Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Сибирский государственный индустриальный университет»

Кафедра прикладных информационных технологий и программирования

УТВЕРЖДАЮ
Директор института
информационных технологий и
автоматизированных систем
_____ А.В. Зимин
подпись
«____» _____ 20___ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Теория искусственных нейронных сетей и машинное обучение

01.03.02 «Прикладная математика и информатика»
(направленность (профиль): «Математическое моделирование и интеллектуальный анализ данных»)

Квалификация выпускника
Бакалавр

Форма обучения
Очная форма

Срок обучения: 4 года

Год начала подготовки 2025

Новокузнецк
2025

СибГИУ
Подлинник электронного документа хранится в системе электронного согласования документов СибГИУ
ДОКУМЕНТ ПОДПИСАН ЭП
Сертификат: daba56d7b892efca3bf5c04ed0251c98b19b099c
Владелец: Директор ИИТиАС Зимин Алексей Валерьевич
Действителен: с 15 января 2025 г. по 10 апреля 2026 г.
Дата и время: 15-04-2025 10:48:01 ПРОВЕРИТЬ ПОДПИСЬ

1 Цели и задачи освоения учебной дисциплины

Целями учебной дисциплины являются:

– формирование у обучающихся знаний и навыков применения методов машинного обучения для задач управления; моделирования и прогнозирования ML-методами, а также оценки построенной модели.

Задачами учебной дисциплины являются:

– изучение математических основ построения и архитектуры искусственных нейронных сетей;
– изучение различных алгоритмов обучения и настройки ИНС;
– изучение методов сбора и подготовки реальных данных для машинного обучения;
– изучение этапов проектирования признаков для построения моделей машинного обучения;
– изучение процесса построения моделей классификации и регрессии методами ML;
– изучение подходов к оценке построенных моделей.

2 Место учебной дисциплины в структуре ООП по направлению подготовки (специальности)

Учебная дисциплина относится к учебным дисциплинам обязательной части **Блока 1 «Дисциплины (модули)»** ООП по направлению подготовки (специальности) 01.03.02 «Прикладная математика и информатика».

Учебная дисциплина базируется на предварительном усвоении обучающимися учебных дисциплин:

– Базы данных;
– Постреляционные базы данных;
– Теория вероятностей и математическая статистика;
– Численные методы решения инженерных задач;
– Анализ дискретных структур;
– Python для анализа данных;
– Обработка и анализ данных.

Учебная дисциплина дополняет знания и умения, получаемые по одновременно изучаемым и последующим учебным дисциплинам:

– Теория прогнозирования;
– Анализ временных рядов и прогнозирование;
– Распознавание образов.

3 Планируемые результаты обучения по учебной дисциплине

Процесс изучения учебной дисциплины направлен на формирование следующих компетенций:

– Профессиональные компетенции

Наименование категории (группы) ПК	Код и наименование ПК	Код и наименование индикатора достижения ПК	Планируемые результаты обучения
	ПК-2: Способен проводить аналитические исследования с применением технологии больших данных в соответствии с требованиями заказчика	ПК-2.1 Выбирает методы и инструментальные средства анализа больших данных для проведения аналитических работ	– знать: архитектуры нейронных сетей, подходы к обучению ML-моделей, алгоритмы построения классификаторов, регрессоров, кластеров данных и ассоциативных правил, способы визуализации больших объемов данных, типовые диаграммы для представления данных. – уметь: выбирать подход к обучению и алгоритм для ML-моделей, выполнять построение классификаторов, регрессоров, кластеров данных и ассоциативных правил, строить графики и диаграммы, визуализирующие связь между входными признаками и целевой переменной.
		ПК-2.2 Разрабатывает математические модели, выполняет поверку и оценку используемых моделей	– знать: методы сбора данных для машинного обучения, подходы к формированию наборов данных, методы разведочного анализа данных, основные методы выделения признаков, типовые методы моделирования и прогнозирования, методы оптимизации параметров модели, методы и алгоритмы машинного обучения, основные этапы базового рабочего процесса ML. – уметь: формировать признаки для

			решаемых задач; выполнять разведочный анализ данных, подбирать наиболее оптимальный набор входных признаков для модели, оптимизировать параметры модели машинного обучения при помощи сеточного поиска, выполнять построение модели машинного обучения с оценкой прогностической точности.

4 Объем и содержание учебной дисциплины

Учебные занятия по учебной дисциплине проводятся в форме контактной работы и в форме самостоятельной работы обучающихся.

Контактная работа включает в себя занятия лекционного типа (лекции), занятия семинарского типа (семинары, практические занятия, практикумы, лабораторные работы, коллоквиумы), промежуточную аттестацию обучающихся и иные формы взаимодействия обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации ООП на иных условиях, в том числе при проведении промежуточной аттестации обучающихся. Контактная работа может проводиться с применением электронного обучения, дистанционных образовательных технологий.

Объем учебной дисциплины

Семестр / курс		***ИТОГО***	***6 семестр***	***7 семестр***
Форма промежуточной аттестации			*экзамен*	*экзамен*
Трудоёмкость	*академ. час.*	***324***	*144*	*180*
	зачетных единиц	***9***	*4*	*5*
Лекции, *академ. час.*		***32***	*16*	*16*
в форме практической подготовки		***0***	*0*	*0*
Лабораторные работы, *академ. час.*		***16***	*0*	*16*
в форме практической подготовки		***0***	*0*	*0*
Практические занятия, *академ. час.*		***48***	*16*	*32*
в форме практической подготовки		***0***	*0*	*0*
Курсовая работа / проект, *академ. час.*		***0***	*0*	*0*
в форме практической подготовки		***0***	*0*	*0*
Консультации, *академ. час.*		***0***	*0*	*0*
в форме практической подготовки		***0***	*0*	*0*

Самостоятельная работа, *академ. час.*	***138***	*76*	*62*
в форме практической подготовки	***0***	*0*	*0*
Контроль, *академ. час.*	***90***	*36*	*54*
в форме практической подготовки	***0***	*0*	*0*

Содержание учебной дисциплины

Раздел 1 Теоретические основы построения ИНС;

Тема 1.1 Основные понятия (Основные термины и понятия. Понятия искусственного нейрона и нейронной сети. Задачи построения нейронной сети. Этапы построения нейронной сети.);

Тема 1.2 Архитектура классических и современных ИНС (Типы нейронных сетей. Архитектура многослойного персептрона, сети Кохонена, Маркова, Хопфилда, сверточной сети, автоэнкодера, сети глубокого обучения (DCN, DN, DCIGN));

Тема 1.3 Способы обучения ИНС (Подходы к обучению. Обучение нейронной сети без учителя. Методы обучения с учителем. Обучение нейронной сети с подкреплением);

Раздел 2 Сбор и подготовка реальных данных для машинного обучения;

Тема 2.1 Сбор данных (Определение набора входных признаков. Наблюдаемое значение целевой переменной. Достаточный объем обучающих данных. Репрезентативность обучающей выборки);

Тема 2.2 Введение в структуры данных pandas и базовая функциональность (Объект Series. Объект DataFrame. Индексные объекты. Переиндексация. Доступ по индексу, выборка и фильтрация. Арифметические операции и выравнивание данных. Применение функций и отображение. Сортировка и ранжирование);

Тема 2.3 Редукция и вычисление описательных статистик (Корреляция и ковариация. Уникальные значения, счетчики значений и членство. Фильтрация отсутствующих данных);

Тема 2.4 Подготовка данных к моделированию (Категориальные признаки. Отсутствующие данные. Нормализация данных. Переформатирование данных: очистка, преобразование, слияние, изменение формы.);

Тема 2.5 Визуализация данных (API библиотеки matplotlib. Функции построения графиков в pandas. Нанесение данных на карту. Инструментальная экосистема визуализации для Python. Мозаичные диаграммы. Диаграммы размаха. Графики плотности. Диаграммы рассеяния);

Раздел 3 Моделирование и прогнозирование;

Тема 3.1 Основы моделирования с машинным обучением (Поиск связи между входными данными и целевой переменной. Типы методов моделирования. Обучение с учителем и без);

Тема 3.2 Классификация (Построение классификатора и получение предсказаний. Классификация сложных нелинейных данных. Классификация в случае множества классов);

Тема 3.3 Регрессия (Построение регрессора и генерация прогнозов. Регрессия для сложных нелинейных данных);

Тема 3.4 Кластеризация и задача ассоциативных правил (Задача кластеризации. Обнаружение аномалий. Решение задачи ассоциации. Автоэнкодеры.);

Раздел 4 Оценка и оптимизация модели;

Тема 4.1 Оценка прогностической точности на новых данных (Проблема переобучения и чрезмерно оптимистической оценки модели. Скользящий контроль. Перекрестная проверка);

Тема 4.2 Оценка моделей классификации (Точность для отдельных классов и таблица сопряженности. Компромиссы при оценке точности и ROC-кривые. Многоклассовая классификация);

Тема 4.3 Оценка моделей регрессии (Простые показатели эффективности регрессионных моделей. Исследование остатков);

Тема 4.4 Оценка моделей кластеризации (Проверка однородности и полноты кластеров. Метрики качества кластеризации. Метрики эталонных кластеров);

Тема 4.5 Оптимизация модели путем подбора параметров (Параметры настройки ML-алгоритмов. Сеточный поиск);

Раздел 5 Основы проектирования признаков;

Тема 5.1 Основные этапы проектирования признаков (Проектирование признаков и знание предметной области. Этапы проектирования признаков. Извлечение признаков из обычного текста);

Тема 5.2 Агрегирование данных и групповые операции (Механизм GroupBy. Агрегирование данных. Групповые операции и преобразования. Сводные таблицы и кросс-табуляция);

Тема 5.3 Выбор признаков (Прямой отбор и обратное исключение. Отбор признаков для исследования данных. Практический пример отбора признаков).

5 Перечень тем лекций

№ раздела / темы дисциплины	**Темы лекций**	**Трудоемкость, *академ.час***	
		всего	**в форме практической подготовки**
Раздел 1.	Теоретические основы построения ИНС		
Тема 1.1.	Основные понятия	2	
Тема 1.2.	Архитектура классических и современных ИНС	2	
Тема 1.3.	Способы обучения ИНС	4	
Раздел 2.	Сбор и подготовка реальных данных для		

	машинного обучения		
Тема 2.1.	Сбор данных	4	
Тема 2.4.	Подготовка данных к моделированию	2	
Тема 2.5.	Визуализация данных	2	
Раздел 3.	Моделирование и прогнозирование		
Тема 3.1.	Основы моделирования с машинным обучением	2	
Тема 3.2.	Классификация	2	
Тема 3.3.	Регрессия	2	
Тема 3.4.	Кластеризация и задача ассоциативных правил	1	
Раздел 4.	Оценка и оптимизация модели		
Тема 4.1.	Оценка прогностической точности на новых данных	2	
Тема 4.2.	Оценка моделей классификации	1	
Тема 4.3.	Оценка моделей регрессии	1	
Тема 4.4.	Оценка моделей кластеризации	1	
Тема 4.5.	Оптимизация модели путем подбора параметров	2	
Раздел 5.	Основы проектирования признаков		
Тема 5.1.	Основные этапы проектирования признаков	1	
Тема 5.3.	Выбор признаков	1	
Итого:		**32**	**0**

6 Перечень тем практических занятий (семинаров)

№ раздела / темы дисциплины	**Темы практических занятий (семинаров)**	**Трудоемкость, *академ.час***	
		всего	**в форме практической подготовки**
Тема 2.1.	Сбор и подготовка данных	4	
Тема 2.2.	Работа с объектами Series и DataFrame	2	
Тема 2.3.	Редукция и вычисление описательных статистик в Jupyter Notebook	2	
Тема 2.4.	Работа с различными типами данных в датасете	4	
Тема 2.5.	Визуализация данных	4	

Тема 3.1.	Поиск связи между входными данными и целевой переменной	4	
Тема 4.1.	Оценка прогностической точности на новых данных	4	
Тема 4.2.	Оценка моделей классификации	4	
Тема 4.3.	Оценка моделей регрессии	4	
Тема 4.4.	Оценка моделей кластеризации	4	
Тема 4.5.	Оптимизация модели путем подбора параметров	4	
Тема 5.1.	Этапы проектирования признаков	2	
Тема 5.2.	Агрегирование данных в Jupyter Notebook	4	
Тема 5.3.	Прямой отбор и обратное исключение признаков для набора данных	2	
Итого:		**48**	**0**

7 Перечень тем лабораторных работ

№ раздела / темы дисциплины	**Темы лабораторных работ**	**Трудоемкость, *академ.час***	
		всего	**в форме практической подготовки**
Раздел 3.	Решение задачи классификации при помощи библиотеки Scikit-learn	4	
Раздел 3.	Решение задачи регрессии Scikit-learn	4	
Раздел 3.	Решение задачи кластеризации Scikit-learn	4	
Раздел 3.	Решение задачи поиска ассоциативных правил Scikit-learn	4	
Итого:		**16**	**0**

8 Перечень тем курсовых работ (проектов)

№ раздела / темы дисциплины	**Темы курсовых работ (проектов)**	**Трудоемкость, *академ.час***	
		всего	**в форме практической подготовки**
	Отсутствуют		
Итого:		**0**	**0**

9 Виды самостоятельной работы

№ раздела / темы дисциплины	Виды самостоятельной работы	Трудоемкость, *академ.час*	
		всего	**в форме практической подготовки**
Раздел 1.	1. Изучение лекционного материала; 2. Прохождение тестирования.	32	
Раздел 2.	1. Изучение лекционного материала; 2. Оформление отчета о практической работе; 3. Подготовка к практическому занятию; 4. Прохождение тестирования.	44	
Раздел 3.	1. Изучение лекционного материала; 2. Оформление отчета о практической работе; 3. Оформление отчета по лабораторной работе; 4. Подготовка к лабораторной работе; 5. Подготовка к практическому занятию; 6. Прохождение тестирования.	32	
Раздел 4.	1. Изучение лекционного материала; 2. Оформление отчета о практической работе; 3. Подготовка к практическому занятию; 4. Прохождение тестирования.	16	
Раздел 5.	1. Изучение лекционного материала; 2. Оформление отчета о практической работе; 3. Подготовка к практическому занятию; 4. Прохождение тестирования.	14	
Контроль	*Подготовка к экзамену (6 семестр)*	36	
Контроль	*Подготовка к экзамену (7 семестр)*	54	
Итого:		**228**	**0**

10 Матрица компетентностных задач по учебной дисциплине

Контролируемые разделы (темы) дисциплины	Контролируемые компетенции	Оценочные средства
Раздел 1.	ПК-2	1. Тестовые задания
Раздел 2.	ПК-2	1. Наблюдение за работой на практическом занятии; 2. Отчет о практической работе; 3. Тестовые задания
Раздел 3.	ПК-2	1. Наблюдение за работой на лабораторной работе; 2. Наблюдение за работой на практическом занятии; 3. Отчет о лабораторной работе; 4. Отчет о практической работе; 5. Тестовые задания
Раздел 4.	ПК-2	1. Наблюдение за работой на практическом занятии; 2. Отчет о практической работе; 3. Тестовые задания
Раздел 5.	ПК-2	1. Наблюдение за работой на практическом занятии; 2. Отчет о практической работе; 3. Тестовые задания
Экзамен (6 семестр).	ПК-2	1. Вопросы к экзамену; 2. Задачи к экзамену
Экзамен (7 семестр).	ПК-2	1. Вопросы к экзамену; 2. Задачи к экзамену

11 Входной контроль

11.1 Цель входного контроля − определить начальный уровень подготовленности обучающихся и выстроить индивидуальную траекторию обучения. В условиях личностно-ориентированной образовательной среды результаты входного оценивания обучающегося используются как начальные значения в индивидуальном профиле академической успешности обучающегося.

11.2 Описание оценочных средств

Форма проведения входного контроля – тестирование.

Длительность тестирования – 30 минут.

Тестирование проводится в течение первых двух недель изучения учебной дисциплины в СУО «Moodle».

Для проведения тестирования используются тестовые задания для входного контроля (для ответа на 1 вопрос, как правило, выделяется 1 минута).

Шкала оценивания результатов тестирования

За каждый правильный ответ выставляется 1 балл. Оценка выставляется в зависимости от процента правильных ответов:

Оценка формируется в соответствии с критериями таблицы:

Процент результативности (правильных ответов)	Оценка
86 до 100	«отлично»
от 67до 85	«хорошо»
от 50 до 66	«удовлетворительно»
менее 50	«неудовлетворительно»

Основываясь на результатах тестирования, формируется индивидуальная траектория обучающегося (группы обучающихся), исходя из следующих соображений:

– 67 – 100 % – обучающийся готов к изучению учебной дисциплины без дополнительного повторения изученного ранее материала;

– 50 – 66 % – обучающийся в целом готов к изучению учебной дисциплины, требуется повторение отдельных вопросов(тем) изученного ранее материала, необходимого для успешного освоения учебной дисциплины;

– 49% и менее – обучающийся имеет слабую подготовку к изучению учебной дисциплины, необходимы корректирующие занятия и консультации.

Тестовые задания для входного контроля знаний обучающихся приведены в приложении.

12 Текущий контроль

12.1 Текущий контроль знаний используется для оперативного и регулярного управления учебной деятельностью (в том числе самостоятельной) обучающихся. Данный вид контроля стимулирует у обучающихся стремление к систематической самостоятельной работе по изучению учебной дисциплины и позволяет оценить степень освоения изученного материала.

Текущий контроль успеваемости осуществляется в течение семестра, в ходе повседневной учебной работы по индивидуальной инициативе педагогического работника, в том числе посредством использования СУО «Moodle».

Формы проведения текущего контроля – наблюдение за работой на лабораторной работе; наблюдение за работой на практическом занятии;

отчет о лабораторной работе; отчет о практической работе; тестовые задания.

12.2 Описание оценочных средств

Шкалы оценивания (методики оценки)

12.2.1 Рекомендации по оцениванию результатов тестирования обучающихся

Тестирование проводится по изученным разделам (темам) учебной дисциплины для выявления уровня знаний и умений обучающегося, мотивирования обучающегося к активизации самостоятельной работы по усвоению учебного материала, выявления и устранения пробелов в знаниях, повышения дисциплины и организации учебной деятельности обучающегося.

Тестирование проводится в СУО «Moodle» в рамках часов, отведенных на самостоятельную работу обучающихся. Для тестирования используется 20 вопросов из базы типовых тестовых заданий по изученным разделам (темам) учебной дисциплины. Обучающийся выбирает один ответ из предложенных вариантов ответа. В случае выбора неправильного ответа обучающийся получает за данный вопрос 0 баллов; в случае правильного ответа – 1 балл.

Оценка формируется в соответствии с критериями таблицы:

Максимальный балл	Проходной балл	Оценка
20	не менее 18	отлично
17	не менее 14	хорошо
13	не менее 10	удовлетворительно
9	–	неудовлетворительно

12.2.2 Рекомендации по оцениванию в процессе наблюдения за работой на практическом занятии

Наблюдение за работой на практическом занятии – важнейший компонент контроля деятельности обучающихся и качества их обучения.

Критерии оценки в процессе наблюдения за работой на практическом занятии:

– самостоятельность выполнения задания и работы с методическими указаниями (учитывается работа в течение занятия, быстрота и способность нахождения в методических указаниях нужной информации);

– правильность выполнения работы (учитывается последовательность выполняемых действий, умение работать с инструментами и/или средствами вычислительной техники);

– рациональность использования времени, отведенного на задание (не одобряется затянутость выполнения работы с учетом индивидуальных особенностей обучающихся);

– активность обучающегося при выполнении практической работы (учитывается индивидуальная работа и работа в группе).

Критерии оценки по результатам наблюдения за работой на практическом занятии

Наименование параметра оценки	Максимальное количество баллов
Самостоятельно выполняет задания; умеет работать с методическими указаниями.	2,5
Указанные в методических указаниях действия выполняет в правильной последовательности; умеет работать с инструментами и/или средствами вычислительной техники.	1,25
Рационально использует время, отведенное на задание; проявляет активность при выполнении практической работы.	1,25
Итого	**5**

Максимальное количество – 5 баллов соответствует оценке «отлично», 4 балла – «хорошо», 3 балла – «удовлетворительно», 2 балла – «неудовлетворительно».

12.2.3 Рекомендации по оцениванию отчета о практической работе

Оформление отчета о проделанной практической работе обучающиеся осуществляют в рамках самостоятельной работы.

Критерии оценки отчета о практической работе:

– правильность выполнения работы (учитывается полнота отчета о практической работе, наличие всех необходимых данных, наличие выводов, наличие ошибок и неточностей);

– аккуратность оформления отчета о практической работе.

Критерии оценки отчета о практической работе

Наименование параметра оценки	Максимальное количество баллов
Отчет о практической работе содержит все необходимые данные и выводы.	2,5
Отчет о практической работе не содержит ошибок и неточностей.	1,25
Отчет о практической работе оформлен аккуратно.	1,25
Итого	**5**

Максимальное количество – 5 баллов соответствует оценке «отлично», 4 балла – «хорошо», 3 балла – «удовлетворительно», 2 балла – «неудовлетворительно».

12.2.4 Рекомендации по оцениванию в процессе наблюдения за работой на лабораторном занятии

Наблюдение за работой на лабораторном занятии – важнейший компонент контроля деятельности обучающихся и качества их обучения.

Критерии оценки в процессе наблюдения за работой на лабораторном занятии:

– самостоятельность выполнения задания и работы с методическими указаниями (учитывается работа в течение занятия, быстрота и способность нахождения в методических указаниях нужной информации);

– правильность выполнения работы (учитывается последовательность выполняемых действий, умение работать с инструментами и/или средствами вычислительной техники);

– рациональность использования времени, отведенного на задание (не одобряется затянутость выполнения работы с учетом индивидуальных особенностей обучающихся);

– активность обучающегося при выполнении лабораторной работы (учитывается индивидуальная работа и работа в группе).

Критерии оценки по результатам наблюдения за работой на лабораторном занятии

Наименование параметра оценки	Максимальное количество баллов
Самостоятельно выполняет задания; умеет работать с методическими указаниями.	2,5
Указанные в методических указаниях действия выполняет в правильной последовательности; умеет работать с инструментами и/или средствами вычислительной техники.	1,25
Рационально использует время, отведенное на задание; проявляет активность при выполнении лабораторной работы.	1,25
Итого	**5**

Максимальное количество – 5 баллов соответствует оценке «отлично», 4 балла – «хорошо», 3 балла – «удовлетворительно», 2 балла – «неудовлетворительно».

12.2.5 Рекомендации по оцениванию отчета о лабораторной работе

Оформление отчета о проделанной лабораторной работе обучающиеся осуществляют в рамках самостоятельной работы.

Критерии оценки отчета о лабораторной работе:

– правильность выполнения работы (учитывается полнота отчета о лабораторной работе, наличие всех необходимых данных, наличие выводов, наличие ошибок и неточностей);

– аккуратность оформления отчета о лабораторной работе.

Критерии оценки отчета о лабораторной работе

Наименование параметра оценки	Максимальное количество баллов
Отчет о лабораторной работе содержит все необходимые данные и выводы.	2,5
Отчет о лабораторной работе не содержит ошибок и неточностей.	1,25
Отчет о лабораторной работе оформлен аккуратно.	1,25
Итого	**5**

Максимальное количество – 5 баллов соответствует оценке «отлично», 4 балла – «хорошо», 3 балла – «удовлетворительно», 2 балла – «неудовлетворительно».

Задания и вопросы для текущего контроля знаний обучающихся по разделам (темам) учебной дисциплины приведены в приложении.

13 Промежуточная аттестация

13.1 Оценочные материалы для промежуточной аттестации обучающихся по учебной дисциплине предназначены для оценки степени достижения запланированных результатов обучения по завершению изучения учебной дисциплины в установленной учебным планом форме и позволяют определить качество усвоения изученного материала.

Итоговой формой контроля сформированности компетенций у обучающихся по учебной дисциплине является: экзамен (6 семестр); экзамен (7 семестр).

Оценочные материалы для промежуточной аттестации обучающихся включают: вопросы к экзамену; задачи к экзамену.

13.2 Оценивание обучающегося на экзамене (6 семестр)

Оценка	Требования к знаниям
отлично	Оценка «отлично» выставляется обучающемуся, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, использует в ответе материал различной литературы, правильно обосновывает принятое нестандартное решение, обладает приемами выполнения практических задач по формированию компетенций.
хорошо	Оценка «хорошо» выставляется обучающемуся, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, обладает необходимыми

Оценка	Требования к знаниям
	умениями и приемами их выполнения.
удовлетворительно	Оценка «удовлетворительно» выставляется обучающемуся, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, испытывает сложности при выполнении практических работ и затрудняется связать теорию вопроса с практикой.
неудовлетворительно	Оценка «неудовлетворительно» выставляется обучающемуся, который не знает значительной части программного материала, неуверенно отвечает, допускает серьезные ошибки, не имеет представлений по методике выполнения практической работы. Как правило, оценка «неудовлетворительно» ставится обучающимся, которые не могут продолжить обучение без дополнительных занятий по данной учебной дисциплине.

13.3 Оценивание обучающегося на экзамене (7 семестр)

Оценка	Требования к знаниям
отлично	Оценка «отлично» выставляется обучающемуся, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, использует в ответе материал различной литературы, правильно обосновывает принятое нестандартное решение, обладает приемами выполнения практических задач по формированию компетенций.
хорошо	Оценка «хорошо» выставляется обучающемуся, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, обладает необходимыми умениями и приемами их выполнения.
удовлетворительно	Оценка «удовлетворительно» выставляется обучающемуся, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, испытывает сложности при выполнении практических работ и затрудняется связать теорию вопроса с практикой.
неудовлетворительно	Оценка «неудовлетворительно» выставляется обучающемуся, который не знает значительной части программного материала, неуверенно отвечает, допускает серьезные ошибки, не имеет представлений по методике выполнения практической работы. Как правило, оценка «неудовлетворительно» ставится обучающимся, которые не могут продолжить обучение без дополнительных занятий по данной учебной дисциплине.

Задания и вопросы для промежуточной аттестации обучающихся приведены в приложении.

14 Учебно-методическое и информационное обеспечение учебной дисциплины

а) литература:

1 Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебное пособие / П. Флах. – Москва : ДМК-пресс, 2015. – 400 с. – ISBN 978-5-97060-273-7. – URL: https://www.studentlibrary.ru/book/ISBN9785970602737.html (дата обращения: 10.04.2025);

2 Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения : практическое пособие / С. Рашка. – Москва : ДМК-пресс, 2017. – 418 с. – ISBN 978-5-97060-409-0. – URL: https://www.studentlibrary.ru/book/ISBN9785970604090.html (дата обращения: 10.04.2025);

3 Миркин, Б. Г. Введение в анализ данных : учебник и практикум. – Москва : Юрайт, 2024. – 174 с. – ISBN 978-5-9916-5009-0. – URL: https://urait.ru/bcode/536117 (дата обращения: 10.04.2025);

4 Маккинли, У. Python и анализ данных : практическое пособие / У. Маккинли. – Москва : ДМК Пресс, 2015. - 482 с. – ISBN 978-5-97060-315-4. – URL: https://www.studentlibrary.ru/ru/book/ISBN9785970603154.html (дата обращения: 10.04.2025).

б) ресурсы информационно-телекоммуникационной сети «Интернет»:

1 Электронный каталог : сайт / Научно-техническая библиотека СибГИУ. – Новокузнецк, [199 –]. – URL: http://libr.sibsiu.ru. – URL: https://libr.sibsiu.ru;

2 Электронная библиотека // Научно-техническая библиотека СибГИУ : сайт. – Новокузнецк, [200 –]. – URL: http://library.sibsiu.ru/LibrELibraryFullText.asp. – Режим доступа: для авторизир. пользователей. – URL: https://library.sibsiu.ru/LibrELibraryFullText.asp;

3 Консультант студента : электронно-библиотечная система / ООО «КОНСУЛЬТАНТ СТУДЕНТА». – Москва, [200 –]. – URL: http://www.studentlibrary.ru. – Режим доступа: для авторизир. пользователей;

4 ЛАНЬ : электронно-библиотечная система : [коллекция «Единая профессиональная база знаний для технических вузов - Издательство Лань" ЭБС»] / ООО «Издательство ЛАНЬ». – Санкт-Петербург, [200 –]. – URL: http://e.lanbook.com. – Режим доступа: для авторизир. пользователей;

5 Образовательная платформа ЮРАЙТ / ООО «Электронное издательство ЮРАЙТ». – Москва, [200 –]. – URL: https://urait.ru. – Режим доступа: для авторизир. пользователей;

6 Университетская библиотека онлайн : электронно-библиотечная система / ООО «Директ-Медиа». – Москва, [200 –]. – URL: https://biblioclub.ru. – Режим доступа: для авторизир. пользователей. – URL: http://www.biblioclub.ru;

7 НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА eLIBRARY.RU : база данных / ООО «НЭБ». – Москва, [200 –]. – URL: http://elibrary.ru. – Режим доступа: по подписке;

8 Электронные периодические издания ИВИС : универсальная база данных / ООО «ИВИС». – Москва, [200 –]. – URL: http://eivis.ru. – Режим доступа: по подписке.

в) лицензионное и свободно распространяемое программное обеспечение:

– 7-Zip;
– Adobe Acrobat Reader;
– Astra Linux Special Edition;
– Kaspersky Endpoint Security;
– Microsoft Office;
– Microsoft Windows;
– Notepad++;
– OnlyOffice;
– PyCharm;
– Р7-Офис.

г) базы данных и информационно-справочные системы:

1 ГАРАНТ : справочно-правовая система / ООО «Правовой центр «Гарант». – Кемерово, [200 –]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.;

2 КонсультантПлюс : справочно-правовая система / ООО «Информационный центр АНВИК». – Новокузнецк, [199 –]. – Режим доступа: компьютерная сеть библиотеки Сиб. гос. индустр. ун-та.;

3 Техэксперт : информационно-справочная система / ООО «Группа компаний «Кодекс». – Кемерово, [200 –]. – Режим доступа: компьютерная сеть Сиб. гос. индустр. ун-та.

15 Материально-техническое обеспечение учебной дисциплины

Материально-техническое обеспечение учебной дисциплины включает учебные аудитории, оснащенные оборудованием, компьютерной техникой, и техническими средствами обучения, в том числе:

- учебную аудиторию для проведения занятий лекционного типа, оборудованную учебной доской, экраном и мультимедийным проектором;

- учебную аудиторию для проведения занятий семинарского типа (практических занятий),оснащенную персональными компьютерами и выходом в сеть «Интернет»;
- учебную аудиторию для проведения занятий семинарского типа (лабораторных работ), оснащенную персональными компьютерами и выходом в сеть «Интернет»;
- учебную аудиторию (помещения) для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду, научно-техническую библиотеку СибГИУ.

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС ВО по направлению подготовки (специальности) 01.03.02 «Прикладная математика и информатика».

Составитель(и):

доцент Бабичева Надежда Борисовна (кафедра прикладных информационных технологий и программирования).

Рабочая программа дисциплины рассмотрена и утверждена на заседании кафедры.

Согласована:

Системный аналитик группы
системной архитектуры департамента
информационных технологий
ООО "СИ СОФТ" Е.А. Шапошникова

Приложение

Тестовые задания для входного контроля знаний обучающихся

Выберите один правильный ответ.

1 Выберите определение термина «Данные».

a. Сведения об окружающем мире, определенные только в рамках какой-либо предметной области;

b. Информация, фиксированная в определенной форме, пригодной для последующей обработки, хранения и передачи;

c. Сведения об окружающем мире, протекающих процессах и явлениях, рассматриваемые с учетом некоторой их семантической сущности;

d. Сведения об окружающем мире, определенным образом истолкованные, обеспечивающие достижение некоторой цели и структуры.

2 Система управления базой данных (СУБД) -это:

a. служебная информация, содержащая упорядоченные сведения о ключевых значениях;

b. комплекс программных и языковых средств, необходимых для создания и модификации базы данных;

c. поименованная совокупность структурированных данных, относящихся к определенной предметной области.

3 Перечислите все известные типы аномалий:

a. Модификации;

b. Перемещения;

c. Вставки;

d. Удаления.

4 Укажите требование к первой нормальной форме:

a. Каждый неключевой атрибут функционально зависит от первичного ключа;

b. Отсутствие транзитивных зависимостей неключевых атрибутов от первичного ключа;

c. Каждый детерминант является ключом-кандидатом;

d. Каждая таблица является отношением.

5 Транзитивная зависимость – это:

a. Прямая зависимость неключевых атрибутов от первичного ключа;

b. Зависимость неключевых атрибутов от первичного ключа через другой атрибут;

c. Прямая зависимость неключевых атрибутов от уникального ключа;

d. Зависимость неключевых атрибутов от уникального ключа через другой атрибут;

e. Зависимость внешнего ключа от первичного ключа.

6 Многозначная зависимость – это:

a. Зависимость, когда, нескольким значениям кортежа соответствует много значений атрибутов;

b. Зависимость, когда, одному значению кортежа соответствует много значений атрибутов;

c. Зависимость, когда, одному значению атрибута соответствует много значений кортежа;

d. Зависимость, когда, одному значению одного атрибута может соответствовать много значений другого атрибута.

7 Отличительной особенностью каких языков программирования является их ориентация не на систему команд той или иной ЭВМ, а на систему операторов, характерных для записи определенного класса алгоритмов?

a. языков программирования низкого уровня;

b. языков программирования высокого уровня;

c. языков программирования сверхвысокого уровня;

d. правильных ответов нет.

8 Что понимают под языком программирования (ЯП)?

a. язык, предназначенный для решения определенного класса задач (проблем);

b. правила представления данных и записи алгоритмов их обработки, которые автоматически выполняются ЭВМ;

c. язык, предназначенный для создания пакетов прикладных программ, в том числе для современных операционных систем;

d. правильных ответов нет.

9 Как называется именованная спецификация одного или более столбцов (для каждого столбца указывается имя, а также его тип или домен)?

a. строчный тип данных;

b. объектный тип данных;

c. комбинированный тип данных;

d. правильных ответов нет.

10 Что понимается под наследованием типов?

a. однородная масса разрядов, имеющая какую-либо структуру;

b. возможность дисциплинированного создания новых типов на основе уже определенных;

c. многовходовой программный модуль, точки входа которого соответствуют набору операций реализуемого типа;

d. правильных ответов нет.

11 Что из перечисленного не относится к наиболее распространенным конструируемым типам данных?

a. тип записи;

b. тип множества;

c. тип массива;

d. тип распределения;

e. правильных ответов нет.

12 Что понимается под классом в объектно-ориентированном программировании (ООП)?

a. процедуры и функции любого языка программирования;

b. структурный тип данных, который включает описание полей данных, процедур и функций, работающих с этими полями данных;

c. списки передаваемых параметров, типы функции;

d. правильных ответов нет.

13 Какие типы в языке C# относятся к значимым:

a. логический;

b. арифметический;

c. структуры;

d. перечисление;

e. массивы;

f. строки;

g. классы;

i. правильных ответов нет.

14 Определение класса в C# и других объектных языках содержит:

a. вершины;

b. методы;

c. приключения;

d. равнины;

e. поля;

f. события;

g. правильных ответов нет.

15 Какие типы в языке C# относятся к ссылочным:

a. логический;

b. арифметический;

c. структуры;

d. перечисление;

e. массивы;

f. строки;

g. классы;

i. правильных ответов нет.

16 Что называют статистической гипотезой?

a. предположение о теоретических значениях параметров распределения;

b. утверждение о виде неизвестного распределения;

c. предположение о виде неизвестного распределения или о параметрах известного распределения;

d. утверждение о параметрах известного распределения.

17 Статистические характеристики – это

a. закон распределения случайной величины;

b. числовые параметры, которые описывают случайную величину полностью;

c. числовые параметры, которые характеризуют степень разбросанности значений случайной величины;

d. числовые параметры, которые отражают существенные черты случайной величины;

e. ничего из перечисленного;

f. все перечисленное.

18 Какие из приведенных высказываний являются свойствами дисперсии?

a. дисперсия произведения независимой случайной величины на постоянную равна произведению дисперсии случайных величины на квадрат постоянной;

b. дисперсия суммы независимых случайных величин равно сумме дисперсий этих случайных величин;

c. дисперсия постоянной равна нулю;

d. дисперсия произведения случайных величин равно произведению дисперсий этих случайных величин;

e. дисперсия суммы случайных величин равно сумме дисперсий этих случайных величин.

19 Коэффициент взаимной корреляции r_x=0,9 – это означает, что

a. между двумя отсчетами одной случайно величины есть сильная линейная связь;

b. между двумя случайными величинами линейная связь отсутствует;

c. между двумя случайными величинами есть сильная линейная связь;

d. между двумя отсчетами одной случайно величины линейная связь отсутствует;

e. между двумя случайными величинами есть сильная нелинейная связь;

f. случайные величины независимы.

20 Обучающая выборка должна

a. быть репрезентативной;

b. удовлетворять условиям качества;

c. быть разбита на две части – обучающую и контрольную;

d. отражать правила и закономерности исследуемого процесса;

e. быть закодирована на машинном языке.

Задания и вопросы для текущего контроля знаний обучающихся по разделам (темам) учебной дисциплины

1 Тестирование по результатам освоения лекционного материала и самостоятельной работы

Тестирование проводится для определения степени усвоения материалов лекций, а также выполнения самостоятельной работы в

рамках часов, отведенных на самостоятельную работу обучающихся, в СУО «Moodle».

Раздел 1. Теоретические основы построения ИНС

Перечень вопросов для тестирования:

1 Логистическая регрессия – это...

Выберите один ответ:

a. модель, в которой предполагаем, как влияет регуляризатор на аппроксимируемую функцию.

b. модель, в которой напрямую делаем предположение о том, как выглядит аргумент сигмоида a.

c. модель, в которой обязательно используется сигмоид a.

d. модель, в которой мы напрямую делаем предположение о том, как выглядит функция ошибки.

2 Что такое условная вероятность?

Выберите один ответ:

a. вероятность одновременного наступления двух событий

b. способ получить обратно из совместной вероятности вероятность того или иного исхода одной из случайных величин

c. отношение «причина–следствие»

d. вероятность наступления одного события, если известно, что произошло другое

3 В машинном обучении нейрон - это...

Выберите один ответ:

a. вычислительная единица, которая получает информацию, производит над ней простые вычисления и передает ее дальше

b. алгоритм вычисления самой функции или производной по одной переменной

c. способ нормализации входных данных

d. связь между двумя функциональными единицами сети

4 В тех случаях когда возникает необходимость оценить вероятность событий, к которым неприменимы подходы классической теории вероятностей (провести эксперименты и подсчитать вероятность) в роли чего выступают вероятности?

Выберите один ответ:

a. степеней доверия

b. вероятностей

c. степеней свободы

5 Вероятностный вывод - это?

Выберите один ответ:

a. вывод из условий задачи, в котором данные используются, чтобы вывести вероятность того, что гипотеза может быть верной

b. вывод из условий задачи, в котором определяется совместная вероятность возникновения двух событий

c. вывод из условий задачи, в котором определяется вероятность возникновения того или иного события

d. вывод из условий задачи, в котором определяется условная независимость двух произошедших событий

6 У каждого из нейронов есть 2 основных параметра. Какие?

Выберите один ответ:

a. смещенные данные

b. входные данные

c. выходные данные

d. скрытые данные

7 Вероятность одновременного наступления двух событий, p(x; y) - это...

Выберите один ответ:

a. условная независимость

b. маргинализация

c. вероятность

d. условная вероятность

e. совместная вероятность

8 Что такое маргинализация?

Выберите один ответ:

a. вероятность наступления одного события, если известно, что произошло другое

b. отношение «причина–следствие»

c. вероятность одновременного наступления двух событий

d. способ получить обратно из совместной вероятности вероятность того или иного исхода одной из случайных величин

9 При помощи какого нейрона производится сдвиг графика функции активации вправо или влево?

Выберите один ответ:

a. скрытый

b. смещения

c. сдвигатель

d. выходной

e. входной

10 Как называется формула, приведенная ниже?

$$p(y \mid x) = \frac{p(x \mid y)p(y)}{p(x)} = \frac{p(x \mid y)p(y)}{\sum_{y' \in Y} p(x \mid y')p(y')}.$$

Выберите один ответ:

a. парадокс Монти Холла

b. условная вероятность

c. совместная вероятность

d. теорема Байеса

e. аксиоматика Колмогорова

11 В машинном обучении функция активации - это...

Выберите один ответ:

a. вычислительная единица, которая получает информацию, производит над ней простые вычисления и передает ее дальше

b. алгоритм вычисления самой функции или производной по одной переменной

c. связь между двумя функциональными единицами сети

d. способ нормализации входных данных

12 Что такое каузальность?

Выберите один ответ:

a. способ получить обратно из совместной вероятности вероятность того или иного исхода одной из случайных величин

b. отношение «причина–следствие»

c. вероятность наступления одного события, если известно, что произошло другое

d. вероятность одновременного наступления двух событий

13 В каких случаях расстояние Кульбака–Лейблера всегда неотрицательно и равно нулю?

Выберите один ответ:

a. когда распределения p и q параллельны друг другу

b. когда распределения p и q совпадают лишь в двух точках

c. когда распределения p и q совпадают почти всюду

d. когда распределения p и q пересекаются

e. когда распределения p и q не совпадают

14 Что отражает корреляция?

Выберите один ответ:

a. нелинейную часть зависимости между случайными величинами

b. вероятность последовательного наступления двух событий

c. вероятность одновременного наступления двух событий

d. вероятность наступления одного события, если известно, что произошло другое

e. линейную часть зависимости между случайными величинами

15 В машинном обучении синапс - это...

Выберите один ответ:

a. алгоритм вычисления самой функции или производной по одной переменной

b. вычислительная единица, которая получает информацию, производит над ней простые вычисления и передает ее дальше

c. способ нормализации входных данных

d. связь между двумя функциональными единицами сети

16 Расстояние Кульбака–Лейблера - это?

Выберите один ответ:

a. мера суммы двух вероятностных распределений P и Q

b. мера произведения двух вероятностных распределений P и Q

c. мера разницы между двумя вероятностными распределениями P и Q

d. мера разницы производных двух вероятностных распределений P и Q

17 Если выполняется равенство ниже, то как называются 2 величины?

$$p(x, y) = p(x)p(y).$$

Выберите один ответ:

a. каузальными

b. зависимыми

c. линейными

d. независимыми

e. условно зависимыми

18 Как формулируется задача классификации?

Выберите один ответ:

a. Требуется построить решающее правило, то есть определить такую функцию f(x), что при f(x)>0 вектор x относится к первому классу, а при f(x)<0 - ко второму

b. Требуется найти ограниченное решение u(x, t) ($t \geq 0$, $0 < x < +\infty$) уравнения, удовлетворяющее некоторому начальному условию

c. Требуется найти набор $a_1,...,a_k$ и разбиение $\{x_p\}=P_1 \leq P_2 \leq ... \leq P_k$, минимизирующие D (сумма расстояний от принадлежащих ему точек выборки до ядра класса)

d. Требуется найти точку условного минимума многочлена второго порядка при условиях вида $x_i = c_i$ для $i = i_1,\ldots,i_k$, тем самым спрогнозировать следующее значение ряда данных

e. Требуется найти значения параметров или функций, реализующих максимум или минимум некоторой зависящей от них величины $z=f(x_1,x_2,\ldots\ldots x_n)$

19 Как формулируется задача прогнозирования значений?

Выберите один ответ:

a. Требуется найти ограниченное решение u(x, t) ($t \geq 0$, $0 < x < +\infty$) уравнения, удовлетворяющее некоторому начальному условию

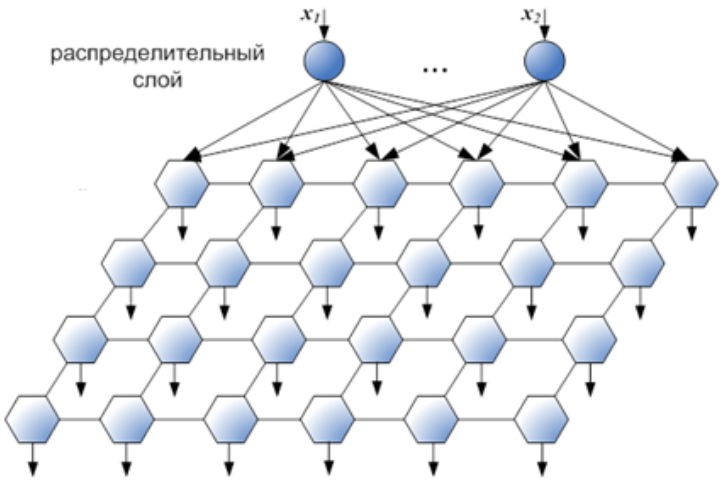
b. Требуется найти значения параметров или функций, реализующих максимум или минимум некоторой зависящей от них величины $z=f(x_1,x_2,\ldots\ldots x_n)$

c. Требуется построить решающее правило, то есть определить такую функцию f(x), что при f(x)>0 вектор x относится к первому классу, а при f(x)<0 - ко второму

d. Требуется найти точку условного минимума многочлена второго порядка при условиях вида $x_i = c_i$ для $i = i_1,\ldots,i_k$, тем самым спрогнозировать следующее значение ряда данных

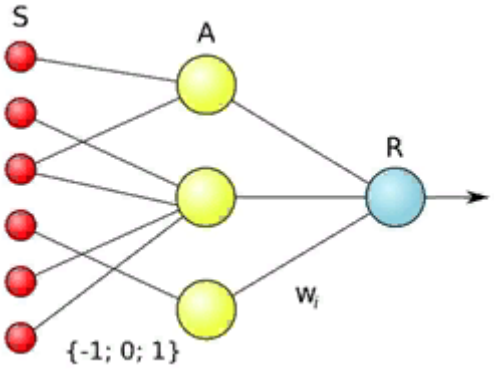
e. Требуется найти набор $a_1,...,a_k$ и разбиение $\{x_p\}=P_1 \leq P_2 \leq ... \leq P_k$, минимизирующие D (сумма расстояний от принадлежащих ему точек выборки до ядра класса)

20 Как называется ИНС, приведенная на изображении ниже?



Выберите один ответ:
a. сверточная нейронная сеть
b. цепь Маркова
c. сеть Кохонена
d. персептрон Розенблатта
e. сеть Хопфилда

21 Введите на месте пропуска текст (регистр не учитывается).
Как называется ИНС, приведенная на изображении ниже?



Ответ: _____

22 Установите соответствие между поставленной задачей и архитектурой ИНС, которая была разработана для решения этой задачи:

Архитектура ИНС	Задача
1) сети Кохонена	А) Классификация и кластеризация (с учителем)
2) персептрон Розенблатта	Б) Восстановление пробелов в данных
3) сети Хопфилда	В) Кластер-анализ и классификация (без учителя)

Выберите вариант ответа:
a. 1-А, 2-Б, 3-В;
b. 1-В, 2-А, 3-Б;
c. 1-В, 2-Б, 3-А;
d. 1-Б, 2-В, 3-А;

23 Введите на месте пропуска текст (регистр не учитывается).

Для чего служит расстояние Кульбака–Лейблера?

Ответ: _____

24 Введите на месте пропуска текст (регистр не учитывается).

Во время инициализации нейронной сети каким образом расставляются веса?

Ответ: _____

25 Введите на месте пропуска текст (регистр не учитывается).

Какая архитектура ИНС была самой первой?

Ответ: _____

Раздел 2. Сбор и подготовка реальных данных для машинного обучения

Перечень вопросов для тестирования:

1 Что визуализируют диаграммы размаха?

Выберите один ответ:

a. визуализируют связи между двумя и более категориальными переменными

b. техника визуализации распределения численных переменных, которая дает представление о центре, разбросе и асимметрии в распределении значений, а также о наличии выбросов

c. отображает распределение одной переменной, для нее рассчитывается сглаженная оценка распределения вероятности переменной

d. визуализация соотношения между двумя численными переменными, показывает наличие зависимости между ними

2 Что такое информативные отсутствующие данные?

Выберите один ответ:

a. специально добавленные пропуски в данных, положительно влияющие на точность модели

b. пропуски в данных входных признаков, которые не имеют зависимости с целевой переменной

c. пропуски в целевой переменной, которые влияют на входные признаки

d. пропуски в данных входных признаков, которые имеют влияние на выходную переменную

3 Какие диапазоны наиболее часто используются в нормализации?

Выберите несколько ответов:

a. от 0 до 10

b. от -1 до 1

c. от 0 до 1

d. любой, зависит от данных

4 Установите соответствие между задачей ML и сферой применения:

Сфера применения	Задача
1) Предложения продуктов	А) классификация
2) Прогнозы на рынке ценных	Б) регрессия

бумаг
3) Анализ тональности текстов В) рекомендация
4) Предложение контента
5) Фильтрация спама
6) Прогноз спроса

Выберите вариант ответа:

a. 1-Б, 2-А, 3-В, 4-А, 5-Б, 6-В;

b. 1-В, 2-Б, 3-А, 4-В, 5-А, 6-Б;

c. 1-В, 2-А, 3-А, 4-Б, 5-В, 6-Б;

d. 1-А, 2-А, 3-А, 4-В, 5-А, 6-Б;

5 Как поступит ML- система, если необходимо определить наилучший порядок действий по отношению к новой задаче?

Выберите один ответ:

a. ML-система собирает новый набор ключевых признаков

b. ML-система использует данные за прошлые периоды, на которых она обучалась

c. ML-система формирует новые данные к задаче

d. ML-система ничего не делает

6 Как называются алгоритмы, которые ищут наиболее точную модель на основе подмножества из набора признаков, сохраняя сигнал и отбрасывая шум?

Выберите один ответ:

a. алгоритмы отбора признаков

b. алгоритмы конструирования

c. алгоритмы машинного обучения

d. алгоритмы прогноза целевой переменной

7 Какая характеристика является ключевой как в человеческом обучении, так и в машинном?

Выберите один ответ:

a. способность обобщать

b. способность думать

c. способность запоминать

d. способность говорить

8 Введите на месте пропуска текст (регистр не учитывается).

Какой процесс в машинном обучении называют выпасом данных?
Ответ: _____

9 Прочтите код, приведенный в листинге ниже. Что он делает?

```
def  cabin_features (data):
     features  =  []
     for cabin in data:
          cabins  = cabin.split("  ")
          n_cabins  =  len(cabins)
          try:
               cabin_char  = cabins[0][0]
```

```
            except  IndexError:
                cabin_char  = "X"
                n_cabins  =  0
            try:
                cabin_num  =  int(cabins [0][1:])
            except:
                cabin_num  =  -1
            features.append([cabin_char, cabin_num, n_cabins])
        return features
```

Выберите один ответ:

a. удаляет пропуски в данных

b. обрабатывает исключения при чтении данных

c. преобразует признаки в другой тип данных

d. извлекает признаки из существующих данных

10 Введите на месте пропуска текст (регистр не учитывается). В чем заключается задача рекомендации? Ответ: _____

11 Входные признаки совместно с известными значениями целевой переменной составляют...

Выберите один ответ:

a. репрезентативную выборку

b. генеральную совокупность данных

c. обучающую выборку

d. тестовую выборку

12 Перечислите какими способами можно увеличить точность ML-модели.

Выберите несколько ответов:

a. Предварительная обработка данных

b. Работа модели на реальных данных

c. Разработка нового варианта алгоритма

d. Выбор подмножества признаков

e. Редактирование параметров модели

13 На рисунке приведен датасет с данными о пингвинах. Установите соответствие между названием столбца и его типом (входной признак или целевая переменная).

	species	island	culmen_length_mm	culmen_depth_mm	flipper_length_mm	body_mass_g	sex
0	Adelie	Torgersen	39.1	18.7	181.0	3750.0	MALE
1	Adelie	Torgersen	39.5	17.4	186.0	3800.0	FEMALE
2	Adelie	Torgersen	40.3	18.0	195.0	3250.0	FEMALE
3	Adelie	Torgersen	NaN	NaN	NaN	NaN	NaN
4	Adelie	Torgersen	36.7	19.3	193.0	3450.0	FEMALE

Набор данных состоит из 7 столбцов.

- species: виды пингвинов (Chinstrap, Adélie, или Gentoo)
- culmen_length_mm: длина кульмена (мм)

- culmen_depth_mm: глубина отверстия (мм)
- flipper_length_mm: длина плавника (мм)
- body_mass_g: масса тела (г)
- island: название острова (Дрим, Торгерсен или Бискоу) в архипелаге Палмер (Антарктида)
- sex: какого пола пингвин в записи

Название столбца	Тип признака
1) culmen_depth_mm	А) отклик
2) island	Б) данные
3) body_mass_g	В) факт
4) species	Г) входной признак
5) culmen_length_mm	
6) flipper_length_mm	
7) sex	

Выберите вариант ответа:
a. 1-Б, 2-А, 3-В, 4-Г, 5-Б, 6-Г, 7-В;
b. 1-Г, 2-Г, 3-Г, 4-А, 5-Г, 6-Гб 7-Г;
c. 1-В, 2-А, 3-А, 4-Б, 5-В, 6-Б, 7-А;
d. 1-А, 2-А, 3-А, 4-В, 5-А, 6-Б, 7-Б;

14 Введите на месте пропуска текст (регистр не учитывается).

Какие данные в машинном обучении считаются отсутствующими?
Ответ: _____

15 Укажите порядок действий при отборе входных признаков.

Описание	Порядок
1) Включите все признаки, которые кажутся хоть как-то связанными с целевой переменной. Выполните обучение ML-модели. Если точность прогноза вас устраивает, останавливайтесь.	А) первый
2) Запустите для набора признаков алгоритм отбора	Б) второй
3) Расширьте набор признаков, добавив туда признаки, связь которых с целевой переменной слабо очевидна. Выполните обучение модели и оцените ее точность. Если она вас устраивает, останавливайтесь	В) третий

Выберите вариант ответа:
a. 1-Б, 2-А, 3-В;
b. 1-В, 2-Б, 3-А;
c. 1-А, 2-В, 3-Б;
d. 1-А, 2-Б, 3-В;

16 Введите на месте пропуска текст (регистр не учитывается).

В наборе данных есть входной признак, в котором есть отсутствующие значения, не несущие информационной нагрузки. Например, в столбце "Возраст" есть пропуски в данных, можно ли вставить в пустые ячейки значение "-1"?

Ответ: _____

17 Дополните код, вставьте в пропуски […] команды.

```
def  cabin_features (data):
     features  =  []
     for cabin in data:
          cabins  = cabin.split("  ")
          n_cabins  =  len(cabins)
          […]:
               cabin_char  = cabins[0][0]
          […]  IndexError:
               cabin_char  = "X"
               n_cabins  =  0
          […]:
               cabin_num  =  int(cabins [0][1:])
          […]:
               cabin_num  =  -1
          features.append([cabin_char, cabin_num, n_cabins])
      return features
```

Выберите несколько ответов:

a. catch;

b. try;

c. except;

18 Какие есть ограничения на использование сведений в качестве входного признака?

Выберите несколько вариантов ответов:

a. признак не должен иметь отношения к целевой переменной

b. значение признака должно быть известно на момент прогноза

c. признак должен быть численным или категориальным

d. значение признака должно быть уникальным для каждой записи

e. признак должен обязательно быть временного типа

19 Введите на месте пропуска текст (регистр не учитывается).

Как называется метод очистки и обработки данных, при котором производится поиск и обработка имен, которые пишутся по-разному, хотя и относятся к одной и той же сущности, а также отсутствующих или недостоверных значений?

Ответ: _____

20 В статистике и теории вычислительных систем целевая переменная называется...

Выберите несколько вариантов ответов:

a. зависимой переменной

b. оттоком

c. независимой переменной

d. откликом

e. объясняющей переменной

21 Установите соответствие с порядком стадий процесса машинного обучения.

Описание	Порядок
1) оценка модели	А) первый этап
2) оптимизация моделей	Б) второй этап
3) построение модели	В) третий этап
4) подготовка данных	Г) четвертый этап
5) прогноз на новых данных	Д) пятый этап

Выберите вариант ответа:

a. 1-Д, 2-Г, 3-В, 4-Б, 5-А;

b. 1-В, 2-Б, 3-Д, 4-А, 5-Г;

c. 1-В, 2-Г, 3-Б, 4-А, 5-Д;

d. 1-А, 2-Б, 3-В, 4-Г, 5-Д;

22 Что значит "проектирование признаков"?

Выберите один ответ:

a. использование существующих признаков для создания новых, увеличивающих ценность исходных данных

b. каждый признак визуализируется для понимания распределения данных

c. каждый признак обрабатывается с целью его подгонки под единую числовую шкалу

d. в каждом признаке убираются отсутствующие значения или заменяются неким оценочным значением

23 По какой причине возникают отсутствующие данные?

Выберите один ответ:

a. это артефакты процесса сбора, возникающие, когда по какой-то причине конкретное значение для экземпляра данных измерить невозможно

b. это случайные флуктуации, возникающие в наборе данных при загрузке в модель

c. это артефакты процесса сбора, возникающие, когда по какой-то причине значение для экземпляра данных выходит из допустимого диапазона данных для выбранного типа данных

d. это артефакты процесса обучения ML-модели, возникающие, когда по какой-то причине данные в датасете не соответствуют требованиям алгоритма

24 Какие существуют способы получения достоверных значений целевой переменной?

Выберите несколько вариантов ответа:

a. по результатам проведенных мероприятий опрашивать клиентов и проводить с ними практические эксперименты

b. случайным образом сгенерировать данные для исследования

c. использовать шаблонные решения и подходы для определения значения целевой переменной

d. запускать управляемые эксперименты и следить за ответами

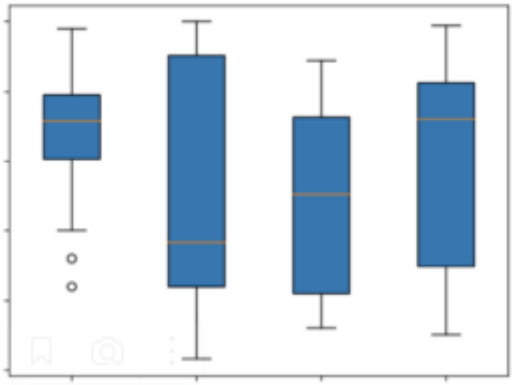
e. нанять аналитика, вручную просматривающего текущие данные и данные за прошлые периоды

f. на краудсорсинговой основе привлечь к оценке значений целевой переменной широкий круг лиц

g. собрать данные, заставляя случайных людей читать и высказывать мнение по поводу прочитанного

h. вычислить значение целевой переменной на основе других данных

25 Какая техника визуализации показана на рисунке?



Выберите один ответ:

a. мозаичные диаграммы

b. графики плотности

c. диаграммы размаха

d. диаграммы рассеяния

Раздел 3. Моделирование и прогнозирование.

Перечень вопросов для тестирования:

1 Введите на месте пропуска текст (регистр не учитывается).

Как называется граница, которая, согласно модели классификатора отделяет экземпляры одного класса от другого?

Ответ: _____

2 Какой алгоритм машинного обучения моделирует скалярную целевую переменную с одним или несколькими количественными признаками, при этом вычисляет линейную комбинацию?

Выберите один ответ:

a. Деревья решений

b. Бустинг

c. Линейная регрессия

d. Логистическая регрессия

e. Метод опорных векторов

3 Какой алгоритм машинного обучения выполняет рекурсивное разбиение обучающей выборки на подмножества по результатам

тестирования значения атрибута, что порождает предсказывающие целевую переменную ветви решений?

Выберите один ответ:

a. Метод опорных векторов

b. Линейная регрессия

c. Деревья решений

d. Логистическая регрессия

e. Бустинг

4 Дополните текст, вставьте в пропуски […] слова.

Функция f обычно называется […], в то время как случайную переменную ε называют […] .

Выберите варианты ответов:

a. целью

b. признаком

c. шумом

d. сигналом

5 Для чего в машинном обучении используется параметр регуляризации?

Выберите один ответ:

a. для снижения риска переобучения модели

b. для восполнения отсутствующих значений

c. для снижения количества выбросов в данных

d. для ускорения работы модели

6 Существует несколько видов целевых признаков. Вы работаете над скоринговой системой, которая в качестве результата будет выдавать ответ "да" или "нет" на вопрос "Выдать текущему клиенту кредит?". С каким видом целевого признака вы работаете?

Выберите один ответ:

a. Количественная дискретная

b. Качественная бинарная

c. Качественная полиномиальная

d. Количественная непрерывная

7 Какой алгоритм машинного обучения импортируется в коде ниже?

from sklearn.svm import SVC as Model

Выберите один ответ:

a. дерево решений

b. логистическая регрессия

c. метод k-ближайших соседей

d. метод опорных векторов с нелинейной функцией ядра

8 Как называется алгоритм обучения нейронной сети, в котором у модели есть набор данных, и нет явных указаний, что с ним делать? Нейронная сеть пытается самостоятельно найти корреляции в данных, извлекая полезные признаки и анализируя их.

Выберите один ответ:

a. обучение с учителем
b. обучение с частичным привлечением учителя
c. обучение с подкреплением
d. обучение без учителя

9 Укажите требует ли алгоритм машинного обучения нормализации или нет.

Алгоритм	Требует нормализации?
1) Логистическая регрессия	А) да
2) Дерево решений	Б) нет
3) Наивный байесовский алгоритм	
4) Метод k-ближайших соседей	
5) Метод опорных векторов с ядром	
6) Бустинг	

Выберите вариант ответа:
a. 1-А, 2-А, 3-А, 4-Б, 5-Б, 6-Б;
b. 1-Б, 2-А, 3-Б, 4-А, 5-Б, 6-А;
c. 1-Б, 2-Б, 3-Б, 4-А, 5-А, 6-А;
d. 1-А, 2-Б, 3-А, 4-А, 5-А, 6-Б;

10 Существует несколько видов целевых признаков. Вы работаете над банковской советующей системой, которая в должна предсказать сколько раз конкретный клиент обратится за кредитом в течение 5 лет. При этом банк не готов выдать клиенту более 10 кредитов за этот период. С каким видом целевого признака вы работаете?

Выберите один ответ:
a. Количественная дискретная
b. Качественная бинарная
c. Качественная полиномиальная
d. Количественная непрерывная

11 К чему в машинном обучении сводится задача регрессии?

Выберите один ответ:
a. к неконтролируемому формированию групп отдельных экземпляров данных со сходными характеристиками
b. к предсказанию численных значений
c. к распределению новых данных по классам (двум и более)
d. к обнаружению связей между переменными в больших объемах данных

12 Существует несколько видов целевых признаков. Вы работаете над системой распознавания цифр, целью является стабильно распознавать все цифры от 0 до 9. С каким видом целевого признака вы работаете?

Выберите один ответ:
a. Количественная непрерывная
b. Качественная полиномиальная
c. Качественная бинарная

d. Количественная дискретная

13 Что такое "статистическое моделирование"?

Выберите один ответ:

a. поиск компромисса между точностью прогнозов и интерпретируемостью модели

b. формирование набора данных для дальнейшего обучения модели

c. проведение различного рода экспериментов, которые опишут зависимости в данных

d. поиск наиболее сложных взаимосвязей в данных для их последующей обработки

14 Введите на месте пропуска текст (регистр не учитывается).

Вам поручено для датасета с параметрами автомобилей предсказать расход количества топлива на километр. В построенной модели функция f описывает истинный расход топлива каждого автомобиля как функцию от множества его входных признаков. Но в данные могут закрасться источники шума. Напишите, что может являться помехой при измерениях?

Ответ: _____

15 Что такое "статистическое моделирование"?

Выберите один ответ:

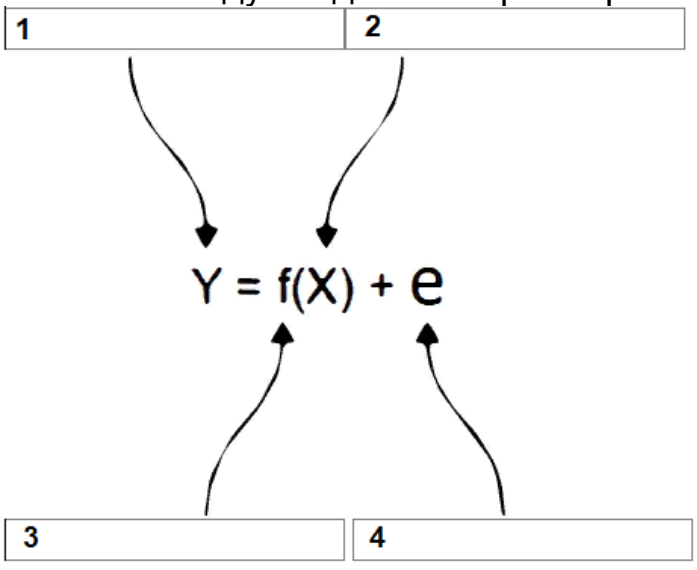
a. поиск компромисса между точностью прогнозов и интерпретируемостью модели

b. формирование набора данных для дальнейшего обучения модели

c. проведение различного рода экспериментов, которые опишут зависимости в данных

d. поиск наиболее сложных взаимосвязей в данных для их последующей обработки

16 Подпишите блоки на изображении, которое иллюстрирует соотношение между входными параметрами и результатом.



Варианты ответов:

a. цель

b. модели или сигнал

c. ошибка или шум

d. информативные признаки X(1,…,n)

17 Относительно метода опорных векторов что называется ядром?

Выберите один ответ:

a. определенный принцип построения программного обеспечения и реализации повторного использования

b. математическая конструкция, которая может «включать» в себя пространство с данными

c. программа, лежащая в основе операционной системы компьютера и, как правило, полностью контролирующая все в системе

d. организованная совокупность данных, предназначенная для длительного хранения

18 У машинного обучения две основные задачи – прогнозы и выводы. Предположим, у нас есть ML-модель fest, описывающая связь между расходом топлива и входными характеристиками автомобиля. Мы задаем ей вопрос «Какие из входных признаков сильнее всего связаны с целевой переменной?». К какой задаче относится данный вопрос?

Выберите один ответ:

a. задача вывода новых зависимостей в данных

b. задача прогнозирования

c. этот вопрос не относится к задачам машинного обучения

19 Установите соответствие между типом задачи машинного обучения и алгоритмом.

Алгоритм	Требует нормализации?
1) Логистическая регрессия	А) классификация
2) Дерево решений	Б) классификация и регрессия
3) Наивный байесовский алгоритм	В) регрессия
4) Метод k-ближайших соседей	
5) Метод опорных векторов с ядром	
6) Линейная регрессия	

Выберите вариант ответа:

a. 1-А, 2-А, 3-А, 4-Б, 5-Б, 6-В;

b. 1-Б, 2-А, 3-Б, 4-А, 5-Б, 6-А;

c. 1-Б, 2-В, 3-Б, 4-В, 5-А, 6-А;

d. 1-А, 2-Б, 3-В, 4-А, 5-В, 6-Б;

20 Как называется тип моделирования, в котором производится поиск компромисса между точностью прогнозов и интерпретируемостью модели?

Выберите один ответ:

a. статистическое моделирование

b. знаковое моделирование

c. натурное моделирование

d. идеальное моделирование

21 Введите на месте пропуска текст (регистр не учитывается).

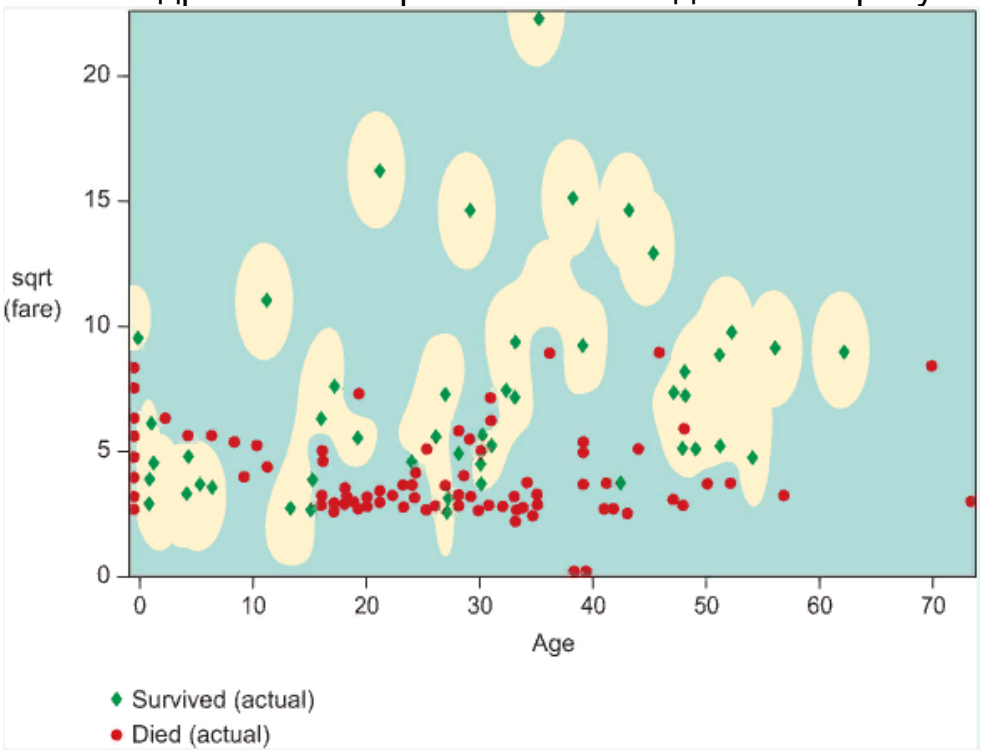
Перед вами стандартная модель линейной регрессии:

$$f(X) = \beta_0 + X_1 \times \beta_1 + X_2 \times \beta_2 + \cdots$$

Как называются ее элементы?

Ответ: _____

22 Ниже на рисунке приведена визуализация работы модели по предсказанию спасения на Титанике методом опорных векторов с нелинейным ядром. Какая проблема наблюдается на рисунке?



Выберите один ответ:

a. модель переобучена

b. модель недообучена

c. модель не сработала

d. проблем нет

23 Введите на месте пропуска текст (регистр не учитывается).

Какая модель использует неразмеченные данные, из которых алгоритм самостоятельно пытается извлечь признаки и зависимости?

Ответ: _____

24 Какой алгоритм машинного обучения распределяет наблюдения по категориям на базе количественных признаков, предсказывает целевой класс или вероятности целевых классов?

Выберите один ответ:

a. Логистическая регрессия

b. Деревья решений

c. Метод опорных векторов

d. Бустинг

e. Линейная регрессия

25 Существует несколько видов целевых признаков. Вы работаете над системой предсказания изменений цены на шанежку в столовой. С каким видом целевого признака вы работаете?

Выберите один ответ:

a. Качественная бинарная

b. Количественная непрерывная

c. Качественная полиномиальная

d. Количественная дискретная

Раздел 4. Оценка и оптимизация модели

Перечень вопросов для тестирования:

1 По виду ROC-кривой можно сделать следующий вывод...

Выберите один ответ:

a. увеличение доли ложно-положительных значений означает уменьшение доли истинно-положительных

b. увеличение доли истинно-положительных значений означает уменьшение доли ложно-отрицательных

c. увеличение доли истинно-положительных значений означает уменьшение доли истинно-отрицательных

d. вид ROC-кривой не позволяет оценить качество предсказаний

2 Стандартный способ оптимизации параметров настройки в ML-моделях сводится к сеточному поиску. Дайте определение этому методу.

Выберите один ответ:

a. алгоритм оптимизации, который позволяет нам при помощи перебора выбирать лучшие параметры для оптимизации проблемы из списка вариантов, который мы предоставляем, тем самым автоматизируя метод «проб и ошибок»

b. алгоритм машинного обучения, строящий предсказание в виде ансамбля слабых предсказывающих моделей, которыми в основном являются деревья решений

c. критерий выбора из класса параметризованных регрессионных моделей

d. измерение, которое исследует, как различия в одной переменной могут быть объяснены различием во второй переменной при прогнозировании исхода данного события

3 Как называется метрика оценки, которая рассматривает разницу всех предсказанных и соответствующих им известных значений и вычисляет среднее способом, позволяющим пренебречь тем фактом, что каждое из предсказанных значений может быть как больше, так и меньше реального?

Выберите один ответ:

a. area under curve

b. рабочая характеристика приемника

c. коэффициент детерминации

d. среднеквадратичная ошибка (MSE)

e. квадратный корень из среднеквадратичной ошибки (RMSE)

4 Как называется алгоритм оценки, который исследует, как различия в одной переменной могут быть объяснены различием во второй переменной при прогнозировании исхода данного события и является всегда относительным в диапазоне от 0 до 1?

Выберите один ответ:

a. коэффициент детерминации R-квадрат

b. квадратный корень из среднеквадратичной ошибки (RMSE)

c. рабочая характеристика приемника

d. среднеквадратичная ошибка (MSE)

e. area under curve

5 В чем заключается метод контроля по k-блокам?

Выберите один ответ:

a. из обучающей выборки случайным образом изымается некоторая небольшая часть, на этой части данных тестируется модель, остальные данные используются для обучения модели

b. данные случайным образом делятся на некоторое количество непересекающихся подмножеств, затем модель раз за разом обучается на всем наборе, за исключением данных очередного подмножества, которые в свою очередь используются для генерации последующих предсказаний

c. обучающая выборка рекурсивно разбивается на подмножества по результатам тестирования значения атрибута, что порождает предсказывающие целевую переменную

d. данные применяются для оценки неизвестных функций на основе большого количества входных переменных через алгоритм обратного распространения

6 За что в алгоритмах машинного обучения отвечает параметр настройки (tuning parameters)?

Выберите один ответ:

a. отвечают за сложность взаимосвязи между входными признаками и целевой переменной

b. отвечают за упрощение подхода к обучению модели

c. отвечают за показатели оценочных метрик

d. ни за что не отвечают

7 Выберите оценочные метрики, подходящие для моделей классификации.

Выберите несколько вариантов ответов:

a. таблица сопряженности

b. площадь под ROC-кривой

c. рабочая характеристика приемника

d. анализ остатков

e. коэффициент детерминации

f. среднеквадратичная ошибка

8 Дополните код перекрестной проверки с контролем по k-блокам, вставьте в пропуски […] команды.

```
N = features.shape[0]
K = 10
preds_kfold = np.[...](N)
folds = np.random.randint(0, K, size=N)

for idx in np.[…](K):
    features_train = features[folds != idx,:]
    target_train = target[folds != idx]
    features_test = features[folds == idx,:]
    model = train(features_train, target_train)
    preds_kfold[folds == idx] = predict(model, features_test)

accuracy = evaluate_acc(preds_kfold, target)
```

Выберите несколько ответов:

a. ones;
b. zeros;
c. empty;
d. arrange;
e. randint;

9 Укажите все способы скользящего контроля.

Выберите несколько вариантов ответов:

a. кроссвалидация
b. градиентный спуск
c. бустинг
d. метод отложенных данных
e. контроль по k-блокам

10 Дополните код скользящего контроля методом отложенных данных, вставьте в пропуски […] команды.

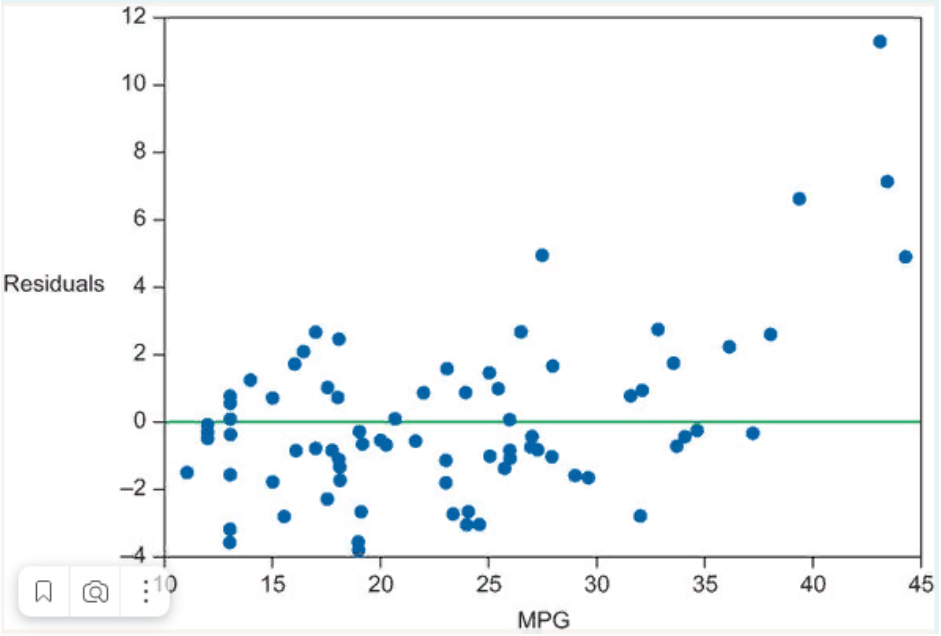
```
features = rand(100,5)
target = rand(100) > 0.5
N = features.shape[0]
N_train = floor(0.7 * N)
idx =random.[…](N)

idx_train = idx[:N_train]
idx_test = idx[N_train:]

features_train = features[idx_train,:]
target_train = target[idx_train]
features_test = features[idx_test,:]
target_test = target[idx_test]

model = train(features_train, target_train)
```

preds_test = predict(model, features_test)
accuracy = evaluate_acc(preds_test, target_test)
Выберите один ответ:
a. permutation
b. randomize
c. arange
11 Какая оценочная метрика показана на рисунке ниже?



Выберите один ответ:
a. байесовский информационный критерий
b. кривая рабочей характеристики приемника
c. таблица сопряженности
d. исследование остатков
12 Введите на месте пропуска текст (регистр не учитывается).
Простейший способ оценки моделей классификации – вычисление доли правильных ответов. Если правильно предсказанными оказались три из четырех строк, то можно сказать, что точность модели на конкретной контрольной выборке составляет... процентов. Впишите только числовой ответ в поле ниже.
Ответ: _____
13 При построении ROC-кривой на какой из осей откладывают значения из таблицы сопряженности?
Выберите один ответ:
a.
Ось Y - доля истинно-положительных значений
Ось X - доля истинно-отрицательных значений
b.
Ось Y - доля ложно-отрицательных значений
Ось X - доля истинно-положительных значений
c.
Ось Y - доля ложно-отрицательных значений

Ось X - доля истинно-отрицательных значений

d.

Ось Y - доля ложно-положительных значений

Ось X - доля истинно-положительных значений

e.

Ось Y - доля истинно-положительных значений

Ось X - доля ложно-положительных значений

14 Какую стратегию необходимо выбирать при построении классификатора и выборе порога для отсечения класса в векторе вероятностей?

Выберите один ответ:

a. максимизирующую долю истинно-отрицательных значений

b. максимизирующую долю ложно-отрицательных значений

c. минимизирующую долю истинно-положительных значений

d. минимизирующую долю ложно-отрицательных значений

15 Какая метрика рассчитывается по формуле, приведенной ниже?

$$\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

Выберите один ответ:

a. информационный критерий Акаике

b. квадратный корень из среднеквадратичной ошибки (RMSE)

c. коэффициент детерминации (R-квадрат)

d. среднеквадратичная ошибка (MSE)

16 Введите на месте пропуска текст (регистр не учитывается).

Напишите, какая метрика вычисляется кодом ниже:

```
def metric(true_values, predicted_values):
    n = len(true_values)
    mean = np.mean(true_values)
    residuals = 0
    total = 0
    for i in range(n):
        residuals += (true_values[i] - predicted_values[i])**2
        total += (true_values[i] - mean)**2
    return 1.0 - residuals/total
```

Ответ: _____

17 На каких данных проводится оценка прогностической точности модели машинного обучения?

Выберите один ответ:

a. новых данных

b. данных из обучающей выборки

c. выбросах

d. отсутствующих данных

18 Установите порядок действий для сеточного поиска по оптимизации параметров настройки в ML-моделях.

Описание этапа	Этап
1) Выберите ML-алгоритм	А) первый
2) Выберите параметр оценки, который будет использоваться как основной	Б) второй
3) Определите сетку как прямое произведение массивов всех параметров настройки	В) третий
4) Выберите параметры, которые собираетесь оптимизировать, и массив значений для тестирования каждого параметра	Г) четвертый
5) Для каждой комбинации настроечных параметров в сетке используйте обучающую выборку для проведения перекрестной проверки и рассчитайте оценочную метрику для полученных предсказаний	Д) пятый
6) Выберите набор параметров настройки, соответствующий максимальному значению оценочного показателя	Е) шестой

Выберите вариант ответа:

a. 1-Б, 2-А, 3-Г, 4-В, 5-Д, 6-Е;

b. 1-Е, 2-Д, 3-Г, 4-В, 5-Б, 6-А;

c. 1-Б, 2-В, 3-А, 4-Д, 5-Г, 6-Е;

d. 1-Г, 2-Е, 3-Д, 4-А, 5-В, 6-Б;

19 Выберите оценочные метрики, подходящие для моделей регрессии.

Выберите несколько вариантов ответов:

a. площадь под ROC-кривой

b. среднеквадратичная ошибка

c. анализ остатков

d. таблица сопряженности

e. рабочая характеристика приемника

f. коэффициент детерминации

20 Введите на месте пропуска текст (регистр не учитывается).

В задачах регрессии численные измерения получают из распределения с некоторой степенью неопределенности. Как она называется?

Ответ: _____

21 Дополните листинг вычисления RMSE, вставьте в пропуски […] команды.

```
def rmse(true_values, predicted_values):
```

```
n = len(true_values)
residuals = 0
for i in range(n):
    residuals […] (true_values[i] - predicted_values[i])[…]2.
return np.sqrt(residuals[…]n)
```

Выберите несколько ответов:

a. **

b. /

c. *

d. +

e. +=

f. -=

22 Введите на месте пропуска текст (регистр не учитывается).

При помощи какого тип инструмента определяется качество предсказаний модели?

Ответ: _____

23 Для чего требуется оценивать точность построенной модели машинного обучения?

Выберите один ответ:

a. для оценки скорости работы модели на новых данных

b. для оценки скорости генерации предсказаний на новых данных

c. для оценки эффективности генерации предсказаний на новых данных

d. для оценки генерации предсказаний на обучающей выборке

24 В чем заключается метод отложенных данных?

Выберите один ответ:

a. данные случайным образом делятся на некоторое количество непересекающихся подмножеств, затем модель раз за разом обучается на всем наборе, за исключением данных очередного подмножества, которые в свою очередь используются для генерации последующих предсказаний

b. обучающая выборка рекурсивно разбивается на подмножества по результатам тестирования значения атрибута, что порождает предсказывающие целевую переменную

c. из обучающей выборки случайным образом изымается некоторая небольшая часть, на этой части данных тестируется модель, остальные данные используются для обучения модели

d. данные применяются для оценки неизвестных функций на основе большого количества входных переменных через алгоритм обратного распространения

25 Введите на месте пропуска текст (регистр не учитывается).

Напишите что является результатом работы вероятностного классификатора?

Ответ: _____

Раздел 5. Основы проектирования признаков

Перечень вопросов для тестирования:

1 Для чего еще, помимо повышения точности, может использоваться отбор признаков в обучении без учителя?

Ответ: _____

2 Проектирование признаков может применяться и для проверки стереотипов, возникающих на базе рабочего опыта. Чтобы узнать, заслуживает ли гипотеза рассмотрения, достаточно использовать ее как признак в ML-модели. Как вы можете понять, что гипотеза верна и заслуживает детального рассмотрения?

Выберите один ответ:

a. После оценки точности работы модели с новым признаком произошло существенное приращение точности работы модели

b. После оценки точности работы модели с новым признаком не было приращения точности работы модели

c. Оценить точность модели с новым признаком не получилось

d. После оценки точности работы модели с новым признаком произошло существенное уменьшение точности работы модели

3 Прямой отбор признаков, это метод, в котором...

Выберите несколько ответов:

a. надо найти разные группы в элементах признаков

b. надо подстроить все значения признаков ML-модели путем перебора

c. надо начать с отсутствия признаков и постепенно найти самые лучшие, которые будут добавлены в подмножество

d. надо начать со всех доступных признаков и последовательными итерациями исключить самые худшие

4 Как изменится средняя арифметическая, если все значения определенного признака увеличить на число A?

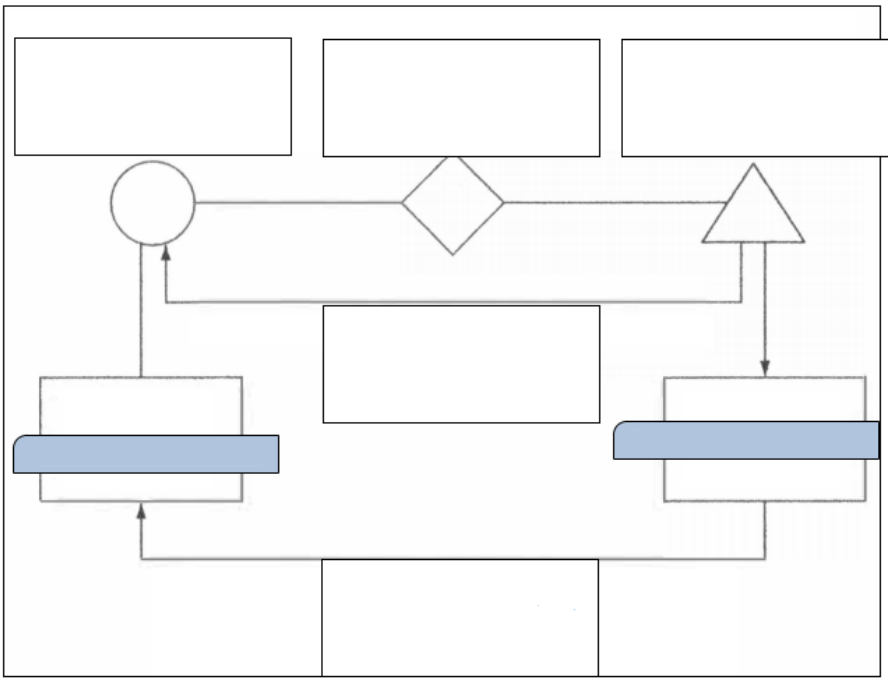
Выберите один ответ:

a. увеличится

b. уменьшится

c. не изменится

5 Подпишите блоки на изображении, которое иллюстрирует схему процесса **прямого** отбора признаков.



Варианты ответов:

a. Фиксируем точность модели после проверки.

b. Добавляем признак из самой лучшей модели к отобранным признакам и удаляем его из доступных. Повторяем.

c. Набор доступных признаков.

d. Выбираем один. Добавляем на время к отобранным признакам.

e. Повторяем для всех доступных признаков.

f. Набор отобранных признаков.

g. Строим и выполняем перекрестную проверку модели на выбранных признаках.

6 Какая концепция связана с выбором признаков, помогающая увидеть взаимосвязи внутри модели и в использовавшихся для ее построения данных?

Выберите один ответ:

a. важности признака

b. отбора признака

c. вторичности признака

d. первичности признака

e. нет такой концепции

7 Как называется метод анализа текстов, в котором подсчитывается количество вхождений каждого слова в текст и вставляется в набор данных столбец с соответствующим числом?

Ответ: _____

8 В алгоритме "мешок слов" есть один серьезный недостаток. Какой?

Выберите один ответ:

a. алгоритм не работает с русским языком

b. набор данных становится меньше и компактнее

c. набор данных становится большим и разреженным

d. алгоритм медленно работает с малыми текстами

9 Установите соответствие между названиями методов и их описаниями

Метод отбора признаков	Описание
1) мешок слов	A Обучающая выборка рекурсивно разбивается на подмножества по результатам тестирования значения атрибута, что порождает предсказывающие целевую переменную в модели
2) случайный лес	B Подсчитывается количество вхождений каждого слова в текст и вставляется в набор данных столбец с соответствующим числом
3) бустинг	C Моделирует скалярную целевую переменную с одним или несколькими количественными признаками. Вычисляет линейную комбинацию, также признаки можно преобразовать с помощью нелинейных функций.
4) линейная регрессия	D Во время работы алгоритма создается ансамбль решающих деревьев, каждое из которых само по себе даёт очень невысокое качество классификации, но за счёт их большого количества результат получается хорошим
5) дерево решений	E Во время работы идет построение ансамбля слабых моделей, в которых модели строятся не независимо, а последовательно, т.е. каждое следующее дерево учится на ошибках предыдущего, затем этот процесс повторяется, наращивая количество слабых моделей

Выберите вариант ответа:

a. 1-A, 2-B, 3-C, 4-D, 5-E;

b. 1-B, 2-D, 3-E, 4-C, 5-A;

c. 1-E, 2-B, 3-D, 4-A, 5-C;

d. 1-C, 2-A, 3-B, 4-E, 5-D;

10 Как при помощи процедур машинного обучения узнать заслуживает ли гипотеза рассмотрения?

Выберите несколько ответов:

a. использовать ее как признак в ML-модели

b. собрать дополнительные данные и проверить ML-модель этими данными

c. оставить в ML-модели только этот признак и оценить точность работы модели

d. использовать ее как признак в ML-модели и проверить точность работы модели с этим признаком и без

11 Перечислите все возможные новые признаки, которые можно получить при проектировании из столбца Дата в приведенном ниже датасете:

	A	B	C	D	E	F	G
1	Ведомость учета заказов						
2	Ф.И.О. заказчика	Наименование товара	Дата заказа	Количество единиц товара в заказе	Количество оплаченных единиц товара в заказе	Цена единицы товара, руб.	Стоимость заказа, руб.
3	Белых А. П.	Пароварка	24.05.14	12	12	3100	37200
4	Михайлов Н. А.	Мультиварка	07.06.14	5	5	4200	21000
5	Седова Н. Р.	Миксер	26.04.14	10	10	600	6000
6	Зотова А. Ф.	Миксер	11.05.14	15	18	600	10800
7	Седова Н. Р.	Пароварка	10.04.14	18	16	3100	49600
8	Зотова А. Ф.	Чайник	28.06.14	8	10	2100	21000
9	Михайлов Н. А.	Блендер	29.06.14	10	12	2300	27600
10	Седова Н. Р.	Кофеварка	03.05.14	15	15	1200	18000
11	Белых А. П.	Чайник	16.05.14	24	24	2100	50400
12	Седова Н. Р.	Блендер	27.05.14	16	20	2300	46000
13	Зотова А. Ф.	Пароварка	06.04.14	10	10	3100	31000
14	Седова Н. Р.	Кофемолка	29.05.14	8	8	900	7200
15	Михайлов Н. А.	Кофеварка	17.05.14	12	10	1200	12000

Выберите несколько ответов:

a. Месяц заказа товара

b. День заказа товара

c. Товар месяца

d. Клиент с неоплаченными товарами

e. Количество дней с последнего заказа клиента

12 Укажите что из перечисленного ниже является проектированием признаков?

Метод отбора признаков	Является проектированием признаков?
1) подсчет числа появлений конкретного слова в тексте	A да
2) разбиение набора данных на k-блоков	B нет
3) разделение писем на спам и не спам	
4) деление общей суммы на количество платежей для получения доли, приходящейся на один платеж	

5) предсказание расхода литра топлива на километры
6) объединение двух таблиц по идентификатору пользователя

Выберите вариант ответа:

a. 1-A, 2-B, 3-A, 4-B, 5-A, 6-B;
b. 1-B, 2-A, 3-A, 4-B, 5-B, 6-A;
c. 1-A, 2-B, 3-B, 4-A, 5-B, 6-A;
d. 1-B, 2-A, 3-B, 4-A, 5-A, 6-A;

13 Установите соответствие между текстовой формулировкой и ML-признаком:

ML-признаком	Текстовое описание
1) числовой признак "Температура воздуха"	A по понедельникам чаще всего покупают набор "Раф и булочка с корицей"
2) категориальный признак "Набор блюд", пример данных: "Раф и булочка с корицей"	B В жаркий день больше покупателей в кофейне
3) булев признак "Сегодня понедельник?"	
4) числовой признак "Средний чек"	
5) числовой признак "Число покупателей"	

Выберите вариант ответа:

a. 1-B, 3-A;
b. 2-A, 5-B;
c. 1-A, 2-B;
d. 4-B, 5-A;

14 Как называется метод отбора признаков, в котором надо начать со всех доступных признаков и последовательными итерациями исключить самые худшие?

Ответ: _____

15 В алгоритме "мешок слов" распространенные, но не несущие информации слова, такими как предлоги, частицы, артикли, называются

Ответ: _____

16 Представьте, что вы руководите службой подписки и хотите при первой авторизации каждого нового клиента спрогнозировать время его пользования вашими услугами. В такой ситуации доступен всего один параметр – географическое положение клиента.

Какое действие при проектировании признаков вам необходимо добавить, чтобы расширить датасет?

Выберите один ответ:

a. подключение сторонних демографических данных
b. применение гистограммы градиентов
c. изменение архитектуры нейронной сети

d. расчет статистических сводок

e. объединение двух таблиц по идентификатору пользователя

17 При обратной связи с увеличением факторного признака результативный признак...

Выберите один ответ:

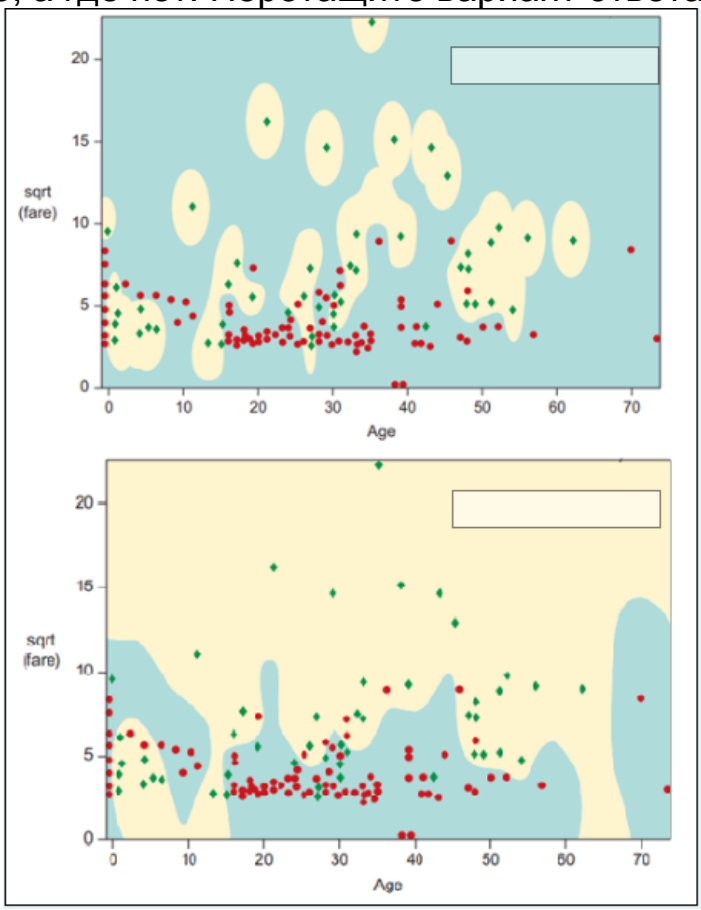
a. увеличивается

b. остается без изменения

c. уменьшается

d. колеблется

18 Перед вами два графика решающих границ двух моделей, обученных на одних и тех же данных. Укажите где модель обучена верно, а где нет. Перетащите вариант ответа на изображение.



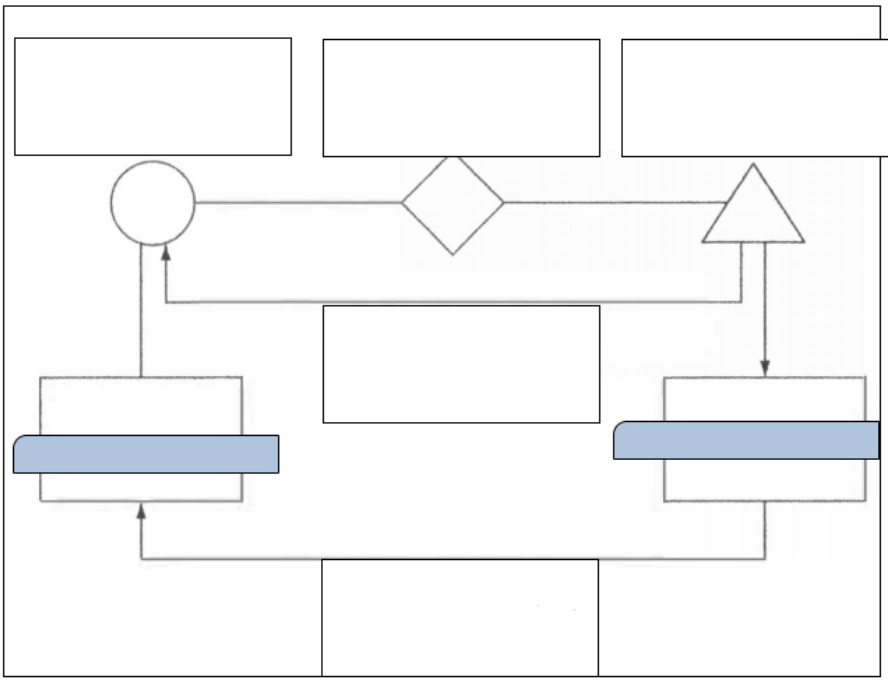
Варианты ответа:

a. модель недообучена

b. модель обучена верно

c. модель переобучена

19 Подпишите блоки на изображении, которое иллюстрирует схему процесса **обратного** отбора признаков.



Варианты ответов:

a. Удаляем признак самой худшей модели из окончательного набора и добавляем в группу удаленных. Повторяем.

b. Повторяем для всех признаков модели.

c. Набор признаков модели.

d. Строим и выполняем скользящий контроль ля признаков модели.

e. Фиксируем точность модели после проверки методом скользящего контроля.

f. Набор удаленных признаков.

g. Выбираем один. На время убираем его из признаков модели.

20 Укажите что является преимуществами методов проектирования признаков из перечисленного ниже:

Выберите несколько ответов:

a. изменение целевой переменной

b. добавление внешних источников данных

c. упрощение интерпретируемых признаков

d. непрерывное совершенствование используемой ML-модели

e. использование источников неструктурированных данных

f. легкое масштабирование ML-модели

g. повышение потенциала информации

h. ускорение оценки качества модели

21 В каком случае останавливается процедура отбора признаков?

Выберите несколько ответов:

a. после того, как добавление или исключение новых признаков приведет к повышению точности до максимума

b. при переполнении стека

c. при достижении заранее указанного размера подмножества

d. после того, как добавление или исключение новых признаков перестает влиять на уровень точности

e. после того, как добавление или исключение новых признаков приведет к снижению точности до минимума

22 Как называется процесс применения математических преобразований к необработанным входным данным с целью получения новых признаков для ML-модели?

Ответ: _____

23 Применение математических преобразований к необработанным входным данным с целью получения новых признаков для ML-модели - это?

Выберите один ответ:

a. проектирование признаков

b. подготовка обучающей выборки

c. сеточный поиск

d. оптимизация параметров модели

24 Признаки для ML-алгоритма "мешок слов" должны быть...

Выберите один ответ:

a. гомогенными

b. гетерогенными

c. гетеродинный

d. шумовыми

25 Простейший способ выбрать оптимальное подмножество признаков - это?

Выберите один ответ:

a. обратное исключение

b. перебор всех комбинаций

c. прямой отбор

d. сеточный поиск

2 Выполнение практических работ

Задания для практических работ, теоретический и методический материал для самостоятельного изучения и вопросы для собеседования представлены в методических указаниях к практическим занятиям.

3 Выполнение лабораторных работ

Задания для лабораторных работ, теоретический и методический материал для самостоятельного изучения и вопросы для собеседования представлены в методических указаниях к лабораторным работам.

4 Вопросы к экзамену для промежуточной аттестации

Вопросы к экзамену в 6 семестре:

1 Основные термины и понятия машинного обучения.

2 Подходы к обучению моделей. Базовый рабочий процесс для ML-модели.

3 Проблематика сбора «больших» данных. Пути решения.

4 Определение набора входных признаков и наблюдаемое значение целевой переменной.

5 Репрезентативность обучающей выборки.
6 Работа с категориальными признаками.
7 Обнаружение и устранение отсутствующих данных.
8 Подходы к нормализации данных.
9 Типовые диаграммы визуализации данных.
10 Построение мозаичных диаграмм и их роль в визуализации.
11 Построение диаграмм размаха и их роль в визуализации.
12 Построение графиков плотности и их роль в визуализации.
13 Построение диаграмм рассеяния и их роль в визуализации.
14 Поиск связи между входными данными и целевой переменной.
15 Методология машинного обучения с учителем.
16 Построение классификатора и получение предсказаний: бинарная классификация.
17 Построение классификатора и получение предсказаний: многоклассовая классификация.
18 Построение регрессора и генерация прогнозов.
19 Проблема переобучения и чрезмерно оптимистической оценки модели.

Вопросы к экзамену в 7 семестре:

20. Методология скользящего контроля.
21 Точность для отдельных классов и таблица сопряженности.
22 Компромиссы при оценке точности и ROC-кривые.
23 Оценка точности многоклассовой классификации.
24 Простые показатели эффективности регрессионных моделей.
25 Исследование остатков.
26 Параметры настройки ML-алгоритмов.
27 Сеточный поиск параметров.
28 Этапы проектирования признаков из входных атрибутов датасета.
29 Извлечение признаков из обычного текста.
30 Прямой отбор признаков из модели.
31 Обратное исключение признаков из модели.
32 Механизм GroupBy.
33 Агрегирование данных.
34 Сводные таблицы и кросс-табуляция.
35 Проверка однородности и полноты кластеров.
36 Метрики качества кластеризации.
37 Метрики эталонных кластеров
38 Типы нейронных сетей.
39 Архитектура многослойного персептрона.
40 Архитектура сети Кохонена.
41 Архитектура сети Маркова.
42 Корреляция и ковариация.
43 Уникальные значения, счетчики значений и членство.
44 Фильтрация отсутствующих данных

45 Нанесение данных на карту в Jupyter Notebook.
46 Задача кластеризации.
47 Обнаружение аномалий в данных.
48 Решение задачи ассоциации.

5 Задачи к экзамену для промежуточной аттестации

Задачи к экзамену в 6 семестре:

1 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета пассажиров Титаника (titanic.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

2 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета Ирисов Фишера (iris.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

3 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета стоимости недвижимости в Бостоне (housing.xlsx) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

4 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета о пациентах медцентра (healthcare-dataset-stroke-data.xlsx) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

5 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета со списком новостей Fake news (fake_news.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

6 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета о пациентах медцентра (parkinsons.data) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

7 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о красном португальском вине «Винью Верде» (winequality-red.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

8 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о белом португальском вине «Винью Верде» (winequality-white.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

9 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о росте и весе 18-летних людей (SOCR-HeightWeight.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

10 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о различных электронных письмах (spambase.data) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

11 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о клиентах банка (application_record.csv и credit_record.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

12 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с о ограненных бриллиантах (diamonds.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

13 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о клиентах медицинской страховой компании (expenses.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

14 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о клиентах банка (100000 CC Records.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

15 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о клиентах телекоммуникационной компании (telecom_churn.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

16 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о сотрудниках компании и их психологическом здоровье (survey.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

17 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о подержанных автомобилях (car details from car dekho.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

18 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о погоде в Лондоне (london_weather.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

19 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о изменении стоимости биткоина в 2012-2021 годах (BTC 2012-2021.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

20 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета Пингвинов Палмера (penguins.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

21 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о параметрах людей (gender_classification_v7.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

22 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о совершенных преступлениях (Crime1.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

23 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с рейтингами фильмов IMDb (film.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

24 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета со вкусами пиццы (pizza.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

25 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с сортами кофе (coffee.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Задачи к экзамену в 7 семестре:

1 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета пассажиров Титаника (titanic.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу бинарной классификации: построить модель, определяющую выживет ли при крушении пассажир.

2 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета Ирисов Фишера (iris.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую к какому из 3 классов ирисов относится текущий цветок.

3 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета Ирисов Фишера (iris.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую вид цветка (setosa, versicolor, virginica) по его параметрам.

4 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета стоимости недвижимости в Бостоне (housing.xlsx) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую цену на конкретную недвижимость в Бостоне.

5 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета о пациентах медцентра (healthcare-dataset-stroke-data.xlsx) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую произойдет ли у пациента инсульт.

6 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета со списком новостей Fake news (fake_news.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую правда эта новость или ложь.

7 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета о пациентах медцентра (parkinsons.data) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую заболеет ли пациент болезнью Паркинсона.

8 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о красном португальском вине «Винью Верде» (winequality-red.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую к какому классу относится вино (плохое 0-4 балла, среднее, 5-7 баллов, хорошее от 8 баллов).

9 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о красном португальском вине «Винью Верде» (winequality-red.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую балльную оценку экспертом качества вина.

10 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о росте и весе 18-летних людей (SOCR-HeightWeight.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую вес в зависимости от роста.

11 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о различных электронных письмах (spambase.data) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую является ли письмо спам-рассылкой или нет.

12 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о клиентах банка (application_record.csv и credit_record.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую возьмет ли клиент кредит в банке.

13 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с о ограненных бриллиантах (diamonds.csv) провести

переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую стоимость бриллианта, в зависимости от его параметров.

14 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о клиентах медицинской страховой компании (expenses.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую стоимость медицинских услуг для клиента.

15 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о клиентах банка (100000 CC Records.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую размер кредитного лимита для текущего клиента.

16 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о клиентах телекоммуникационной компании (telecom_churn.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую расторгнет ли контракт клиент.

17 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о сотрудниках компании и их психологическом здоровье (survey.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую есть ли психическое расстройство у сотрудника или нет.

18 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о подержанных автомобилях (car details from car dekho.csv) провести переформатирование и визуализацию

данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую стоимость подержанного автомобиля.

19 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о погоде в Лондоне (london_weather.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую среднюю температуру в градусах Цельсия в определенную дату.

20 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о изменении стоимости биткоина в 2012-2021 годах (BTC 2012-2021.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую стоимость биткоина в дальнейшем периоде.

21 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета Пингвинов Палмера (penguins.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую к какому из 3 видов пингвинов, относится текущая особь (Адели, Генту и Чинстрап).

22 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета Пингвинов Палмера (penguins.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую изменение массы тела пингвина исходя из его других параметров.

23 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о параметрах людей (gender_classification_v7.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить

сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую к какому из двух полов, относится человек на базе замеров его лица.

24 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о совершенных преступлениях (Crime1.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу классификации: построить модель, определяющую в какой день недели скорее всего будет совершено преступление.

25 Решите поставленную задачу.

Работа выполняется в отдельном блокноте ПО Jupyter Notebook. Для датасета с информацией о стоимости медицинской страховки (expenses.csv) провести переформатирование и визуализацию данных, выполнить агрегирование данных и построить сводную таблицу. Каждое действие производить в отдельной ячейке Jupyter Notebook.

Решить задачу регрессии: построить регрессионную модель, предсказывающую стоимость медицинской страховки на год для клиента.