

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Галунин Сергей Александрович
Должность: проректор по учебной работе
Дата подписания: 12.07.2023 11:59:50
Уникальный программный ключ:
08ef34338325bdb0ac5a47baa5472ce36cc3fc3b

Приложение к ОПОП
«Информационные системы и
технологии в управлении цифро-
вым предприятием»



СПбГЭТУ «ЛЭТИ»

ПЕРВЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ

МИНОБРНАУКИ РОССИИ

федеральное государственное автономное образовательное учреждение высшего образования
**«Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И.Ульянова (Ленина)»
(СПбГЭТУ «ЛЭТИ»)**

РАБОЧАЯ ПРОГРАММА

дисциплины

«ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ»

для подготовки магистров

по направлению

09.04.02 «Информационные системы и технологии»

по программе

«Информационные системы и технологии в управлении цифровым
предприятием»

Санкт-Петербург

2022

ЛИСТ СОГЛАСОВАНИЯ

Разработчики:

доцент, к.т.н., доцент Шилов Н.Г.

Рабочая программа рассмотрена и одобрена на заседании кафедры ИС

21.02.2022, протокол № 2

Рабочая программа рассмотрена и одобрена учебно-методической комиссией

ФКТИ, 24.02.2022, протокол № 2

Согласовано в ИС ИОТ

Начальник ОМОЛА Загороднюк О.В.

1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	ИС
Общая трудоемкость (ЗЕТ)	7
Курс	1
Семестр	2

Виды занятий

Лекции (академ. часов)	34
Практические занятия (академ. часов)	51
Иная контактная работа (академ. часов)	1
Все контактные часы (академ. часов)	86
Самостоятельная работа, включая часы на контроль (академ. часов)	166
Всего (академ. часов)	252

Вид промежуточной аттестации

Экзамен (курс) 1

2 АННОТАЦИЯ ДИСЦИПЛИНЫ

«ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ»

Содержание дисциплины охватывает круг вопросов, связанных с интеллектуальным анализом данных, а именно его основными характеристиками и отличительными особенностями, историей, основными преимуществами и ограничениями, используемыми методами, а также прикладными задачами и классами систем, в которых его применение необходимо или позволяет добиться существенного повышения эффективности работы.

SUBJECT SUMMARY

«DATA MINING»

The content of the discipline covers a range of issues related to data mining, namely its main characteristics and distinctive features, history, main advantages and limitations, the methods used, as well as applied tasks and classes of systems in which its use is necessary or allows to achieve a significant increase in efficiency. work.

3 ОБЩИЕ ПОЛОЖЕНИЯ

3.1 Цели и задачи дисциплины

1. Цели дисциплины:

-изучение основ интеллектуального анализа данных, его основных характеристик и отличительных особенностей, его истории, используемых методов, моделей представления данных и технологий;
-формирование навыков и умений использования методов и моделей интеллектуального анализа данных в профессиональной деятельности.

2. Задачами изучения дисциплины является приобретение теоретических знаний в области интеллектуального анализа данных, его основных характеристик и отличительных особенностей, его истории и формирование практических умений и навыков использования методов, моделей представления данных и технологий.

3. Знания сущности интеллектуального анализа данных, его основных характеристик и отличительных особенностей, его истории, используемых методов, моделей представления данных и технологий.

4. Формирование базовых умений в области выбора и применения методов интеллектуального анализа данных для решения практических задач на примерах нейронных сетей, байесовских сетей доверия, корреляционного и регрессионного анализа, классификации и деревьев решений, кластерного анализа, ассоциативных правил и методов визуализации данных.

5. Формирование базовых навыков в области выбора и принципов использования методов для решения различных задач интеллектуального анализа данных и создания информационных систем, использующих интеллектуальных анализ данных.

3.2 Место дисциплины в структуре ОПОП

Дисциплина изучается на основе ранее освоенных дисциплин учебного плана:

1. «Системы поддержки принятия решений»
2. «Специальные главы математики»

и обеспечивает изучение последующих дисциплин:

1. «Междисциплинарный проект ”Разработка сложной информационной системы”»
2. «Производственная практика (научно-исследовательская работа)»
3. «Современные методы и средства проектирования информационных систем»
4. «Производственная практика (преддипломная практика)»

3.3 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения образовательной программы обучающийся должен достичь следующие результаты обучения по дисциплине:

Код компетенции/ индикатора компетенции	Наименование компетенции/индикатора компетенции
ОПК-1	Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте;
<i>ОПК-1.1</i>	<i>Знает математические, естественнонаучные и социально-экономические методы для использования в профессиональной деятельности</i>
<i>ОПК-1.2</i>	<i>Умеет решать нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных, социально-экономических и профессиональных знаний</i>
<i>ОПК-1.3</i>	<i>Имеет навыки теоретического и экспериментального исследования объектов профессиональной деятельности, в том числе в новой или незнакомой среде и в междисциплинарном контексте</i>
ОПК-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями;
<i>ОПК-3.1</i>	<i>Знает принципы, методы и средства анализа и структурирования профессиональной информации</i>
<i>ОПК-3.2</i>	<i>Умеет анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров</i>
<i>ОПК-3.3</i>	<i>Имеет навыки подготовки научных докладов, публикаций и аналитических обзоров с обоснованными выводами и рекомендациями</i>

4 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1 Содержание разделов дисциплины

4.1.1 Наименование тем и часы на все виды нагрузки

№ п/п	Наименование темы дисциплины	Лек, ач	Пр, ач	ИКР, ач	СР, ач
1	История интеллектуального анализа данных.	4			15
2	Использование методов визуализации данных для интеллектуального анализа данных.	4	8		25
3	Корреляционный и регрессионный анализ в интеллектуальном анализе данных.	4	8		25
4	Классификация и деревья решений в интеллектуальном анализе данных.	6	8	1	26
5	Нейронные сети.	6	10		25
6	Байесовские сети доверия и их применение для интеллектуального анализа данных.	4	8		25
7	Кластерный анализ в интеллектуальном анализе данных.	6	9		25
	Итого, ач	34	51	1	166
	Из них ач на контроль	0	0	0	35
	Общая трудоемкость освоения, ач/зе				252/7

4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	История интеллектуального анализа данных.	эволюция подходов к работе с данными, «данные, информация, знания», получение новых знаний, поддержка принятия решений, рекомендующие системы.
2	Использование методов визуализации данных для интеллектуального анализа данных.	возможности визуализации, кластер, тренд, диаграмма рассеивания.
3	Корреляционный и регрессионный анализ в интеллектуальном анализе данных.	линейная аппроксимация, минимизация среднеквадратического отклонения, примеры.
4	Классификация и деревья решений в интеллектуальном анализе данных.	деревья решений как метод классификации, преимущества и недостатки, критерии расщепления, индекс Gini, критерии остановки, сокращение, алгоритм CART.

№ п/п	Наименование темы дисциплины	Содержание
5	Нейронные сети.	нейрон, математическая модель нейрона, функция активации, вычислительная возможность нейрона, многослойные нейронные сети, алгоритм прямого распространения, обучение с учителем, обучение без учителя, алгоритм обратного распространения ошибки, обучение нейрона, области применения, примеры.
6	Байесовские сети доверия и их применение для интеллектуального анализа данных.	элементы теории вероятности, условная вероятность, теорема Байеса, примеры.
7	Кластерный анализ в интеллектуальном анализе данных.	кластеризация, иерархическая кластеризация, кластеризация методами теории графов, алгоритм Краскала, алгоритм k-средних.

4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Использование методов визуализации данных для интеллектуального анализа данных.	8
2. Корреляционный и регрессионный анализ в интеллектуальном анализе данных.	8
3. Классификация и деревья решений в интеллектуальном анализе данных.	8
4. Нейронные сети и их применение для интеллектуального анализа данных.	10
5. Байесовские сети доверия и их применение для интеллектуального анализа данных.	8
6. Кластерный анализ в интеллектуальном анализе данных.	9
Итого	51

4.4 Курсовое проектирование

Курсовая работа (проект) не предусмотрены.

4.5 Реферат

Реферат не предусмотрен.

4.6 Индивидуальное домашнее задание

Индивидуальное домашнее задание не предусмотрено.

4.7 Доклад

Доклад не предусмотрен.

4.8 Кейс

Кейс не предусмотрен.

4.9 Организация и учебно-методическое обеспечение самостоятельной работы

Изучение дисциплины сопровождается самостоятельной работой студентов с рекомендованными преподавателем литературными источниками и информационными ресурсами сети Интернет.

Планирование времени для изучения дисциплины осуществляется на весь период обучения, предусматривая при этом регулярное повторение пройденного материала. Обучающимся, в рамках внеаудиторной самостоятельной работы, необходимо регулярно дополнять сведениями из литературных источников материал, законспектированный на лекциях. При этом на основе изучения рекомендованной литературы целесообразно составить конспект основных положений, терминов и определений, необходимых для освоения разделов учебной дисциплины.

Особое место уделяется консультированию, как одной из форм обучения и контроля самостоятельной работы. Консультирование предполагает особым образом организованное взаимодействие между преподавателем и студентами, при этом предполагается, что консультант либо знает готовое решение, которое он может предписать консультируемому, либо он владеет способами деятель-

ности, которые указывают путь решения проблемы.

Самостоятельное изучение студентами теоретических основ дисциплины обеспечено необходимыми учебно-методическими материалами (учебники, учебные пособия, конспект лекций и т.п.), выполненными в печатном или электронном виде.

Изучение студентами дисциплины сопровождается проведением регулярных консультаций преподавателей, обеспечивающих практические занятия по дисциплине, за счет бюджета времени, отводимого на консультации (внеаудиторные занятия, относящиеся к разделу «Самостоятельные часы для изучения дисциплины»).

Самостоятельной записи на курс нет. Студент заходит на курс, используя логин/пароль от единой учетной записи университета (единий логин и пароль). Каждую неделю будет доступна новая тема курса: видеолекции, кратко раскрывающие содержание каждой темы, презентации и конспекты, с которыми обучающиеся смогут ознакомиться в любое удобное время. Все темы включают практические занятия, которые предусматривают самостоятельное выполнение заданий, а также задания с автоматической проверкой, результаты которых учитываются при общей аттестации полученных знаний. В конце каждой лекции необходимо пройти небольшой контрольный тест, который покажет насколько усвоен предложенный материал. Рекомендуем изучать материал последовательно, что существенно облегчит работу. У каждого контрольного задания имеется своя форма (тест или практическое задание) есть срок выполнения (окончательный срок), по истечении которого даже правильные ответы система принимать не будет! В расписании курса указан окончательный срок каждого задания, который варьируется от двух до четырех недель в зависимости от его сложности.

Текущая СРС	Примерная трудоемкость, ач
Работа с лекционным материалом, с учебной литературой	40

Текущая СРС	Примерная трудоемкость, ач
Опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	15
Самостоятельное изучение разделов дисциплины	20
Выполнение домашних заданий, домашних контрольных работ	0
Подготовка к лабораторным работам, к практическим и семинарским занятиям	56
Подготовка к контрольным работам, коллоквиумам	0
Выполнение расчетно-графических работ	0
Выполнение курсового проекта или курсовой работы	0
Поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	0
Работа над междисциплинарным проектом	0
Анализ данных по заданной теме, выполнение расчетов, составление схем и моделей, на основе собранных данных	0
Подготовка к зачету, дифференцированному зачету, экзамену	35
ИТОГО СРС	166

5 Учебно-методическое обеспечение дисциплины

5.1 Перечень основной и дополнительной литературы, необходимой для освоения дисциплины

№ п/п	Название, библиографическое описание	К-во экз. в библ.
Основная литература		
1	Цехановский, Владислав Владимирович . Интеллектуальный анализ данных [Текст] : учеб. пособие / В. В. Цехановский, В. Д. Чертовской, 2019. - 55 с.	23
2	Элбон, Крис. Машинное обучение с использованием Python. Сборник рецептов [Текст] : [пер. с англ.] / К. Элбон, 2020. -369 с.	20
Дополнительная литература		
1	Интеллектуальный анализ распределенных данных на базе облачных вычислений [Текст] / [М.С. Куприянов [и др.], 2011. -147 с.	9

5.2 Перечень ресурсов информационно-телеинформационной сети «Интернет», используемых при освоении дисциплины

№ п/п	Электронный адрес
1	Нейронные сети и компьютерное зрениe https://stepik.org/course/50352

5.3 Адрес сайта курса

Адрес сайта курса: <https://vec.etu.ru/moodle/course/view.php?id=13088>

6 Критерии оценивания и оценочные материалы

6.1 Критерии оценивания

Для дисциплины «Интеллектуальный анализ данных» формой промежуточной аттестации является экзамен. Оценивание качества освоения дисциплины производится с использованием рейтинговой системы.

Экзамен

Оценка	Количество баллов	Описание
Неудовлетворительно	0 – 39	теоретическое содержание курса не освоено, необходимые практически навыки и умения не сформированы, выполненные учебные задания содержат грубые ошибки, дополнительная самостоятельная работа над курсом не приведет к существенному повышению качества выполнения учебных заданий
Удовлетворительно	40 – 59	теоретическое содержание курса освоено частично, но пробелы не носят существенного характера, необходимые практические навыки и умения работы с освоенным материалом в основном сформированы, большинство предусмотренных программой обучения учебных заданий выполнено, некоторые из выполненных заданий содержат ошибки
Хорошо	60 – 79	теоретическое содержание курса освоено полностью, без пробелов, некоторые практические навыки и умения сформированы недостаточно, все предусмотренные программой обучения учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками
Отлично	80 – 100	теоретическое содержание курса освоено полностью, без пробелов, необходимые практические навыки и умения сформированы, все предусмотренные программой обучения учебные задания выполнены, качество их выполнения оценено количеством баллов, близким к максимальному

Особенности допуска

Промежуточная аттестация проводится на основе мини-тестов после каждой пройденной темы.

К экзамену допускаются студенты, успешно выполнившие мини-тесты.

Экзамен проводится в виде теста по всем темам курса, пример которого приведен в п.6.2

6.2 Оценочные материалы для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине

Вопросы к экзамену

№ п/п	Описание
1	Знания с точки зрения интеллектуальных систем.
2	Примеры использования интеллектуального анализа данных.
3	Что такое точечная диаграмма? Нарисуйте пример точечной диаграммы с линейным трендом.
4	Группировка, кластеризация, классификация. Опишите различия понятий
5	Классификационная модель. Линейное уравнение, являющееся классификационной моделью.
6	Алгоритмы построения деревьев решений
7	Определить значение выходного импульса искусственного нейрона. Входные параметры: $X_1=1$; $X_2=0$. Синаптические коэффициенты: $w_1=0,3$; $w_2=0,7$. Функция активации: единичный скачок, $a=0,5$.
8	Определить, является ли сообщение «надо купить молоко» спамом, если известно, что -вероятность того, что слова «надо», «купить » и «молоко» встречаются в спам-сообщениях, составляет соответственно 0,5, 0,9 и 0,1 -спам сообщения составляют 30% от всех сообщений.
9	Минимальное оставное дерево и его построение.
10	Дайте определение термину "группировка".
11	Дайте определение термину "кластеризация".
12	Дайте определение термину "классификация".
13	Деревья решений как метод классификации. Преимущества и недостатки
14	Критерии расщепления, индекс Gini
15	Критерии остановки, сокращение, алгоритм CART.
16	Опишите математическую модель нейрона
17	Объясните разницу и преимущества обучения с учителем и без учителя
18	Алгоритм обратного распространения ошибки
19	Нарисуйте пример точечной диаграммы с линейным трендом
20	Приведите пример дерева решений

Вариант экзаменационного теста

1. Знания с точки зрения интеллектуальных систем это
 - данные в контексте
 - умения использовать информацию для решения задачи
 - информация, используемая для решения задачи (правильный ответ)
2. Информация отличается от данных
 - наличием методов решения задач
 - наличием логических выражений
 - наличием контекста (правильный ответ)
3. Интеллектуальный анализ данных подразумевает
 - обработку данных для определения функциональных зависимостей
 - обнаружение в наборах данных новых знаний (правильный ответ)
 - создание и изучение информационных систем, способных обучаться
4. Что такое интеллектуальный анализ данных?
 - обнаружение в наборах данных нетривиальных новых знаний, которые являются практически полезными и доступными для интерпретации и обработки с целью принятия решений в различных областях
5. Зачем нужен интеллектуальный анализ данных?
 - для обнаружения новых знаний и закономерностей в целях поддержки принятия решений, использования в рекомендующих системах и т.п.
6. Примером использования интеллектуального анализа данных является
 - выявление закономерности между объемами продаж различных товаров (правильный ответ)
 - игра «Угадай понятие»

- рекомендация сопутствующих товаров при осуществлении покупки в Интернет - магазине

7. Интеллектуальный анализ данных

- основан на использовании только эвристик

- основан на использовании только теоретических методов

- объединяет теоретические и эвристические методы (правильный ответ)

8. Интеллектуальный анализ данных подразумевает

- сбор данных

- обнаружение в наборах данных новых знаний (первый правильный ответ)

- создание и изучение информационных систем, способных обучаться
- выявление неочевидных закономерностей между объемами продаж различных товаров (второй правильный ответ)

- рекомендация сопутствующих товаров при осуществлении покупки в Интернет - магазине

9. Какое из высказываний является верным?

- Машинное обучение в большей степени ориентировано на использование эвристик, чем интеллектуальный анализ данных (правильный ответ)

- Интеллектуальный анализ данных основан на использовании рекомендующих систем

- Системы поддержки принятия решений основаны на использовании статистических методов

10. Какое из высказываний является верным?

- Интеллектуальный анализ данных основан на использовании систем поддержки принятия решений

- Статистика отличается от интеллектуального анализа данных тем, что в ней широко используются эвристики

- В рекомендующих системах используются методы интеллектуального анализа данных для подготовки данных (правильный ответ)

11. Интеллектуальный анализ данных

- основан на использовании теории

- в большей степени, чем статистика использует эвристические методы (правильный ответ)

- преследует те же цели, что и машинное обучение, но использует другие методы

12. Что не является одной из возможностей визуализации?

- Снижение размерности информации

- Восстановление пробелов в наборах данных

- Расчет коэффициента корреляции (правильный ответ)

13. Способ представления количественных данных в виде прямоугольников, площади которых пропорциональны значениям представляемых данных, называется

- Диаграммой рассеяния

- Точечной диаграммой

- Гистограммой (правильный ответ)

14. Нарисуйте пример точечной диаграммы

15. Нарисуйте пример точечной диаграммы с линейным трендом

16. Отнесение объектов (наблюдений, событий) к одной из заранее известных групп называется

- классификацией (правильный ответ)

- кластеризацией

17. Разбиение объектов (наблюдений, событий) на непересекающиеся подмножества так, чтобы элементы одного подмножества были схожими, а разных – существенно различались, называется

- классификацией

- кластеризацией (правильный ответ)

18. Предсказание категориальной зависимой переменной (т.е. зависимой переменной, являющейся категорией) на основе выборки непрерывных и/или категориальных переменных называется

- классификацией (правильный ответ)

- кластеризацией

19. Упорядоченное по некоторому принципу множество объектов, которые имеют сходные признаки (одно или несколько свойств), выбранных для определения сходства или различия между этими объектами, называется

- классификацией

- кластеризацией (правильный ответ)

20. Множество, которое включает данные, использующиеся для построения модели, называется

- обучающим (правильный ответ)

тестовым

непрерывным

21. Множество, которое включает данные, использующиеся для проверки работоспособности модели, называется

- обучающим

- тестовым (правильный ответ)

- множество наблюдений

22. Напишите линейное уравнение, являющееся классификационной моделью для следующего определения классов

$$C1: x < y + 5$$

$$C2: x > y + 5$$

Ответ: $x = y + 5$

23. Напишите линейное уравнение, являющееся классификационной моделью для следующего определения классов

$$C1: x > y$$

$$C2: x < y$$

Ответ: $x = y$

24. Изобразите на диаграмме и напишите линейное уравнение, являющееся классификационной моделью для следующих данных:

X	Y	C
---	---	---

1	2	1
---	---	---

2	3	1
---	---	---

2	1	2
---	---	---

4	3	2
---	---	---

25. Изобразите на диаграмме и напишите линейное уравнение, являющееся классификационной моделью для следующих данных:

X	Y	C
---	---	---

1	2	1
---	---	---

2	3	2
---	---	---

2	1	1
---	---	---

4 3 2

26. Деревья решений могут использоваться для наборов данных, с атрибутами имеющими

- только категориальные значения
- только числовые значения
- как категориальные, так и числовые значения (правильный ответ)

27. Какое из утверждений является истинным

- деревья решений редко используются при решении задач интеллектуального анализа данных
- деревья решений являются одним из наиболее популярных методов решения задач интеллектуального анализа данных (правильный ответ)

28. Какое из утверждений является истинным

- Правила деревьев решений невозможно описать на естественном языке
- Правила деревьев решений достаточно легко описываются на естественном языке (правильный ответ)

29. Деревья решений являются интуитивно понятными и упрощают понимание задачи

- да (правильный ответ)
- нет

30. Какое из утверждений является истинным

- Обучение модели на основе деревьев решений не требует выбора входных атрибутов (правильный ответ)
- Выбор входных атрибутов существенно влияет на время обучения модели на основе деревьев решений

31. Какое из утверждений является истинным

- Большое дерево решений не означает, что оно "подходящее" (правильный ответ)

- Чем больше построенное дерево решений, тем оно лучше

32. Какое из утверждений является истинным

- Модели, основанные на деревьях решений, обладают высокой точностью (правильный ответ)

- Модели, основанные на деревьях решений, обладают низкой точностью

33. Точность моделей, созданных при помощи деревьев решений,

- сопоставима с другими методами построения классификационных моделей (правильный ответ)

- существенно ниже, чем у других методов построения классификационных моделей

34. Какое из утверждений является истинным

- Чем больше построенное дерево решений, тем оно лучше

- Величина дерева решений не всегда является критерием его качества (правильный ответ)

35. Большинство алгоритмов построения деревьев решений

- требуют для своей работы выбор исходных параметров аналитиком

- не требуют для своей работы выбора исходных параметров (правильный ответ)

36. Какое значение параметра X наиболее удачно подходит для расщепления при построении дерева решений?

X Gsplit

1 2,14

2 3,20

4 2,27

5 1,06

37. Какое значение параметра X наиболее удачно подходит для расщепления при построении дерева решений?

X Gsplit

1 2,50

2 2,87

4 2,95

5 3,16

38. Число синапсов (входных значений) нейрона искусственной нейронной сети

- всегда равно 1
- всегда меньше 5
- является произвольным (правильный ответ)

39. Число аксонов (выходных значений) нейрона искусственной нейронной сети

- всегда равно 1 (правильный ответ)
- всегда меньше 5
- является произвольным

40. Определить значение выходного импульса искусственного нейрона

Входные параметры: $X_1=1; X_2=0$.

Синаптические коэффициенты: $w_1=0,3; w_2=0,7$.

Функция активации: единичный скачок, $a=0,5$

41. Определить значение выходного импульса искусственного нейрона

Входные параметры: X1=1; X2=0.

Синаптические коэффициенты: w1=0,7; w2=0,3.

Функция активации: единичный скачок, a=0,5

42. Может ли искусственный нейрон моделировать логическую функцию XOR? Почему?

43. Может ли искусственный нейрон моделировать функцию логического И? Почему?

44. Теорема Байеса представляет интерес

- с точки зрения математического доказательства
- с точки зрения ее интерпретации для задач исследования вероятностей событий (правильный ответ)
- с числовыми значениями переменных

45. Наивный байесовский классификатор подходит для задач интеллектуального анализа

- с категориальными значениями переменных (правильный ответ)
- с числовыми значениями переменных

46. Определить, является ли сообщение «надо купить молоко» спамом, если известно, что

- вероятность того, что слова «надо», «купить » и «молоко» встречаются в спам-сообщениях, составляет соответственно 0,5, 0,9 и 0,1

- спам сообщения составляют 30% от всех сообщений

47. Определить, является ли сообщение «уникальное выгодное предложение» спамом, если известно, что

- вероятность того, что слова «уникальное», «выгодное» и «предложение» встречаются в спам-сообщениях, составляет соответственно 0,5, 0,8 и 0,5

- спам сообщения составляют 30% от всех сообщений

48. Иерархическая кластеризация подразумевает

- построение графов с дугами, веса которых соответствуют расстоянию между объектами

- рассмотрение каждого объекта как кластера с последующим объединением близких кластеров (правильный ответ)

49. Минимальное оствовное дерево может быть построено при использовании

- иерархической кластеризации

- кластеризации методами теории графов (правильный ответ)

50. Алгоритм Краскала

- позволяет находить локальный оптимум кластеризации

- позволяет находить глобальный оптимум кластеризации (правильный ответ)

51. Алгоритм k-средних

- позволяет находить локальный оптимум кластеризации (правильный ответ)

- позволяет находить глобальный оптимум кластеризации

52. Какой из алгоритмов следует использовать, если кластеризация части данных уже была проведена

- алгоритм Краскала

- алгоритм k-средних (правильный ответ)

53. В каком из алгоритмов используется понятие «центра» кластера

- в алгоритме Краскала

- в алгоритме k-средних (правильный ответ)

54. Определить 2 кластера при помощи алгоритма Краскала для следующих точек: A(1;2), B(1;1), C(3;2), D(2;1). В качестве весов ребер использовать расстояния между точками.

55. Определить 2 кластера при помощи алгоритма Краскала для следующих точек: A(1;1), B(1;3), C(2;4), D(2;1). В качестве весов ребер использовать расстояния между точками.

Вариант промежуточного мини-теста по теме "Кластеризация"

1. Иерархическая кластеризация подразумевает

- построение графов с дугами, веса которых соответствуют расстоянию между объектами

- рассмотрение каждого объекта как кластера с последующим объединением близких кластеров (правильный ответ)

2. Минимальное оствовное дерево может быть построено при использовании

- иерархической кластеризации

- кластеризации методами теории графов (правильный ответ)

3. Алгоритм Краскала

- позволяет находить локальный оптимум кластеризации

- позволяет находить глобальный оптимум кластеризации (правильный ответ)

4. Какой из алгоритмов следует использовать, если кластеризация части данных уже была проведена

- алгоритм Краскала

- алгоритм k-средних (правильный ответ)

5. Определить 2 кластера при помощи алгоритма Краскала для следующих точек: A(1;2), B(1;1), C(3;2), D(2;1). В качестве весов ребер использовать расстояния между точками.

Весь комплект контрольно-измерительных материалов для проверки сформированности компетенции (индикатора компетенции) размещен в закрытой части по адресу, указанному в п. 5.3

6.3 График текущего контроля успеваемости

Неделя	Темы занятий	Вид контроля
1	Введение.	Тест
2	История интеллектуального анализа данных.	Тест
3	Использование методов визуализации данных для интеллектуального анализа данных.	Тест
4	Корреляционный и регрессионный анализ в интеллектуальном анализе данных.	Тест
5	Классификация и деревья решений в интеллектуальном анализе данных.	Тест
6	Нейронные сети.	Тест
7	Байесовские сети доверия и их применение для интеллектуального анализа данных.	Тест
8	Кластерный анализ в интеллектуальном анализе данных.	Тест

6.4 Методика текущего контроля

Лекционные занятия

Текущий контроль включает в себя контроль посещаемости (не менее **80** % занятий), по результатам которого студент получает допуск на экзамен.

Практические занятия

Текущий контроль включает в себя контроль посещаемости (не менее **80** % занятий), по результатам которого студент получает допуск на экзамен.

В ходе проведения семинарских и практических занятий целесообразно привлечение студентов к как можно более активному участию в дискуссиях, решении задач, обсуждениях и т. д. При этом активность студентов также может учитываться преподавателем, как один из способов текущего контроля на практических занятиях.

Самостоятельная работа студентов

Контроль самостоятельной работы студентов осуществляется на лекционных и практических занятиях студентов по методикам, описанным выше.

Критерии оценки тестов

Как в экзаменационном, так и в мини-тестах, вопросы имеют различный "вес":

- вопрос, связанный с решением задачи имеет вес 3,
- вопрос, требующий логических рассуждений имеет вес 2,
- вопрос, ответ на который был в явном виде озвучен на лекции, имеет вес 1.

Суммируются баллы, набранные за правильные ответы на вопросы, затем результат нормализуется (например, если набрано 8 баллов из 10 возможных, а максимальное количество баллов за тест - 5, то студент получает за тест 4 балла).

7 Описание информационных технологий и материально-технической базы

Тип занятий	Тип помещения	Требования к помещению	Требования к программному обеспечению
Лекция	Лекционная аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя. Доска, экран, проектор, ПК или ноутбук.	Microsoft PowerPoint, или Adobe Acrobat, или альтернативные средства просмотра файлов презентаций в формате Microsoft PowerPoint или pdf.
Практические занятия	Аудитория	Количество посадочных мест – в соответствии с контингентом, рабочее место преподавателя. Доска, экран, проектор, компьютер.	Наличие Microsoft Excel (или аналог), JupyterHub/JupyterLab (с библиотеками pytorch, matplotlib, pandas, scipy), или доступ к Google Colab, или аналоги.
Самостоятельная работа	Помещение для самостоятельной работы	Оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше

8 Адаптация рабочей программы для лиц с ОВЗ

Адаптированная программа разрабатывается при наличии заявления со стороны обучающегося (родителей, законных представителей) и медицинских показаний (рекомендациями психолого-медико-педагогической комиссии). Для инвалидов адаптированная образовательная программа разрабатывается в соответствии с индивидуальной программой реабилитации.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Дата	Изменение	Дата и номер протокола заседания УМК	Автор	Начальник ОМОЛА