекты и будущее нейронных сетей

Перечень изучаемых вопросов

1. Глубокая архитектура нейронных сетей.
2. Преимущества глубокого обучения.
3. Использование предобученных моделей.
4. Fine-tuning.
5. Генеративные состязательные сети.

6. Компьютерное зрение.
7. Обработка естественного языка (NLP).
8. Рекомендательные системы.
9. Предвзятость и этика в ИИ.
10. Будущее нейронных сетей

Методические указания

Глубокие нейронные сети характеризуются наличием большого количества слоёв, что позволяет им решать сложные задачи, такие как распознавание изображений, обработка естественного языка и генерация текста. В отличие от традиционных нейронных сетей, ГНС способны выполнять нелинейные преобразования на высоком уровне абстракции, что делает их мощным инструментом для решения задач, требующих высокой точности и детализированного анализа. Приводятся преимущества глубокого обучения. Глубокое обучение предоставляет ряд значительных преимуществ: высокая точность, универсальность, масштабируемость. Предобученные модели, такие как AlexNet, VGG и ResNet, предоставляют готовые решения для многих задач, что снижает затраты на разработку и ускоряет внедрение технологий. Такие модели уже прошли предварительное обучение на огромных наборах данных, что позволяет использовать их в качестве фундамента для решения специализированных задач. Описывается процесс тонкой настройки (fine-tuning) предварительно обученной модели для решения конкретной задачи. Этот подход позволяет сохранить большую часть первоначальной архитектуры и веса, адаптируя их к новым условиям. Это уменьшает потребность в полном пересмотре всей модели, обеспечивая значительное сокращение затрат на обучение. Изучаются основные сферы реализации глубокого обучения. Обработка естественного языка (NLP) включает в себя задачи перевода, синтеза текста и анализа текста. Нейронные сети, такие как LSTM и трансформеры, показали значительный прогресс в решении задач NLP, предоставляя более качественные переводы и анализ текстов, чем предыдущие подходы. Рекомендательные системы используют нейронные сети для предоставления персонализированных предложений пользователям. Они учатся на пользовательском поведении и исторических данных, чтобы предлагать контент, который наиболее вероятен для конкретного пользователя. Приводятся правовые и этические аспекты в ИИ. Предвзятость и этические проблемы становятся всё более важными вопросами в развитии ИИ. Разработчики обязаны учитывать потенциальные риски предвзятости и дискриминации, а также обеспечивать прозрачность и объяснимость моделей. Этические аспекты включают в себя вопросы конфиденциальности, защиты персональных данных и ответственности

за возможные негативные последствия использования ИИ. Приводятся перспективы развития нейронных сетей. Будущее нейронных сетей направлено на улучшение их производительности, точности и надёжности. Продолжающееся совершенствование аппаратуры и программного обеспечения, а также интеграция новых подходов, таких как квантовые вычисления, помогут преодолеть существующие ограничения и расширить область применения нейронных сетей.

Контрольные вопросы

1. Что такое глубокие нейронные сети (ГНС)?
2. В чем преимущество ГНС перед традиционными нейронными сетями?
3. Какие преимущества предоставляет глубокое обучение?
4. Что такое предобученные модели и как они используются?
5. Что такое тонкая настройка (fine-tuning) и почему она полезна?
6. Какие задачи решает обработка естественного языка (NLP) с помощью нейронных сетей?
7. Какие нейронные сети используются в рекомендательных системах?
8. Какие правовые и этические аспекты следует учитывать при разработке ИИ?
9. Как предвзятость и дискриминация могут влиять на работу ИИ?
10. Как обеспечивается прозрачность и объяснимость моделей ИИ?
11. Какие перспективы развития нейронных сетей вы видите в будущем?
12. Как квантовые вычисления могут повлиять на развитие нейронных сетей?
13. Какие конкретные задачи можно решить с помощью предобученных моделей?
14. Какие подходы используются для предотвращения предвзятости в ИИ?
15. Как обеспечить конфиденциальность и защиту персональных данных в ИИ?
16. Как ответственность за возможные негативные последствия использования ИИ распределяется среди разработчиков и пользователей?
17. Какие новые подходы к обучению нейронных сетей появляются в настоящее время?
18. Какие ограничения существуют у текущих нейронных сетей и как их можно преодолеть?
19. Какие специализированные задачи можно решить с помощью глубокой нейронной сети?
20. Как нейронные сети могут улучшить качество жизни людей в будущем?

Рекомендуемая литература по теме: [2, гл. 3], [7, гл. 4–6].

Раздел VII. Рекуррентные и сверточные нейронные сети

Тема 10. Рекуррентные нейронные сети

Перечень изучаемых вопросов

1. Введение в рекуррентные нейронные сети (RNN).
2. Отличие RNN от традиционных нейронных сетей.
3. Современное состояние и развитие RNN.
4. Основные компоненты RNN.
5. Базовая структура RNN.
6. Виды RNN.
7. Обучение RNN.
8. Применение RNN.
9. Современные модификации RNN. LSTM и GRU.

Методические указания

Изучаются основные определения, связанные с рекуррентными нейронными сетями (RNN). RNN представляют собой особый класс архитектур искусственных нейронных сетей, предназначенных для обработки последовательной информации. Они позволяют учитывать контекст предыдущих шагов при обработке текущих данных, что делает их особенно полезными для задач, связанных с временными рядами, текстовыми данными и другими последовательностями. Производится сравнение с другими видами нейронных сетей. В отличие от традиционных нейронных сетей, где каждый вход обрабатывается независимо от других, рекуррентные нейронные сети обладают памятью о предыдущих состояниях системы. Это позволяет им сохранять информацию о последовательности событий и использовать её для прогнозирования будущих состояний. Такой подход значительно расширяет возможности применения нейронных сетей в задачах, требующих учета временных зависимостей. Рассматриваются современное состояние и развитие RNN. современные рекуррентные нейронные сети активно развиваются и находят применение в самых разных областях, таких как обработка естественного языка, распознавание речи, управление роботами и многие другие. С развитием вычислительных мощностей и появлением новых методов оптимизации, RNN становятся все более эффективными и точными. Изучаются основные компоненты RNN. Основными компонентами рекуррентных нейронных сетей являются: входной слой, который принимает текущую

последовательность данных, скрытый слой, содержащий память о предыдущих шагах и вычисляющий текущее состояние сети, выходной слой, генерирующий результат на основе текущего состояния сети. Разбирается базовая структура RNN. Базовая структура рекуррентной нейронной сети включает в себя несколько слоев, соединенных между собой таким образом, чтобы информация могла передаваться от одного шага к другому. На каждом шаге сеть обновляет свое внутреннее состояние, используя информацию о предыдущем состоянии и текущем входе. Рассматриваются основные виды RNN. Существует несколько видов рекуррентных нейронных сетей, включая однонаправленные RNN, двунаправленные RNN и многослойные RNN. Однонаправленные RNN учитывают только предыдущие шаги, тогда как двунаправленные RNN могут обрабатывать информацию как вперед, так и назад по времени. Многослойные RNN содержат несколько скрытых слоев, что позволяет им лучше улавливать сложные зависимости в данных. Приводится описание процесса обучения RNN. Обучение рекуррентных нейронных сетей осуществляется путем использования метода обратного распространения ошибки через время (Backpropagation Through Time). Этот метод позволяет корректировать веса сети на основе ошибок, накопленных за несколько временных шагов. Рассматриваются сферы применения и современные модификации RNN. Рекуррентные нейронные сети широко применяются в различных областях, включая обработку текстов, генерацию музыки, предсказание временных рядов и многое другое. Они способны эффективно справляться с задачами, требующими учета временной динамики данных. Для улучшения производительности рекуррентных нейронных сетей были разработаны специальные архитектуры, такие как долговременная краткосрочная память (LSTM) и гейтированный рекуррентный блок (GRU). Эти архитектуры позволяют избежать проблемы исчезающего градиента, которая часто возникает при обучении стандартных RNN, и обеспечивают более стабильную работу сети на длинных последовательностях.

Контрольные вопросы

1. Что такое рекуррентные нейронные сети (RNN)?
2. В чем отличие RNN от традиционных нейронных сетей?
3. Как RNN учитывает контекст предыдущих шагов при обработке данных?
4. Какие задачи можно решать с помощью RNN?
5. Какие компоненты входят в состав RNN?
6. Как устроена базовая структура RNN?
7. Какие виды RNN существуют?

8. В чем заключается метод обратного распространения ошибки через время (BPTT)?

9. Какие сферы применения есть у RNN?

10. Какие современные модификации RNN существуют?

11. Что такое проблема исчезающего градиента в RNN?

12. Как работают долговременная краткосрочная память (LSTM) и гейтированный рекуррентный блок (GRU)?

13. Как LSTM и GRU решают проблему исчезающего градиента?

14. Как многослойные RNN улучшают производительность?

15. Какие данные чаще всего обрабатываются с помощью RNN?

16. Как RNN используются в обработке естественного языка?

17. Как RNN применяются в распознавании речи?

18. Как RNN используются в управлении роботами?

19. Какие перспективы развития RNN вы видите в будущем?

Рекомендуемая литература по теме: [2, гл. 3], [7, гл. 5], [8, гл. 3].

Тема 11. Сверточные нейронные сети

Перечень изучаемых вопросов

1. Введение в сверточные нейронные сети (CNN).

2. История развития CNN.

3. Основные компоненты CNN.

4. Архитектура CNN.

5. Обучение CNN.

6. Применение CNN.

7. Современные модификации CNN.

Методические указания

Рассматриваются ключевые аспекты сверточных нейронных сетей (CNN). Сверточные нейронные сети (Convolutional Neural Networks, CNN) представляют собой специализированный вид нейронных сетей, разработанный для эффективного извлечения признаков из многомерных данных, таких как изображения. Основная идея заключается в использовании сверток для обнаружения локальных особенностей, что позволяет сетям автоматически выделять важные признаки без необходимости явного программирования вручную. Изучается история развития CNN. История сверточных нейронных сетей начинается с работ Хьюбела и Визеля, которые исследовали механизмы зрительного восприятия в мозге млекопитающих. Их открытия легли в основу

создания первых сверточных нейронных сетей, известных как Neocognitron, предложенного Кунихико Фукушимой в 1980 г. Однако широкое распространение CNN получили лишь в начале XXI века благодаря успеху сети LeNet-5, разработанной Яном Лекуном для распознавания рукописных цифр. Разбираются основные компоненты и архитектура CNN. Основные компоненты сверточных нейронных сетей включают: слои свертки: выполняют операцию свертки, которая выявляет локальные особенности в изображении, пульсирующие слои: уменьшают размерность данных, сохраняя наиболее значимые характеристики, полносвязанные слои: аналогичны слоям в обычных нейронных сетях и используются для классификации или регрессии. Архитектура сверточной нейронной сети обычно состоит из чередующихся слоев свертки и пулирования, за которыми следуют один или несколько полносвязанных слоев. Такая архитектура позволяет постепенно увеличивать уровень абстракции признаков, начиная с простых деталей и заканчивая сложными объектами. Приводится принцип обучения CNN. Обучение сверточных нейронных сетей происходит посредством метода обратного распространения ошибки, как и в случае с обычными нейронными сетями. Однако из-за большого количества параметров и высокой вычислительной сложности, обучение CNN требует значительных ресурсов и времени. Для ускорения процесса часто используют графические процессоры (GPU) и распределенное обучение. Рассматриваются сферы применения и современные модификации CNN. Сверточные нейронные сети нашли широкое применение в области компьютерного зрения, включая задачи распознавания изображений, сегментации объектов, детекции лиц и многих других. Кроме того, CNN успешно применяются в обработке сигналов, анализе текстов и даже в музыке. Современные модификации сверточных нейронных сетей направлены на улучшение их эффективности и точности. К ним относятся: Residual Networks (ResNets), Densely Connected Convolutional Networks (DenseNets), Capsule Networks.

Контрольные вопросы

1. Что такое сверточные нейронные сети (CNN)?
2. Какова основная идея использования сверток в CNN?
3. Кто был автором первых сверточных нейронных сетей и когда они появились?
4. Какую роль сыграли работы Хьюбела и Визеля в развитии CNN?
5. Какие основные компоненты входят в архитектуру CNN?
6. Как работает операция свертки в слое свертки?

7. Для чего нужны пульсирующие слои в CNN?
8. Какова роль полносвязанных слоев в архитектуре CNN?
9. Как происходит обучение сверточных нейронных сетей?
10. Какие ресурсы необходимы для обучения CNN?
11. В каких областях применяются сверточные нейронные сети?
12. Какие современные модификации CNN вы знаете?
13. Что такое Residual Networks (ResNets) и как они улучшают эффективность CNN?
14. Что такое Densely Connected Convolutional Networks (DenseNets) и как они отличаются от классических CNN?
15. Что такое Capsule Networks и как они изменили подход к обработке данных в CNN?
16. Какие задачи в области компьютерного зрения решаются с помощью CNN?
17. Как CNN используются в обработке сигналов?
18. Как CNN применяются в анализе текстов?
19. Как CNN используются в музыке?
20. Какие перспективы развития CNN вы видите в будущем?

Рекомендуемая литература по теме: [2, гл. 3], [7, гл. 5], [8, гл. 3].

Раздел VIII. Применение искусственного интеллекта и нейронных сетей для решения профессиональных задач

Тема 12. Назначение

Перечень изучаемых вопросов

1. Основные сферы применения ИИ и нейронных сетей.
2. Профессиональные задачи и требования.
3. Применение ИИ и нейронных сетей в финансах.
4. Применение ИИ и нейронных сетей в здравоохранении.
5. Применение ИИ и нейронных сетей в производстве.
6. Применение ИИ и нейронных сетей в транспорте.
7. Применение ИИ и нейронных сетей в энергетике.
8. Применение ИИ и нейронных сетей в образовании.

Методические указания

В этой теме рассматриваются основные сферы применения нейронных сетей и искусственного интеллекта. Искусственный интеллект (ИИ) и нейронные сети находят применение в широком спектре областей, охватывающих практически все аспекты современной жизни. Среди основных сфер применения

можно выделить: компьютерное зрение, обработка естественного языка: перевод, синтез речи, чат-боты, анализ текстов, робототехника, медицинская диагностика: анализ медицинских изображений, выявление заболеваний, персонализированная медицина, финансы, производство, транспорт, энергетика, образование. Работа с искусственным интеллектом и нейронными сетями требует глубоких знаний в области математики, статистики, информатики и программирования. Специалисты должны уметь разрабатывать и внедрять алгоритмы машинного обучения, работать с большими объемами данных, а также обладать навыками визуализации и интерпретации результатов. Важным аспектом является понимание этических вопросов, связанных с применением ИИ, и умение оценивать потенциальные риски и последствия внедрения технологий. Финансовый сектор активно использует искусственный интеллект и нейронные сети для повышения эффективности операций и снижения рисков. Здравоохранение является одной из ключевых областей, где искусственный интеллект и нейронные сети демонстрируют значительный потенциал. Производство становится всё более автоматизированным и интеллектуальным благодаря использованию искусственного интеллекта и нейронных сетей. Транспортная отрасль также активно внедряет технологии искусственного интеллекта и нейронных сетей. Энергетика является важной областью, где искусственный интеллект и нейронные сети помогают улучшить эффективность и устойчивость энергоснабжения. Образование также претерпевает значительные изменения под влиянием технологий искусственного интеллекта и нейронных сетей. Таким образом, искусственный интеллект и нейронные сети играют ключевую роль в трансформации различных отраслей экономики и общества, открывая новые возможности для повышения эффективности, безопасности и комфорта нашей жизни.

Контрольные вопросы

1. Какие основные сферы применения нейронных сетей и искусственного интеллекта рассматриваются в данном тексте?
2. Какие навыки требуются специалистам для работы с искусственным интеллектом и нейронными сетями?
3. Какие этические вопросы связаны с применением искусственного интеллекта?
4. Как финансовый сектор использует искусственный интеллект и нейронные сети?
5. Каковы перспективы применения искусственного интеллекта и нейронных сетей в здравоохранении?

6. Как производство становится более автоматизированным и интеллектуальным благодаря искусственному интеллекту?

7. Какие изменения происходят в транспортной отрасли под влиянием технологий искусственного интеллекта?

8. Как искусственный интеллект и нейронные сети помогают улучшить энергетику?

9. Как образование трансформируется под воздействием технологий искусственного интеллекта?

10. Какие риски и последствия могут возникнуть при внедрении технологий искусственного интеллекта?

11. Какие задачи в области компьютерного зрения решают нейронные сети?

12. Как нейронные сети используются в обработке естественного языка?

13. Как ИИ и нейронные сети применяются в медицинской диагностике?

14. Какие возможности открываются для производства благодаря применению ИИ?

15. Как ИИ и нейронные сети способствуют повышению безопасности в транспорте?

16. Как ИИ помогает снизить риски в финансовой сфере?

17. Какие образовательные технологии основаны на искусственном интеллекте?

18. Как ИИ и нейронные сети повышают комфорт в повседневной жизни?

19. Какие математические и статистические знания необходимы для работы с ИИ?

20. Как специалисты визуализируют и интерпретируют результаты работы нейронных сетей?

Рекомендуемая литература по теме: [1, гл. 1, § 1.1, 1.2], [7, гл. 3], [8, гл. 2, 3].

4 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ПОДГОТОВКЕ К ПРАКТИЧЕСКИМ ЗАНЯТИЯМ

В процессе выполнения практических занятий (ПЗ) обучающиеся систематизируют и закрепляют полученные теоретические знания, развивают интеллектуальные и профессиональные умения, формируют элементы компетенций будущих специалистов.

Перед началом выполнения заданий рекомендуется повторить теоретический материал, относящийся к теме практического занятия, ознакомиться с рекомендуемой литературой.

В каждой практической работе имеется:

- тема;
- цель;
- практические задания;
- вывод;
- контрольные вопросы.

Практические задания по дисциплине «Анализ данных и искусственный интеллект» выполняются с использованием ПК, офисного и прикладного программного обеспечения (ПО Logiom, ПО Anaconda), сети Интернет.

Выполненные практические задания оформляются в соответствии с требованиями в виде отчета в электронном виде (шрифт 12 пт). Титульный лист для всех работ содержит номер, тему практического занятия, группу, вариант и фамилию студента. В отчете должна быть указана цель работы, размещены скриншоты выполненных заданий, сделан вывод, даны ответы на контрольные вопросы.

Отчет вместе с файлами, содержащими выполненные задания, размещается в облачном хранилище в папке студента, доступ на редактирование которой имеет преподаватель.

Закончив выполнение работы, обучающийся через специальный элемент курса в ЭИОС КГТУ отправляет ссылку на папку с заданиями и отчетом в облачном хранилище на проверку.

Для студентов заочной формы обучения в аудиторное время в тематическом плане предусмотрено выполнение заданий практических занятий № 1, 4, 10, 12, 16. Остальные занятия предполагают самостоятельное выполнение.

Варианты заданий и методические рекомендации по их выполнению представлены на соответствующем курсе дисциплины в ЭИОС.

Тематический план ПЗ представлен в таблице 2.

Таблица 2 – Тематический план ПЗ

№ ПЗ	Раздел дисциплины / Тема ПЗ
I. Сбор, хранение и обработка информации	
1	Сбор и хранение информации
2	Обработка информации
3	Визуализация данных
II. Регрессионный анализ	
4	Линейная регрессия
5	Логистическая регрессия
III. Классификация	
6	Бинарная классификация
7	Линейный дискриминантный анализ
8	Метод опорных векторов
9	Деревья решений
10	Метод k-соседей
IV. Кластеризация	
11	Метод K-ближайших соседей
12	DBSCAN
V. Искусственные нейронные сети	
13	Нейронные сети. Перцептрон
14	Рекуррентные нейронные сети
15	Сверточные нейронные сети
16	Применение нейронных сетей для решения широкого спектра задач

Практическое занятие № 1. Сбор и хранение информации

Цель

Освоить навыки работы с разными типами данных, осуществить их правильный сбор и хранение в специализированных форматах.

Задание

1. Запустить ПО Loginom. Создать пустой сценарий.
2. Загрузить один и тот же набор данных в разных форматах (xlsx, csv, lgd, txt).
3. Проверить корректность загрузки, выставить необходимые свойства для каждого столба и таблицы.
4. Выполнить визуализацию полученных результатов.
5. Сохранить файл.

6. Запустить ПО Anaconda (Jupyter Notebook).
7. Загрузить тот же самый набор данных в разных форматах (xlsx, csv, txt).
8. Проверить корректность загрузки, выставить необходимые свойства для каждого столба и таблицы.
9. Выполнить визуализацию полученных результатов.
10. Сохранить файл.
11. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Какие методы сбора данных существуют?
2. В чем заключается разница между структурированными и неструктурированными данными?
3. Каковы основные этапы процесса сбора данных?
4. Что такое API и как оно используется для получения данных?
5. Какие инструменты используются для хранения больших объемов данных?
6. В чем отличие реляционных баз данных от NoSQL-баз данных?
7. Какие проблемы могут возникнуть при сборе данных из различных источников?
8. Как обеспечить безопасность данных при их передаче и хранении?
9. Какие форматы файлов наиболее часто используются для хранения данных?
10. Как выбрать подходящую базу данных для конкретного проекта?
11. Какие меры предосторожности следует принимать при работе с конфиденциальными данными?
12. Какова роль облачных технологий в хранении и обработке данных?
13. Какие задачи решает система управления базами данных (СУБД)?
14. Какие современные тенденции наблюдаются в области сбора и хранения данных?

Практическое занятие № 2. Обработка информации

Цель

Получить навыки первичной обработки данных: удаление шумов и выбросов, стандартизация данных, заполнение пропусков.

Задание

1. Запустить ПО Loginom.

2. Загрузить готовый набор данных.
3. Удалить пустые строки.
4. Обработать выбросы и шумы с помощью имеющихся инструментов.
5. Заполнить пропуски.
6. Выполнить визуализацию полученных данных.
7. Сохранить файл.
8. Запустить ПО Anaconda (Jupyter Notebook).
9. Загрузить готовый набор данных.
10. Удалить пустые строки.
11. Обработать выбросы и шумы с помощью имеющихся инструментов.
12. Заполнить пропуски.
13. Выполнить визуализацию полученных данных.
14. Сохранить файл.
15. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Какие этапы включает в себя предварительная обработка данных?
2. Почему важно проводить предварительную обработку данных перед анализом?
3. Какую роль играет удаление выбросов в процессе подготовки данных?
4. Какие методы заполнения пропусков в данных существуют?
5. Что такое нормализация данных и почему она важна?
6. Какие методы стандартизации данных существуют?
7. В каких случаях применяется метод биннинга (дискретизация)?
8. Какой подход используется для уменьшения размерности данных?
9. Какие преимущества и недостатки имеют существующие методы предварительной обработки?
10. Какие ошибки допускаются чаще всего при проведении предварительной обработки данных?
11. Какие рекомендации можно дать для оптимизации процесса предварительной обработки данных?

Практическое занятие № 3. Визуализация данных

Цель

Получить навыки визуализации. Научиться строить диаграммы и графики.

Задание

1. Запустить ПО Loginom.
2. Загрузить готовый набор данных.
3. Построить гистограммы распределения данных.
4. Построить круговые диаграммы распределения данных.
5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Загрузить готовый набор данных.
8. Построить гистограммы распределения данных.
9. Построить круговые диаграммы распределения данных.
10. Сохранить файл.
11. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Какие цели преследует визуализация данных?
2. Какие основные виды графиков и диаграмм существуют?
3. Какую информацию можно извлечь из гистограмм?
4. Для чего нужны box-plots (ящики с усами) и как они помогают в анализе данных?
5. Каким образом scatter plots (точечные графики) помогают выявить корреляции между переменными?
6. Какую роль играет цветовая палитра в визуализации данных?
7. Какие принципы дизайна необходимо учитывать при создании инфографики?
8. Как выбрать подходящий график для представления определенного типа данных?
9. Как визуализировать временные ряды данных?
10. Какие особенности данных влияют на выбор типа графика?
11. Как интерпретируются графики плотности распределения?
12. Какие преимущества и недостатки имеют различные типы графиков?
13. Какие виды данных требуют особого подхода к визуализации?
14. Какие последствия имеет некачественная визуализация данных на дальнейший анализ?
15. Какие ошибки допускаются чаще всего при создании визуализаций данных?
16. Какие перспективы развития методов визуализации данных обсуждаются в научной литературе?

Практическое занятие № 4. Линейная регрессия

Цель

Получение навыков линейного регрессионного анализа.

Задание

1. Запустить ПО Loginom.
2. Загрузить готовый набор данных.
3. Выполнить линейный регрессионный анализ.
4. Выполнить визуализацию полученных результатов.
5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Выполнить линейный регрессионный анализ.
8. Выполнить визуализацию полученных результатов.
9. Сохранить файл.
10. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Что такое линейная регрессия и для чего она используется?
2. Какие предпосылки должны выполняться для использования линейной регрессии?
3. Как выбираются независимые переменные в модели линейной регрессии?
4. Какой смысл имеют коэффициенты в уравнении линейной регрессии?
5. Как вычисляется коэффициент детерминации R^2 и что он означает?
6. Как интерпретируется значение p-value в контексте линейной регрессии?
7. Какие меры были предприняты для проверки мультиколлинеарности в модели?
8. Как выявляются и устраняются выбросы в данных при использовании линейной регрессии?
9. Какие подходы используются для тестирования значимости коэффициентов модели?
10. Какие метрики используются для оценки качества модели линейной регрессии?
11. Какие особенности данных влияют на выбор метода оценки параметров модели?

12. Какие преимущества и недостатки имеет использование линейной регрессии?

13. Какие виды данных требуют особого подхода при применении линейной регрессии?

14. Какие альтернативные методы моделирования можно использовать вместо линейной регрессии?

15. Какие последствия имеет неправильный выбор независимых переменных на точность модели?

16. Какие ошибки допускаются чаще всего при построении модели линейной регрессии?

17. Какие рекомендации можно дать для оптимизации процесса построения модели линейной регрессии?

Практическое занятие № 5. Логистическая регрессия

Цель

Получение навыков логистического регрессионного анализа.

Задание

1. Запустить ПО Loginom.
2. Загрузить готовый набор данных.
3. Выполнить логистический регрессионный анализ.
4. Выполнить визуализацию полученных результатов.
5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Выполнить логистический регрессионный анализ.
8. Выполнить визуализацию полученных результатов.
9. Сохранить файл.
10. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Что такое логистическая регрессия и для чего она используется?
2. Какие предпосылки должны выполняться для использования логистической регрессии?
3. Как выбираются независимые переменные в модели логистической регрессии?
4. Какой смысл имеют коэффициенты в уравнении логистической регрессии?

5. Как вычисляется вероятность принадлежности к классу в модели логистической регрессии?
6. Какие методы оценки параметров модели логистической регрессии существуют?
7. Как интерпретируется значение p -value в контексте логистической регрессии?
8. Какие меры предпринимаются для проверки мультиколлинеарности в модели?
9. Как выявляются и устраняются выбросы в данных при использовании логистической регрессии?
10. Какие метрики используются для оценки качества модели логистической регрессии?
11. Какие преимущества и недостатки имеет использование логистической регрессии?
12. Какие виды данных требуют особого подхода при применении логистической регрессии?
13. Какие выводы можно сделать о качестве модели логистической регрессии после её построения?
14. Какие альтернативные методы классификации могли бы быть использованы вместо логистической регрессии?
15. Какие последствия имеет неправильный выбор независимых переменных на точность модели?
16. Какие ошибки допускаются чаще всего при построении модели логистической регрессии?

Практическое занятие № 6. Бинарная классификация

Цель

Получение навыков бинарной классификации.

Задание

1. Запустить ПО Loginom.
2. Загрузить готовый набор данных.
3. Выполнить бинарную классификацию.
4. Выполнить визуализацию полученных результатов.
5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Выполнить бинарную классификацию.
8. Выполнить визуализацию полученных результатов.

9. Сохранить файл.
10. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Что такое бинарная классификация и для чего она используется?
2. Какие предпосылки должны выполняться для использования методов бинарной классификации?
3. Как выбираются признаки (features) в модели бинарной классификации?
4. Какой смысл имеют коэффициенты в уравнениях классификационных моделей?
5. Как вычисляется вероятность принадлежности к одному из двух классов в моделях бинарной классификации?
6. Какие методы оценки параметров модели бинарной классификации используются?
7. Как интерпретируется значение p-value в контексте бинарной классификации?
8. Какие меры предпринимаются для проверки мультиколлинеарности в модели?
9. Как выявляются и устраняются выбросы в данных при использовании методов бинарной классификации?
10. Какие метрики используются для оценки качества модели бинарной классификации?
11. Какие преимущества и недостатки имеет использование методов бинарной классификации?
12. Какие виды данных требуют особого подхода при применении методов бинарной классификации?
13. Какие выводы можно сделать о качестве модели бинарной классификации после её построения?
14. Какие последствия имеет неправильный выбор признаков на точность модели?
15. Какие ошибки допускаются чаще всего при построении модели бинарной классификации?

Практическое занятие № 7. Линейный дискриминантный анализ

Цель

Получение навыков классификации с помощью метода линейного дискриминантного анализа.

Задание

1. Запустить ПО Loginom.
2. Загрузить готовый набор данных.
3. Выполнить классификацию методом линейного дискриминантного анализа.
4. Выполнить визуализацию полученных результатов.
5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Выполнить классификацию методом линейного дискриминантного анализа.
8. Выполнить визуализацию полученных результатов.
9. Сохранить файл.
10. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Как определяется матрица ковариации в ЛДА и как она влияет на разделение классов?
2. Объясните математический принцип разделения гиперплоскости в пространстве признаков в ЛДА.
3. Как связаны понятия максимальной вероятности и минимизации дисперсии внутри класса в ЛДА?
4. Какие ограничения накладываются на распределение данных в классе при использовании ЛДА?
5. Опишите процедуру поиска оптимальной проекции в многомерном пространстве признаков в ЛДА.
6. Как ЛДА справляется с проблемой высокой размерности данных?
7. Как ЛДА учитывает априорную вероятность классов при принятии решений?
8. Можно ли использовать ЛДА для многоклассовой классификации? Если да, то каким образом?
9. Как ЛДА может использоваться для уменьшения размерности пространства признаков без потери важной информации?
10. Почему ЛДА предпочтительнее других методов классификации в случае малого количества образцов и большого числа признаков?
11. Как ЛДА обрабатывает случаи, когда классы сильно перекрываются?
12. Чем обусловлен выбор критерия Фишера в ЛДА для максимизации межклассовых различий?
13. Оцените устойчивость ЛДА к шумам и выбросам в данных.

14. Объясните, как ЛДА может быть использована для сегментации изображений.

Практическое занятие № 8. Метод опорных векторов

Цель

Получение навыков классификации с помощью метода опорных векторов.

Задание

1. Запустить ПО Loginom.
2. Загрузить готовый набор данных.
3. Выполнить классификацию методом опорных векторов.
4. Выполнить визуализацию полученных результатов.
5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Выполнить классификацию методом опорных векторов.
8. Выполнить визуализацию полученных результатов.
9. Сохранить файл.
10. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Как определяется оптимальное решение задачи квадратичного программирования в МОВ?
2. Объясните понятие мягкого и жесткого маржинального условия в МОВ.
3. Как влияет выбор ядра на производительность МОВ? Приведите примеры различных ядер и их применения.
4. Каково математическое обоснование использования метода Лагранжа для нахождения оптимального решения в МОВ?
5. В чем заключается проблема переноса знаний в МОВ и как ее решают?
6. Какая связь существует между МОВ и статистическими методами обучения?
7. Какова роль регуляризации в МОВ и как она влияет на обобщающую способность модели?
8. Как МОВ справляется с задачей многоклассовой классификации?
9. Как можно интерпретировать весовые коэффициенты в МОВ?
10. Какое влияние оказывает выбор функции потерь на обучение МОВ?
11. Как реализовать МОВ для задач регрессии?

12. Как изменяется сложность алгоритма МОВ при увеличении размера выборки?

13. Как происходит выбор опорных векторов в МОВ? Объясните математические аспекты этого процесса.

14. Как применить МОВ для задачи ранжирования?

15. Как использовать МОВ для классификации временных рядов?

16. Как сравнить SVM с другими методами классификации, такими как случайные леса и градиентный бустинг?

17. Каковы преимущества и недостатки использования МОВ по сравнению с нейронными сетями?

18. Как применять МОВ для анализа текста?

19. Как оптимизировать гиперпараметры МОВ с использованием методов перекрестной проверки?

Практическое занятие № 9. Деревья решений

Цель

Получение навыков классификации с помощью метода деревьев решений.

Задание

1. Запустить ПО Loginom.
2. Загрузить готовый набор данных.
3. Выполнить классификацию методом деревьев решений.
4. Выполнить визуализацию полученных результатов.
5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Выполнить классификацию методом деревьев решений.
8. Выполнить визуализацию полученных результатов.
9. Сохранить файл.
10. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Как определяются узлы ветвления и критерии расщепления в дереве решений?

2. Объясните математически критерий Джини и энтропию, используемые для измерения неоднородности в узлах дерева.

3. Как дерево решений справляется с проблемой высокой размерности данных?

4. Как избежать переобучения в деревьях решений? Обсудите методы ранней остановки и обрезки ветвей.
5. Как выбирается глубина дерева решений и как она влияет на его производительность?
6. Какова роль рекурсивного расщепления в построении дерева решений?
7. Как деревья решений справляются с категоричными и числовыми признаками одновременно?
8. Как построить ансамбль деревьев решений и повысить точность классификации?
9. Как влияет выбор метрики на оценку качества узлов дерева решений?
10. Как происходит оценка важности признаков в деревьях решений?
11. Как решается проблема мульти-классового обучения в деревьях решений?
12. Как деревья решений могут справляться с проблемами несбалансированных классов?
13. Как можно интерпретировать правила, полученные из дерева решений?
14. Как деревья решений работают с пропусками в данных?
15. Как выбирать стратегию расщепления узлов в деревьях решений?
16. Как деревья решений обрабатывают нелинейные зависимости в данных?
17. Как производится оценка вероятности принадлежности к классу в листьях дерева решений?
18. Каковы различия между алгоритмами ID3, C4.5 и CART?
19. Как использовать деревья решений для анализа временных рядов?

Практическое занятие № 10. Метод k-соседей

Цель

Получение навыков классификации с помощью метода k-соседей.

Задание

1. Запустить ПО Loginom.
2. Загрузить готовый набор данных.
3. Выполнить классификацию методом k-соседей.
4. Выполнить визуализацию полученных результатов.
5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Выполнить классификацию методом k-соседей.
8. Выполнить визуализацию полученных результатов.

9. Сохранить файл.

10. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Как определяется расстояние между объектами в методе k-соседей? Какие метрики расстояния обычно используются и почему?

2. Как влияет выбор значения k на точность классификации? Почему слишком маленькое или большое значение k может привести к ухудшению результата?

3. Как метод k-соседей справляется с проблемой высокой размерности данных? Какие техники снижения размерности могут быть полезны в этом случае?

4. Как метод k-соседей обрабатывает категориальные признаки? Существуют ли специальные методы для работы с ними?

5. Как метод k-соседей справляется с несбалансированными классами? Какие подходы можно использовать для улучшения классификации в таких ситуациях?

6. Как метод k-соседей может быть использован для регрессионных задач? В чем отличия от классификации?

7. Как метод k-соседей работает с временными рядами? Какие модификации необходимы для эффективного применения?

8. Как метод k-соседей применяется в задачах рекомендательных систем? Какие особенности нужно учесть при его использовании?

9. Как метод k-соседей используется в задачах обнаружения аномалий? Какие ключевые моменты стоит учитывать?

10. Как метод k-соседей реагирует на изменение масштаба данных? Нужно ли нормализовать данные перед применением метода?

Практическое занятие № 11. Метод k-ближайших соседей

Цель

Получение навыков кластеризации с помощью метода k-ближайших соседей.

Задание

1. Запустить ПО Loginom.

2. Загрузить готовый набор данных.

3. Выполнить кластеризацию методом k-ближайших соседей.

4. Выполнить визуализацию полученных результатов.

5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Выполнить кластеризацию методом k-ближайших соседей.
8. Выполнить визуализацию полученных результатов.
9. Сохранить файл.
10. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Как определяется начальная центроидная точка в методе k-means? Какие подходы к инициализации центров кластеров существуют?
2. Как влияет выбор значения k на качество кластеризации? Почему слишком малое или большое значение k может привести к неудовлетворительному результату?
3. Какие подходы можно использовать для улучшения кластеризации?
4. Как метод k-means применяется в задачах рекомендательных систем? Какие особенности нужно учесть при его использовании?
5. Как метод k-means соотносится с другими методами кластеризации, такими как иерархическая кластеризация и DBSCAN? В каких случаях k-means будет предпочтительным выбором?

Практическое занятие № 12. DBSCAN

Цель

Получение навыков кластеризации с помощью метода DBSCAN.

Задание

1. Запустить ПО Loginom.
2. Загрузить готовый набор данных.
3. Выполнить кластеризацию методом DBSCAN.
4. Выполнить визуализацию полученных результатов.
5. Сохранить файл.
6. Запустить ПО Anaconda (Jupyter Notebook).
7. Выполнить кластеризацию методом DBSCAN.
8. Выполнить визуализацию полученных результатов.
9. Сохранить файл.
10. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Как определяется плотность окрестностей точек в методе DBSCAN? Какие параметры влияют на эту плотность?
2. Как влияет выбор значений ϵ и minPts на качество кластеризации? Почему слишком маленькие или большие значения могут привести к неудовлетворительному результату?
3. Как метод DBSCAN справляется с проблемой высокой размерности данных? Какие техники снижения размерности могут быть полезны в этом случае?
4. Как метод DBSCAN обрабатывает категориальные признаки? Существуют ли специальные методы для работы с ними?
5. Как метод DBSCAN справляется с несбалансированными кластерами? Какие подходы можно использовать для улучшения кластеризации в таких ситуациях?
6. Как метод DBSCAN относится к проблеме шума и выбросов в данных? Какие методы фильтрации могут быть применены?
7. Как метод DBSCAN применяется в задачах рекомендательных систем? Какие особенности нужно учесть при его использовании?
8. Как метод DBSCAN решает задачу мультикластерного обучения? Какие подходы могут быть использованы для расширения метода на несколько кластеров?

Практическое занятие № 13. Нейронные сети. Перцептрон

Цель

Изучить особенности однослойного перцептрона и его ограничения.

Задание

1. Запустить ПО Loginom.
2. Изучить математическую модель, структурную схему простого перцептрона и правило обучения.
3. Изучить математическую модель, структурную схему сигмоидального нейрона.
4. Сформировать нейронную сеть.
5. Обучить нейронную сеть для распознавания образов.
6. Сохранить результат.
7. Запустить ПО Anaconda (Jupyter Notebook).
8. Сформировать нейронную сеть.
9. Обучить нейронную сеть для распознавания образов.

10. Сохранить результат.

11. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Как определяется функция активации в перцептроне и как она влияет на его обучаемость?

2. Объясните математику градиентного спуска в контексте обучения перцептрона.

3. Как перцептроны решают проблему XOR и почему они не способны справиться с ней без дополнительных слоев?

4. Как связано количество входных сигналов и число синаптических весов в перцептроне?

5. Какова роль порогового элемента в архитектуре перцептрона?

6. Как влияет величина шага обучения (learning rate) на скорость и стабильность обучения перцептрона?

7. Как перцептрон справляется с задачами, где данные не линейно разделяемы?

8. Как реализуется обратное распространение ошибки в многослойном перцептроне?

9. Как выбрать оптимальное количество скрытых слоев и нейронов в каждом слое для конкретной задачи?

10. Как минимизировать эффект переобучения в перцептроне?

11. Каковы основные ограничения однослойного перцептрона и как их преодолеть?

12. Как связаны архитектура перцептрона и теорема о представлении Ципфа?

13. Как организовать параллельное обучение нескольких перцептронов для увеличения скорости вычислений?

14. Как оценить общую ошибку перцептрона на этапе тестирования?

15. Как интерпретировать веса нейронов в перцептроне?

16. Как предотвратить насыщение нейронов в перцептроне?

17. Как применять регуляризацию в перцептроне для улучшения обобщающей способности?

18. Как обрабатывать пропуски в данных при обучении перцептрона?

19. Как реализовать динамическое изменение топологии перцептрона в процессе обучения?

20. Как бороться с проблемой локальных минимумов в перцептроне?

21. Как работать с категориальными признаками в перцептроне?

Практическое занятие № 14. Рекуррентные нейронные сети

Цель

Получение навыков с работы рекуррентной нейронной сетью. Изучить алгоритм LSTM для генерации текстов.

Задание

1. Воспользовавшись теоретической информацией, провести предварительную обработку текстовых данных, на основе предложенной выборки.
2. Реализовать алгоритм генерации текстов на основе рекуррентной нейронной сети LSTM.
3. Запустить ПО Anaconda (Jupyter Notebook).
4. Сохранить результат.
5. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Как определяется зависимость текущего состояния от предыдущих состояний в рекуррентных нейронных сетях (RNN)?
2. Как рекуррентные нейронные сети решают проблему затухания градиента и взрыва градиента?
3. Как LSTM (Long Short-Term Memory) отличается от стандартной RNN и как эта архитектура помогает сохранять долгосрочную память?
4. Как GRU (Gated Recurrent Unit) сравнивается с LSTM и в каких случаях его использование предпочтительно?
5. Как RNN справляется с задачами, где данные представляют собой временные ряды?
6. Как влияет длина последовательности ввода на производительность RNN?
7. Как выбрать оптимальное количество скрытых слоев и нейронов в каждом слое для конкретной задачи?
8. Как минимизировать эффект переобучения в RNN?
9. Как организованы процессы обновления памяти и сброса в LSTM?
10. Как оценить общую ошибку RNN на этапе тестирования?
11. Как интерпретировать веса нейронов в RNN?
12. Как предотвратить насыщение нейронов в RNN?
13. Как применять регуляризацию в RNN для улучшения обобщающей способности?

14. Как использовать RNN для задач классификации многомерных данных?

Практическое занятие № 15. Сверточные нейронные сети

Цель

Получение практических навыков по реализации пайплайна для решения задачи машинного обучения с использованием сверточной нейросетевой модели.

Задание

1. Ознакомиться с работой сверточных нейронных сетей в библиотеке Keras.
2. Научиться решать задачу сегментации изображений на основе нейронных сетей.
3. Запустить ПО Anaconda (Jupyter Notebook).
4. Сохранить результат.
5. Сделать выводы о проделанной работе, подготовить отчет.

Контрольные вопросы

1. Как определяется свертка в сверточных нейронных сетях (CNN) и как она влияет на обучение?
2. Объясните математику пулинга (pooling) в CNN и как он влияет на производительность и интерпретацию результатов?
3. Как CNN справляется с проблемой избыточной информации и высокочастотной информацией в изображениях?
4. Как выбирается оптимальное количество фильтров и их размеры в слоях свертки в CNN?
5. Как влияет величина шага обучения (learning rate) на скорость и стабильность обучения CNN?
6. Как минимизировать эффект переобучения в CNN?
7. Как выбрать оптимальное количество сверточных слоев и нейронов в каждом слое для конкретной задачи?
8. Как оценить общую ошибку CNN на этапе тестирования?
9. Как интерпретировать фильтры в CNN?
10. Как предотвратить насыщение нейронов в CNN?
11. Как применять регуляризацию в CNN для улучшения обобщающей способности?
12. Как использовать CNN для задач классификации многомерных данных?

13. Как работает CNN с различными функциями активации, такими как ReLU и Leaky ReLU?

14. Как обучить CNN на потоковых данных?

15. Как уменьшить вычислительные затраты в CNN при сохранении производительности?

Практическое занятие № 16. Применение нейронных сетей для решения широкого спектра задач

Цель

Изучить работу современных генеративных нейронных сетей. Получить навыки применения генеративных нейронных сетей для решения профессиональных задач.

Задание

1. В любой доступной среде генеративной нейронной сети (GigaChat и т. п.) выполнить генерацию различных текстов с помощью запросов (пром프트) согласно заданию.

2. Выполнить генерацию различных изображений помощью запросов (пром프트) согласно заданию.

3. Сравнить полученные результаты в зависимости от запроса.

4. Сделать выводы, подготовить отчет.

Контрольные вопросы

1. Как выбрать оптимальную архитектуру нейронной сети для конкретной задачи? Какие факторы необходимо учитывать при выборе архитектуры?

2. Как определить оптимальное количество слоев и нейронов в нейронной сети? Какие эвристики можно использовать для настройки архитектуры?

3. Как оценить общую ошибку нейронной сети на этапе тестирования? Какие метрики лучше всего подходят для оценки качества модели?

4. Как улучшить производительность нейронных сетей при работе с большими объемами данных? Какие ускорители и распределённые системы можно использовать?

5. Как комбинировать несколько типов данных в одной нейронной сети? Какие методы объединения данных можно использовать?

6. Как уменьшить вычислительные затраты в нейронных сетях при сохранении производительности? Какие методы сжатия и квантования можно использовать?

7. Как генеративные нейронные сети решают проблему генерации реалистичных изображений? Какие методы обработки данных используют для улучшения качества генерации?

8. Как генеративные нейронные сети справляются с проблемами зашумления и выбросов в данных? Какие подходы можно использовать для устранения этих проблем?

9. Как генеративные нейронные сети обеспечивают высокую точность воспроизведения деталей? Какие методы оптимизации параметров используют для повышения качества генерации?

10. Как генеративные нейронные сети предотвращают потерю важных особенностей изображения? Какие методы сохранения деталей применяют генеративные нейронные сети?

11. Как генеративные нейронные сети управляют качеством генерации в режиме реального времени? Какие механизмы обратной связи используют генеративные нейронные сети для контроля качества генерируемого контента?

12. Как генеративные нейронные сети избегают генерации повторяющихся фрагментов? Какие методы предотвращения дублирования используют генеративные нейронные сети ?

13. Как генеративные нейронные сети сохраняют естественную динамику сцены? Какие подходы используют генеративные нейронные сети для поддержания плавного движения объектов?

14. Как генеративные нейронные сети обрабатывают пропажи в данных? Какие методы заполнения пропусков используют генеративные нейронные сети?

15. Как генеративные нейронные сети работают с большим количеством данных? Какие методы ускорения и оптимизации вычислений применяют генеративные нейронные сети для обработки больших объемов данных?

16. Как генеративные нейронные сети обучаются на потоковых данных? Какие особенности работы с потоковыми данными выделяют генеративные нейронные сети?

5 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО САМОСТОЯТЕЛЬНОЙ ПОДГОТОВКЕ

Внеаудиторная самостоятельная работа (СР) проводится с целью:

- систематизации и закрепления полученных теоретических знаний и практических умений;
- углубления и расширения теоретических знаний;
- формирования умений использовать нормативную, правовую, справочную документацию и специальную литературу;
- формирования самостоятельности мышления, способностей к саморазвитию, самосовершенствованию и самореализации;
- развития исследовательских умений;
- использования материала, собранного и полученного в ходе самостоятельных практических занятий для эффективной подготовки к итоговой аттестации.

Внеаудиторная СР выполняется студентом без непосредственного участия преподавателя, но по его заданию, и предполагает:

- подготовку к аудиторным занятиям (лекциям, практическим занятиям) и выполнение соответствующих заданий;
- самостоятельную работу над отдельными темами учебной дисциплины в соответствии с тематическим планом;
- подготовку к промежуточной и итоговой аттестации (зачету/экзамену).

При подготовке к лекции рекомендуется повторить ранее изученный материал, что дает возможность получить необходимые разъяснения преподавателя непосредственно в ходе занятия. В часы самостоятельной подготовки рекомендуется повторить изучаемый материал и доработать конспект.

Подготовка к практическим занятиям предусматривает детальную проработку учебного материала, рекомендованной литературы и методической разработки на предстоящее занятие.

При организации самостоятельного изучения ряда тем студент работает в соответствии с указаниями, выданными преподавателем. Указания по изучению материала дисциплины по каждой теме включают в себя:

- название, цели и задачи изучения;
- основные вопросы, характеристику основных понятий и определений, необходимых для усвоения данной темы;

- краткие выводы, ориентирующие обучаемого на определенную совокупность сведений, основных идей и ключевых положений;
- список рекомендуемой литературы;
- контрольные вопросы для самопроверки.

В ходе самостоятельной подготовки к экзамену при анализе имеющегося теоретического и практического материала студенту также рекомендуется проводить постановку различного рода задач по изучаемой теме.

При изучении дисциплины предусмотрены следующие виды внеаудиторной СР:

- работа с конспектами лекций;
- выполнение самостоятельных практических заданий;
- подготовка видеорефератов на основе презентаций;
- работа с электронными ресурсами в сети Интернет;
- чтение основной и дополнительной литературы по дисциплине.

Понятие «видеореферат» означает, что презентация должна быть непрерывно выполняющейся (в виде видеофайла), где на каждом кадре виден слайд и студент, озвучивающий этот слайд. Альтернативный вариант – записать выступление на телефон так, чтобы в кадре был виден слайд и студент. Озвучивание материала должно совпадать с изображением на слайде.

По всем темам дисциплины в качестве СР предусмотрена подготовка и отправка в ЭИОС отчетов установленной формы с ответами на контрольные вопросы с приложением фотографий письменных конспектов лекций.

Виды СР по разделам и темам дисциплины указаны в таблице 3.

Таблица 3 – Перечень разделов и тем СР по дисциплине

Раздел/Тема	Вид СР	Форма контроля
Раздел I Тема 2	Подготовка видеореферата на основе презентации	Отправка в ЭИОС отчета установленной формы, файла с выполненным заданием
Раздел V Тема 7	Подготовка видеореферата на основе презентации	Отправка в ЭИОС отчета установленной формы, файла с выполненным заданием
Раздел VI Тема 9	Подготовка видеореферата на основе презентации	Отправка в ЭИОС отчета установленной формы, файла с выполненным заданием
Раздел VII Тема 11	Подготовка видеореферата на основе презентации	Отправка в ЭИОС отчета установленной формы, файла с выполненным заданием

Варианты заданий и методические рекомендации по их выполнению представлены на соответствующем курсе дисциплины в ЭИОС.

Заочная форма обучения характеризуется большой долей самостоятельной работы студента с учебной литературой. Современные тенденции развития информационных технологий позволяют в широкой степени использовать в качестве источника информации сеть Интернет. Использование данных источников при изучении дисциплины приветствуется, но следует иметь в виду, что не вся полученная таким образом информация является достоверной. В первую очередь, следует обратить внимание на список литературы и электронные ресурсы, указанные в соответствующем разделе настоящих методических указаний. Это касается и студентов очной формы обучения.

6 МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ВЫПОЛНЕНИЮ КОНТРОЛЬНОЙ РАБОТЫ (ДЛЯ ЗАОЧНОЙ ФОРМЫ ОБУЧЕНИЯ)

Контрольная работа (КР) является обязательной частью учебного плана для заочной формы обучения и выполняется самостоятельно во внеаудиторное время (время, отводимое на самостоятельную работу).

Для освоения некоторых тем требуется использовать общее или специализированное программное обеспечение. При необходимости можно обратиться на кафедру, где предоставляется место в компьютерном классе для самостоятельной работы. Также на период обучения возможна установка пробных версий программных продуктов с ограниченным функционалом или сроком использования. Большинство предложенных к изучению программ имеют как платные, так и бесплатные версии. Обратитесь за консультацией к своему преподавателю.

В КР предлагается выполнить несколько заданий с индивидуальными вариантами и контрольные вопросы по указанным темам дисциплины. Варианты заданий и методические рекомендации по их выполнению представлены на соответствующем курсе дисциплины в ЭИОС.

Приведенный образец решения не является догмой, допускается использование и других методов решения с их подробным описанием. Если преподаватель не указал иного, то вариант выбирается по последним двум цифрам шифра студента (или номера студенческого билета), а в случае отсутствия такого номера варианта в таблице – по последней цифре.

КР должна быть оформлена в электронном виде с использованием текстового редактора (рекомендуется MS Word или аналогичное офисное

приложение). Обязательно наличие титульного листа, нумерации страниц и оглавления. Файл с КР и вспомогательные файлы (при их наличии) загружаются в облачное хранилище и отправляются в виде ссылки через отдельное задание курса дисциплины в ЭИОС на проверку преподавателю. Регистрация КР проходит в электронном виде. В отдельных случаях по указанию преподавателя КР должна быть распечатана и зарегистрирована на кафедре прикладной математики и информационных технологий Института цифровых технологий.

Цель выполнения КР:

- ⑩ формирование теоретических знаний в области анализа данных и искусственного интеллекта через подготовку реферата по заданной теме, развитие навыков самостоятельного поиска, систематизации и критического осмысления информации;
- ⑩ развитие практических умений использования современных инструментов анализа данных, таких как платформа Loginom Community и MS Excel, для решения задач обработки и визуализации данных;
- ⑩ закрепление навыков применения методов анализа данных к реальным данным, а также формирование умения интерпретировать результаты анализа и делать обоснованные выводы
- ⑩ оценка уровня подготовки студентов в части владения теоретическими основами дисциплины и практическими навыками работы с инструментами анализа данных.

Контрольная работа состоит из двух заданий: теоретического – подготовка реферата, и практического – получение прикладных навыков анализа данных и работы со специализированным программным обеспечением.

Для успешного выполнения контрольной работы необходимо ознакомиться с перечнем основной литературы, а также выполнить практические задания в рамках курса.

Типовой вариант заданий КР представлен в приложении.

7 ТЕКУЩАЯ И ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ДИСЦИПЛИНЕ

Формы текущего контроля

Текущий контроль предназначен для проверки хода и качества усвоения студентами учебного материала и стимулирования их учебной работы. Он может осуществляться в ходе всех видов занятий в форме, избранной преподавателем или предусмотренной рабочей программой дисциплины.

Текущий контроль предполагает постоянный контроль преподавателем качества усвоения учебного материала, активизацию учебной деятельности студентов на занятиях, побуждение их к самостоятельной систематической работе. Он необходим обучающимся для самоконтроля на разных этапах обучения. Результаты учитываются выставлением преподавателем оценок в журнале учета успеваемости и в ходе ежемесячной аттестации.

При текущем контроле успеваемости учитывается выполнение обучающимся всех видов работ и заданий, предусмотренных рабочей программой дисциплины, посещаемость аудиторных занятий (занятий с применением дистанционных образовательных технологий).

К формам текущего контроля успеваемости дисциплины относятся отчеты с ответами на контрольные вопросы по лекциям, практическим занятиям и самостоятельным работам, устный опрос, индивидуальные задания, тестирование.

Оценочными средствами текущего контроля успеваемости являются защита отчетов по лекциям, практическим и самостоятельным работам, тестовые задания открытого и закрытого типов.

Требования к аттестации по дисциплине

Промежуточная аттестация в форме дифференцированного зачета проходит по результатам прохождения всех видов текущего контроля успеваемости. В отдельных случаях (при не прохождении всех видов текущего контроля) зачет может быть проведен в виде тестирования.

К оценочным средствам для промежуточной аттестации в форме экзамена относятся:

- экзаменационные задания по дисциплине, представленные в виде тестовых заданий закрытого и открытого типов;
- контрольная работа (для заочной формы обучения).

Студенты очной формы обучения допускаются к экзамену при положительной аттестации по результатам текущего контроля. Для студентов заочной формы обучения дополнительно необходимо защитить контрольную работу.

На усмотрение преподавателя экзамен может быть проведен в письменной, устной или комбинированной форме, а также в форме тестирования. При наличии сомнений в отношении знаний и умений студента экзаменатор может (имеет право) задать дополнительные вопросы, а также дать дополнительное задание.

Итоговая аттестация по дисциплине проводится в соответствии с действующим Положением о текущем контроле успеваемости и промежуточной аттестации студентов. В случае возникновения непонимания между студентом и ведущим преподавателем, как в период обучения, так и в период зачётно - экзаменационной сессии следует незамедлительно обратиться к заведующему кафедрой прикладной математики и информационных технологий для разрешения спорных ситуаций.

Критерии оценивания результатов освоения дисциплины

Универсальная система оценивания результатов обучения включает в себя системы оценок: 1) «отлично», «хорошо», «удовлетворительно», «неудовлетворительно»; 2) «зачтено», «не зачтено»; 3) 100 – балльную/процентную систему и правило перевода оценок в пятибалльную систему (таблица 4).

Таблица 4 – Система оценок и критерии выставления оценки

Система оценок	0–40 %	41–60 %	61–80 %	81–100 %
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
Критерий	«не зачтено»	«зачтено»		
Системность и полнота знаний в отношении изучаемых объектов	Обладает частичными и разрозненными знаниями, которые не может научно-корректно связывать между собой (только некоторые из которых может связывать между собой)	Обладает минимальным набором знаний, необходимым для системного взгляда на изучаемый объект	Обладает набором знаний, достаточным для системного взгляда на изучаемый объект	Обладает полнотой знаний и системным взглядом на изучаемый объект

Система оценок				
	0–40 %	41–60 %	61–80 %	81–100 %
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	Критерий	«не зачтено»	«зачтено»	
Работа с информацией	Не в состоянии находить необходимую информацию, либо в состоянии находить отдельные фрагменты информации в рамках поставленной задачи	Может найти необходимую информацию в рамках поставленной задачи	Может найти, интерпретировать и систематизировать необходимую информацию в рамках поставленной задачи	Может найти, систематизировать необходимую информацию, а также выявить новые, дополнительные источники информации в рамках поставленной задачи
Научное осмысление изучаемого явления, процесса, объекта	Не может делать научно-корректных выводов из имеющихся у него сведений, в состоянии проанализировать только некоторые из имеющихся у него сведений	В состоянии осуществлять научно-корректный анализ предоставленной информации	В состоянии осуществлять систематический и научно-корректный анализ предоставленной информации, вовлекает в исследование новые релевантные задаче данные	В состоянии осуществлять систематический и научно-корректный анализ предоставленной информации, вовлекает в исследование новые релевантные поставленной задаче данные, предлагает новые ракурсы поставленной задачи
4. Освоение стандартных алгоритмов решения профессиональных задач	В состоянии решать только фрагменты поставленной задачи в соответствии с заданным алгоритмом, не освоил предложенный алгоритм, допускает ошибки	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом, понимает основы предложенного алгоритма	Не только владеет алгоритмом и понимает его основы, но и предлагает новые решения в рамках поставленной задачи

Оценивание тестовых заданий закрытого типа осуществляется по системе зачтено/не зачтено («зачтено» – 41–100 % правильных ответов; «не зачтено» – менее 40 % правильных ответов) или пятибалльной системе (оценка «неудовлетворительно» – менее 40 % правильных ответов; оценка «удовлетворительно» – от 41 до 60 % правильных ответов; оценка «хорошо» – от 61 до 80 % правильных ответов; оценка «отлично» – от 81 до 100 % правильных ответов).

Тестовые задания открытого типа оцениваются по системе «зачтено/ не зачтено».

Оценивается верность ответа по существу вопроса, при этом не учитывается порядок слов в словосочетании, верность окончаний, падежи.

Компетенции в той части, в которой они должны быть сформированы в рамках изучения дисциплины, могут считаться сформированными в случае, если студент получил на экзамене положительную оценку.

8 БИБЛИОГРАФИЧЕСКИЙ СПИСОК

Основная литература

1. Хачумов, М. В. Введение в интеллектуальный анализ данных: учеб. пособие / М. В. Хачумов. – Москва: РТУ МИРЭА, 2023. – 123 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/398240> (дата обращения: 29.05.2024). – ISBN 978-5-7339-2073-3. – Текст : электронный.

2. Алексеев, Д. С. Технологии интеллектуального анализа данных / Д. С. Алексеев, О. В. Щекочихин. – 2-е изд., стер. – Санкт-Петербург: Лань, 2024. – 176 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/362915> (дата обращения: 29.05.2024). – ISBN 978-5-507-48763-9. – Текст: электронный.

3. Основы искусственного интеллекта: учеб. пособие / Ю. А. Антохина, А. А. Оводенко, М. Л. Кричевский, Ю. А. Мартынова. – Санкт-Петербург: ГУАП, 2022. – 169 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/263933> (дата обращения: 21.05.2024). – ISBN 978-5-8088-1720-3. – Текст: электронный.

4. Колмогорова, С. С. Основы искусственного интеллекта: учеб. пособие для студентов / С. С. Колмогорова. – Санкт-Петербург: СПбГЛТУ, 2022. – 108 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/257804> (дата обращения: 24.05.2024). – ISBN 978-5-9239-1308-8. – Текст: электронный.

Дополнительная литература

5. Ланских, Ю. В. Интеллектуальный анализ данных: учеб. пособие / Ю. В. Ланских, В. Г. Ланских. – Киров: ВятГУ, 2023. – 240 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/408569> (дата обращения: 29.05.2024). – Текст: электронный.

6. Кузьмин, В. И. Методы анализа данных: учеб. пособие / В. И. Кузьмин, А. Ф. Гадзаов. – 2-е изд., перераб. и доп. – Москва: РТУ МИРЭА, 2020. – 155 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/171433> (дата обращения: 29.05.2024). – Текст: электронный.

7. Остроух, А. В. Системы искусственного интеллекта: монография / А. В. Остроух, Н. Е. Суркова. – 4-е изд., стер. – Санкт-Петербург: Лань, 2024. – 228 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/379988> (дата обращения: 24.05.2024). – ISBN 978-5-507-47478-3. – Текст: электронный.

8. Искусственный интеллект. Инноватика: учеб. пособие / Ю. А. Антохина, М. Л. Кричевский, Ю. А. Мартынова, А. А. Оводенко. – Санкт-Петербург: ГУАП, 2023. – 320 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/341003> (дата обращения: 21.05.2024). – ISBN 978-5-8088-1830-9. – Текст: электронный.

9. Пенькова, Т. Г. Модели и методы искусственного интеллекта: учеб. пособие / Т. Г. Пенькова, Ю. В. Вайнштейн. – Красноярск: СФУ, 2019. – 116 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/157579> (дата обращения: 21.05.2024). – ISBN 978-5-7638-4043-8. – Текст: электронный.

10. Галыгина, И. В. Основы искусственного интеллекта. Лабораторный практикум / И. В. Галыгина, Л. В. Галыгина. – 2-е изд., стер. – Санкт-Петербург: Лань, 2024. – 364 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/362927> (дата обращения: 21.05.2024). – ISBN 978-5-507-48767-7. – Текст: электронный.

11. Вольфсон, М. Б. Анализ данных: учеб.-метод. пособие / М. Б. Вольфсон. – Санкт-Петербург: СПбГУТ им. М. А. Бонч-Бруевича, 2023. – 69 с. – Режим доступа: для авториз. пользователей. – Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/381533> (дата обращения: 24.05.2024). – Текст: электронный.

12. Газанова, Н. Ш. Методы искусственного интеллекта: учеб.-метод. пособие / Н. Ш. Газанова, С. Н. Никольский. – Москва: РТУ МИРЭА, 2023. – 102 с. – Режим доступа: для авториз. пользователей. – Лань: электронно-библиотечная система. – URL: <https://e.lanbook.com/book/368756> (дата обращения: 24.05.2024). – ISBN 978-5-7339-1805-1. – Текст: электронный.

Электронные ресурсы для изучения дисциплины

1. Национальный портал в сфере искусственного интеллекта:
<https://ai.gov.ru/>

2. Искусственный интеллект Университете 20.35:
<https://ai.2035.university/>
3. LOW-CODE платформа для реализации аналитических процессов:
www.loginom.ru
4. Python: <https://www.python.org/>, <https://pythonworld.ru/>
5. Anaconda: The Most Popular Python Data Science Platform –
<https://www.anaconda.com>
6. Университетская библиотека онлайн – <http://www.biblioclub.ru>
7. ЭБС Лань – <https://e.lanbook.com/>
8. Портал «Калининградский государственный технический университет» – www.klgtu.ru
9. Библиотека КГТУ – www.klgtu.ru/library
10. ЭИОС КГТУ – eios.klgtu.ru
11. Школа больших данных: <https://bigdataschool.ru>
12. Нейронные сети: <https://neurohive.io/>

ПРИЛОЖЕНИЕ

ТИПОВОЙ ВАРИАНТ ЗАДАНИЙ КОНТРОЛЬНОЙ РАБОТЫ (ДЛЯ ЗАОЧНОЙ ФОРМЫ ОБУЧЕНИЯ)

Задание 1. Подготовить реферат по заданной теме (по вариантам)

1. Понятие анализа данных. Задачи анализа данных. Методы анализа данных. Принципы анализа данных.
2. Инструменты анализа данных. Области применения анализа данных.
3. Очистка данных. Преобразование данных. Интеграция данных. Разделение данных. Определение типов данных.
4. Обработка отсутствующих значений. Обработка выбросов. Масштабирование данных. Нормализация данных.
5. Кодирование категориальных признаков. Слияние данных. Удаление дубликатов. Индексирование данных.
6. Визуализация данных. Оценка качества данных. Сохранение предобработанных данных.
7. Линейная регрессия. Модель линейной регрессии. Коэффициенты линейной регрессии.
8. Метод наименьших квадратов.
9. Регрессионный анализ. Критерий оптимальности. Предсказание.
10. Множественная линейная регрессия.
11. Интерпретация результатов. Применение линейной регрессии в реальных задачах.
12. Классификация. Класс. Объект. Признак. Алгоритм.
13. Кластеризация. Разбиение. Дерево решений.
14. Нейронные сети.
15. Метод опорных векторов. Метрики качества. Переобучение. Регуляризация.
16. Иерархическая кластеризация.
17. Алгоритм k-средних.
18. Алгоритм DBSCAN.
19. Проблемы кластеризации и их решения.
20. Применение кластеризации в машинном обучении.
21. Алгоритм Apriori. Алгоритм FP-growth. Параллельные алгоритмы извлечения правил ассоциации.
22. Построение дерева принятия решений на основе правил ассоциации.

23. Искусственные нейронные сети. Математические модели.

Распознавание образов. Классификация данных.

24. Многослойный перцептрон.

25. Сверточные нейронные сети.

26. Компьютерное зрение.

27. Обработка естественного языка.

28. Архитектура нейронных сетей.

29. Распознавание речи.

Задание 2. Использование возможностей платформы Loginom Community и MS Excel для обработки данных

Создать сценарий Loginom, включающий в себя:

- импорт данных (формат *.csv, *.xlsx) на платформу Loginom Community;
- обработку данных в соответствии с заданием (данные о продажах товаров по группам и категориям):

1. Исключить из набора записи с отсутствующими данными.

2. Определить количество продаж товаров по категориям, сумму выручки в каждой категории за определенный период (по вариантам).

3. Рассчитать прибыль ($N\%$ – по вариантам) от продаж определенной (по вариантам) группы товаров за определенный период (по вариантам).

- визуализацию результатов обработки соответствующих узлов сценария;
- экспорт результатов обработки в файл MS Excel.

Локальный электронный методический материал

В. Е. Еремичева

М. А. Романов

АНАЛИЗ ДАННЫХ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

Редактор С. Кондрашова

Корректор Т. Звада

Уч.-изд. л. 5,1. Печ. л. 4,4.

Издательство федерального государственного бюджетного
образовательного учреждения высшего образования
«Калининградский государственный технический университет».
236022, Калининград, Советский проспект, 1