

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

**УТВЕРЖДАЮ**

Заведующий кафедрой



Матвеев Михаил Григорьевич

Кафедра информационных технологий управления

03.05.2023

**РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ**

Б1.В.02 Алгоритмы и методы машинного обучения

**1. Код и наименование направления подготовки/специальности:**

09.04.02 Информационные системы и технологии

**2. Профиль подготовки/специализация:** Информационные технологии в менеджменте

**3. Квалификация (степень) выпускника:**

Магистратура

**4. Форма обучения:**

Очная

**5. Кафедра, отвечающая за реализацию дисциплины:**

Кафедра информационных технологий управления

**6. Составители программы:**

доцент Копытин Алексей Вячеславович

**7. Рекомендована:** НМС ФКН 03.05.2023 протокол №7

**8. Учебный год:**

2023-2024 1 семестр

**9. Цели и задачи учебной дисциплины:**

Формирование знаний, умений и компетенций в области машинного обучения. Изучение различных математических моделей данных и алгоритмов анализа данных. Формирование практических навыков реализации алгоритмов машинного обучения на языке программирования Python с использованием пакетов NumPy, Pandas, Matplotlib и Scikit-Learn.

**10. Место учебной дисциплины в структуре ООП:**

дисциплина относится к вариативной части блока Б1. Для успешного освоения дисциплины необходимо предварительное изучение курса математического анализа, алгебры, теории вероятностей и математической статистики.

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

Код и название компетенции	Код и название индикатора компетенции	Знания, умения, навыки
<p>ПК-10 Способен определять варианты структур программного обеспечения информационных систем (программного средства), необходимые информационные потоки и исследовать варианты структур с использованием моделей различного уровня</p>	<p>ПК-10.1 Умеет проводить анализ внешнесистемных требований, возможностей их реализации, определяет концептуальный и функциональный облик системы (программного средства), выявление и анализ известных аналогов</p>	<p>Уметь проводить анализ внешнесистемных требований, возможностей их реализации, определяет концептуальный и функциональный облик системы (программного средства), выявление и анализ известных аналогов.</p>
<p>ПК-10 Способен определять варианты структур программного обеспечения информационных систем (программного средства), необходимые информационные потоки и исследовать варианты структур с использованием моделей различного уровня</p>	<p>ПК-10.2 Умеет проводить формирование вариантов структуры системы (программного средства) и разрабатывает варианты реализации их реализации в рамках предлагаемых алгоритмических и программных решений</p>	<p>Уметь проводить формирование вариантов структуры системы (программного средства) и разрабатывает варианты реализации их реализации в рамках предлагаемых алгоритмических и программных решений.</p>
<p>ПК-14 Способен проектировать архитектуру программного средства</p>	<p>ПК-14.2 Умеет определять способы взаимодействия между программными подсистемами программного средства</p>	<p>Уметь определять способы взаимодействия между программными подсистемами программного средства.</p>
<p>ПК-5 Способен организационно и технологически обеспечивать проектирование и дизайн ИС</p>	<p>ПК-5.1 Умеет обеспечивать соответствие проектирования и дизайна ИС принятым стандартам и технологиям</p>	<p>Уметь обеспечивать соответствие проектирования и дизайна ИС принятым стандартам и технологиям.</p>
<p>ПК-5 Способен организационно и технологически обеспечивать проектирование и дизайн ИС</p>	<p>ПК-5.2 Умеет назначать и распределять ресурсы проектирования и дизайна ИС</p>	<p>Уметь назначать и распределять ресурсы проектирования и дизайна ИС.</p>

ПК-15 Способен разрабатывать и исследовать модели объектов профессиональной деятельности, предлагать и адаптировать методики решения научноисследовательских задач, планировать и проводить исследования	ПК-15.1 Умеет обеспечивать сбор научно-технической (научной) информации, необходимой для постановки и решения задач исследования	Уметь обеспечивать сбор научно-технической (научной) информации, необходимой для постановки и решения задач исследования.
--	--	---

## 12. Объем дисциплины в зачетных единицах/час:

3/108

## Форма промежуточной аттестации:

Зачет с оценкой, Контрольная работа

## 13. Трудоемкость по видам учебной работы

Вид учебной работы	Семестр 1	Всего
Аудиторные занятия	54	54
Лекционные занятия	18	18
Практические занятия		0
Лабораторные занятия	36	36
Самостоятельная работа	54	54
Курсовая работа		0
Промежуточная аттестация	0	0
Часы на контроль		0
Всего	108	108

### 13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
<b>1. Лекционные занятия</b>			
1.1	Знакомство с машинным обучением	Знакомство с машинным обучением. Обучение на размеченных данных. Постановка задачи. Обучение без учителя. Признаки в машинном обучении	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
1.2	Линейные модели	Линейные модели в задачах регрессии. Обучение модели линейной регрессии. Градиентный спуск для линейной регрессии. Стохастический градиентный спуск. Линейная классификация. Функции потерь в задачах классификации.	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>

1.3	Проблема переобучения и борьба с ней	Проблема переобучения. Регуляризация. Оценка качества алгоритмов. Кроссвалидация. Выбор гиперпараметров и сравнение алгоритмов	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
1.4	Метрики качества	Метрики качества в задачах регрессии. Метрика качества классификации. Точность и полнота. Объединение точности и полноты. Качество оценок принадлежности классу	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
1.5	Линейные модели: статистический взгляд	Задача регрессии. Метод максимизации правдоподобия. Регрессия как максимизация правдоподобия. Регрессия как оценка среднего. Регуляризация. Логистическая регрессия	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
1.6	Практические рекомендации по линейным моделям	Масштабирование признаков. Нелинейные зависимости. Работа с категориальными признаками. Несбалансированные данные. Многоклассовая классификация	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
1.7	Решающие деревья	Решающие деревья. Обучение решающих деревьев. Критерии информативности. Критерий останова и стрижка деревьев. Решающие деревья и категориальные признаки	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
1.8	Случайные леса	Композиции деревьев. Смещение и разброс. Случайные леса. Трюки со случайными лесами	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
1.9	Градиентный бустинг	Композиции простых алгоритмов. Градиентный бустинг. Описание алгоритма градиентного бустинга. Проблема переобучения градиентного бустинга. Градиентный бустинг для регрессии и классификации. Градиентный бустинг для решающих деревьев	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
<b>2.</b>	<b>Практические занятия</b>		
нет			
<b>3.</b>	<b>Лабораторные занятия</b>		

3.1	Знакомство с машинным обучением	Знакомство с машинным обучением. Обучение на размеченных данных. Постановка задачи. Обучение без учителя. Признаки в машинном обучении	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
3.2	Линейные модели	Линейные модели в задачах регрессии. Обучение модели линейной регрессии. Градиентный спуск для линейной регрессии. Стохастический градиентный спуск. Линейная классификация. Функции потерь в задачах классификации.	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
3.3	Проблема переобучения и борьба с ней	Проблема переобучения. Регуляризация. Оценка качества алгоритмов. Кроссвалидация. Выбор гиперпараметров и сравнение алгоритмов	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
3.4	Метрики качества	Метрики качества в задачах регрессии. Метрика качества классификации. Точность и полнота. Объединение точности и полноты. Качество оценок принадлежности классу	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
3.5	Линейные модели: статистический взгляд	Задача регрессии. Метод максимизации правдоподобия. Регрессия как максимизация правдоподобия. Регрессия как оценка среднего. Регуляризация. Логистическая регрессия	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
3.6	Практические рекомендации по линейным моделям	Масштабирование признаков. Нелинейные зависимости. Работа с категориальными признаками. Несбалансированные данные. Многоклассовая классификация	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
3.7	Решающие деревья	Решающие деревья. Обучение решающих деревьев. Критерии информативности. Критерий останова и стрижка деревьев. Решающие деревья и категориальные признаки	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
3.8	Случайные леса	Композиции деревьев. Смещение и разброс. Случайные леса. Трюки со случайными лесами	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>
3.9	Градиентный бустинг	Композиции простых алгоритмов. Градиентный бустинг. Описание алгоритма градиентного бустинга. Проблема переобучения градиентного бустинга. Градиентный бустинг для регрессии и классификации. Градиентный бустинг для решающих деревьев	<a href="https://edu.vsu.ru/course/view.php?id=11798">https://edu.vsu.ru/course/view.php?id=11798</a>

**13.2. Темы (разделы) дисциплины и виды занятий**

№ п/п	Наименование темы (раздела)	Лекционные занятия	Практические занятия	Лабораторные занятия	Самостоятельная работа	Всего
1	Знакомство с машинным обучением	2		4	6	12
2	Линейные модели	2		4	6	12
3	Проблема переобучения и борьба с ней	2		4	6	12
4	Метрики качества	2		4	6	12
5	Линейные модели: статистический взгляд	2		4	6	12
№ п/п	Наименование темы (раздела)	Лекционные занятия	Практические занятия	Лабораторные занятия	Самостоятельная работа	Всего
6	Практические рекомендации по линейным моделям	2		4	6	12
7	Решающие деревья	2		4	6	12
8	Случайные леса	2		4	6	12
9	Градиентный бустинг	2		4	6	12

		18	0	36	54	108
--	--	----	---	----	----	-----

#### 14. Методические указания для обучающихся по освоению дисциплины

При изучении дисциплины рекомендуется использовать следующие средства:

- рекомендуемую основную и дополнительную литературу;
- методические указания и пособия;
- контрольные задания для закрепления теоретического материала;
- электронные версии учебников и методических указаний для выполнения практических работ. Форма организации самостоятельной работы: подготовка к аудиторным занятиям; выполнение домашних заданий; выполнение контрольных работ.

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

№ п/п	Источник
1	Златопольский, Д. М. Основы программирования на языке Python / Д. М. Златопольский .— Москва : ДМК Пресс, 2017 .— ISBN 978-5-97060-552-3 .— <URL:https://e.lanbook.com/book/97359>.
2	Кук, Д. Машинное обучение с использованием библиотеки H2O / Д. Кук .— Москва : ДМК Пресс, 2018 .— ISBN 978-5-97060-508-0 .— <URL:https://e.lanbook.com/book/97353>.

б) дополнительная литература:

№ п/п	Источник
1	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.
№ п/п	Источник
2	Андреас Мюллер, Сара Гвидо Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными - ИД Вильямс, O'Reilly, 2017. - 480 с.

в) информационные электронно-образовательные ресурсы:

№ п/п	Источник
1	www.lib.vsu.ru - ЗНБ ВГУ
2	https://edu.vsu.ru/course/view.php?id=11798

#### 16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
-------	----------

1	<a href="https://github.com/Gewissta/Introduction_to_ML_with_Python_russian_translation">https://github.com/Gewissta/Introduction_to_ML_with_Python_russian_translation</a>
2	Саммерфилд, М. Python на практике / М. Саммерфилд .— Москва : Лань, 2014 .— 338 с. — ISBN 978-5-97060-095-5 .— <URL: <a href="http://e.lanbook.com/book/66480#authors">http://e.lanbook.com/book/66480#authors</a> >

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

Обучение происходит с использованием электронного обучения и дистанционных образовательных технологий (ДОТ) на портале «Электронный университет ВГУ» (платформа Moodle: <https://edu.vsu.ru/course/view.php?id=11798>).

Учебные материалы размещаются в электронной информационно-образовательной среде вуза «Электронный университет ВГУ – Moodle» для обеспечения возможности дистанционного освоения учебного материала и самостоятельной работы слушателей.

**18. Материально-техническое обеспечение дисциплины:**

Курс реализуется на основе материально-технической базы факультета компьютерных наук Воронежского государственного университета.

Аудитории для проведения занятий: 477, 479, 380, 381, 382, 383, 384, 385, 387, 290, 291, 292, 293, 295, 297, 301п, 303п, 305п, 307п, 314п, 316п, 505п;

Материально-техническое оснащений аудиторий

Наименование помещения (номер аудитории)	Имеющееся оборудование
479	Учебная аудитория: компьютер преподавателя i5-8400-2,8ГГц, монитор с ЖК 19", мультимедийный проектор, экран. Специализированная мебель.
380	Учебная аудитория: компьютер преподавателя i3-3240-3,4ГГц, монитор с ЖК 22", мультимедийный проектор, экран. Специализированная мебель. Система Интернет-видеоконференцсвязи (корп. 1а ауд. 380) Состав системы Интернет-видеоконференцсвязи: ВКС LifeSize Team220 Camera 200 Dual, аудиосистема Defender Mercury 34 SPK-705, интерактивная доска со встроенным проектором "SmartBoard 480iv V25" Лабораторное оборудование по теоретической механике и оптике: машина Атвуда, маятник Максвелла, универсальный маятник, маятник Обербека, крутильный маятник, наклонный маятник, прибор для исследования столкновения шаров, определение скорости полета пули с помощью крутильно-баллистического маятника, изучение законов вращательного движения тел, исследование сложных колебаний, установка для измерения модуля упругости проволоки.
505п	Учебная аудитория: компьютер преподавателя i5-3220-3.3ГГц, монитор с ЖК 17", мультимедийный проектор, экран. Специализированная мебель.
477	Учебная аудитория: ноутбук HP Pavilion Dv9000-er, мультимедийный проектор, экран. Специализированная мебель.
292	Учебная аудитория: компьютер преподавателя Pentium-G3420-3,2ГГц, монитор с ЖК 17", мультимедийный проектор, экран. Специализированная мебель. Система для видеоконференций Logitech ConferenceCam Group и ноутбук 15.6" FHD Lenovo V155-15API.
297	Учебная аудитория: ноутбуки HP EliteBook на базе Intel Core i5-8250U-3.4 ГГц, мониторы ЖК 24" (16 шт.), мультимедийный проектор, экран. Специализированная мебель.



290	<p>Учебная аудитория: персональные компьютеры на базе i7-7800x-4ГГц, мониторы ЖК 27" (12 шт.), мультимедийный проектор, экран. Специализированная мебель.</p> <p>Лабораторное оборудование искусственного интеллекта: рабочие места - персональные компьютеры на базе i7-7800x-4ГГц, мониторы ЖК 27" (12 шт.); модули АО НПЦ "ЭЛВИС" : процессорный Салют-ЭЛ24ПМ2 (9 шт.), отладочный Салют-ЭЛ24ОМ1 (9 шт.), эмулятор MC-USB-JTAG (9 шт.).</p> <p>Лабораторное оборудование электроники, электротехники и схемотехники: рабочие места - персональные компьютеры на базе i7-7800x-4ГГц, мониторы ЖК 27" (12 шт.); стенд для практических занятий по электрическим цепям (KL-100); стенд для изучения аналоговых электрических схем (KL-200); стенд для изучения цифровых схем (KL-300).</p>
291	<p>Учебная аудитория: персональные компьютеры на базе i3-3220-3,3ГГц, мониторы ЖК 19" (16 шт.), мультимедийный проектор, экран. Специализированная мебель.</p>
293	<p>Учебная аудитория: персональные компьютеры на базе Core i7-11700K-3.6 ГГц, мониторы ЖК 24" (15 шт.), мультимедийный проектор, экран. Специализированная мебель.</p> <p>Лабораторное оборудование компьютерной графики видеоадаптеры GeForce RTX 3070.</p>
295	<p>Учебная аудитория: персональные компьютеры на базе i3-9100-3,6ГГц, мониторы ЖК 24" (14 шт.), мультимедийный проектор, экран. Специализированная мебель.</p> <p>Лабораторное оборудование информационной безопасности операционных систем и программных средств защиты информации от несанкционированного доступа: рабочие места - персональные компьютеры на базе Intel i3-9100-3,6ГГц, , мониторы ЖК 24" (14 шт.); учебный стенд «Программные средства защиты информации от несанкционированного доступа».</p>
305п	<p>Учебная аудитория: ноутбук HP Pavilion Dv9000-er, мультимедийный проектор, экран. Специализированная мебель.</p>
307п	<p>Учебная аудитория: ноутбук HP Pavilion Dv9000-er, мультимедийный проектор, экран. Специализированная мебель.</p>
303п	<p>Учебная аудитория: персональные компьютеры на базе i3-8100-3,9ГГц, мониторы ЖК 24" (13 шт.), мультимедийный проектор, экран. Специализированная мебель.</p> <p>Лабораторное оборудование программно-аппаратных средств обеспечения информационной безопасности: персональные компьютеры на базе Intel i3-8100 3.60ГГц, мониторы ЖК 19" (10 шт.), стойка (коммуникационный шкаф), управляемый коммутатор HP Procurve 2524, аппаратный межсетевой экран D-Link DFL-260E, аппаратный межсетевой экран CISCO ASA-5505. лабораторная виртуальная сеть на базе Linux-KVM/LibVirt, взаимодействующая с сетевыми экранами. USB-считыватели смарт-карт ACR1281U-C1 и ACR38U-NEO, смарт-карты ACOS3 72K+MIFARE, карты памяти SLE4428/SLE5528. Учебно-методический комплекс "Программно-аппаратная защита сетей с защитой от HCD" ОАО "ИнфоТекС".</p> <p>Лабораторное оборудование технической защиты информации, состав ST033P "Пиранья" - многофункциональный поисковый прибор, ST03.DA - дифференциальный низкочастотный усилитель, ST03.TEST - контрольное устройство; комплекс виброакустической защиты "Соната": Соната-ИПЗ, Соната-СА-65М, Соната-СВ-45М; генератор-виброизлучатель ( 5 октав) "ГШ-1000У"; генератор шума для защиты объектов вычислительной техники 1, 2 и 3 категорий от утечки информации; система автоматизированная оценки защищенности технических средств от утечки информации по каналу побочных электромагнитных излучений и наводок &lt;Сигурд&gt;. Программно-аппаратный комплекс для мониторинга радиообстановки в диапазоне 9 кГц - 21 ГГц «Кассандра К21». Комплекс оценки эффективности защиты речевой информации от утечки по акустическому и виброакустическому каналам, 20 – 12500 Гц.</p>
314п	<p>Учебная аудитория: персональные компьютеры на базе i3-7100-3,6ГГц, мониторы ЖК 19" (16 шт.), мультимедийный проектор, экран. Специализированная мебель.</p>
316п	<p>Учебная аудитория: персональные компьютеры на базе i3-9100-3,6ГГц, мониторы ЖК 19" (30 шт.), мультимедийный проектор, экран. Специализированная мебель.</p>
381	<p>Учебная аудитория: компьютер преподавателя i3-540-3ГГц, мультимедийный проектор, экран. Специализированная мебель.</p>

382	Учебная аудитория: персональные компьютеры на базе i5-9600KF-3,7ГГц, мониторы ЖК 24" (16 шт.), ТВ панель-флипчарт. Специализированная мебель.
383	Учебная аудитория: персональные компьютеры на базе i7-9700F-3ГГц, мониторы ЖК 27" (16 шт.), мультимедийный проектор, экран. Специализированная мебель.  Лабораторное оборудование мобильных приложений и игр: рабочие места - персональные компьютеры на базе Intel i7-9700F, видеоадаптеры nVidia GeForce RTX2070, мониторы ЖК 27" (16 шт.); Системы виртуальной реальности HTC Vive Cosmos (2шт.); Беспроводной маршрутизатор TP-Link Archer C7.  Лабораторное оборудование безопасности компьютерных сетей: рабочие места - персональные компьютеры HP-3500-PRO на базе Intel i3-2120, мониторы ЖК 22" (16 шт.), стойка (коммуникационный шкаф), управляемый коммутатор CISCO Catalyst 2950, маршрутизатор CISCO 2811-ISR, аппаратный межсетевой экран CISCO серии ASA-5500. лабораторная виртуальная сеть на базе Linux-KVM/LibVirt, взаимодействующая с перечисленным сетевым оборудованием. Программный анализатор сетевого трафика WireShark. Программный симулятор Packet Tracer, для создания виртуальных стендов, включающих коммутаторы 2 и 3 уровней, маршрутизаторы, сетевые экраны и COB. Учебно-методический комплекс "Безопасность компьютерных сетей" ОАО "ИнфоТекС".
384	Учебная аудитория: персональные компьютеры на базе i3-2120-3,3ГГц, мониторы ЖК 22" (16 шт.), мультимедийный проектор, экран. Специализированная мебель.
385	Учебная аудитория: персональные компьютеры на базе i3-2120-3,3ГГц, мониторы ЖК 19" (16 шт.), мультимедийный проектор, экран. Специализированная мебель.
387	Учебная аудитория: компьютер преподавателя Core2Duo-E7600-3ГГц, монитор с ЖК 22", мультимедийный проектор, экран. Персональные компьютеры студентов на базе i5-10400-2,9ГГц, мониторы ЖК 27" (11 шт.). Специализированная мебель.
301п	Учебная аудитория: персональные компьютеры на базе i3-2120-3,3ГГц, мониторы ЖК 17" (15 шт.), мультимедийный проектор, экран. Специализированная мебель.  Лабораторное оборудование суперкомпьютерного центра: кластер с пиковой производительностью 40 Tflops. Состав кластера: 10 узлов, каждый имеет два 12-ядерных процессора Intel Xeon E5-2680V3, 128 Гбайт ОЗУ, SSD 256 Гбайт. 7 узлов из 10 содержат по 2 ускорителя Intel Xeon Phi 7120, 3 узла - 2 ускорителя Tesla K80M. Все узлы объединены высокоскоростной сетью InfiniBand 56 Gbps; управляющий узел кластера (также сервером для хранения файлов): два 6-ядерных процессора, 64 Гбайт оперативной памяти и дисковую подсистему объемом 14 Тбайт; сервер для занятий по параллельному программированию: Intel X5650@2.67GHz 12 ядер 24 потоков, ОЗУ 36ГБ, дисковая подсистема объемом 300ГБ.
190a	Лабораторное оборудование медицинской кибернетики: рабочие места - персональные компьютеры на базе Intel i3-2120, мониторы ЖК 19" (3 шт.); электроэнцефалограф Нейрон-спектр-4 (2 шт.); кардиограф Полиспектр-12 (1 шт.); оптические микроскопы Р-1 (2 шт.); 3D-принтер (1 шт.); паяльные станции (2 шт.). Специализированная мебель.
403п	Учебная аудитория: персональные компьютеры на базе i3-2320-3,3ГГц, мониторы ЖК 22" (7 шт.), мультимедийный проектор, экран. Специализированная мебель.  Лабораторное оборудование физической лаборатории с комплектом оборудования по квантовой физике: Установка для изучения космических лучей (ФПК-01); установка для определения резонансного потенциала методом Франка и Герца (ФПК-02); установка для определения длины свободного пробега частиц в воздухе (ФПК-03); установка для изучения энергетического спектра электронов (ФПК-05); установка для изучения р-п перехода (ФПК-06); установка для изучения температурной зависимости электропроводности металлов и полупроводников (ФПК-07); установка для изучения эффекта Холла в полупроводниках (ФПК-08); установка для изучения спектра атома водорода (ФПК-09); установка для изучения внешнего фотоэффекта (ФПК-10); установка для изучения абсолютно черного тела (ФПК-11); установка для изучения работы сцинтилляционного счетчика (ФПК-12); установка для изучения и анализа свойств материалов с помощью сцинтилляционного счетчика (ФПК-13).
420	Лабораторное оборудование по электротехнике и электронике: лабораторные стенды: полупроводниковые диоды, фотодиод, биполярный транзистор, полевой транзистор, операционный усилитель, многокаскадовый RC-усилитель, амплитудный модулятор и демодулятор, LC-генератор с индуктивной обратной связью, кварцевый генератор, RC-генератор с фазосдвигающей цепью, мультивибратор, триггер на биполярном транзисторе, основные схемы выпрямителей, универсальные логические элементы ТТЛ, регистр сдвига, счетчик Специализированная мебель.
425	Лабораторное оборудование сетей и систем передачи информации: стойка (коммуникационный шкаф), 3 коммутатора CISCO WS-C2960-24TT-L, 3 маршрутизатора CISCO 2801, 2 WiFi-маршрутизатора Linksys WRT54G. Специализированная мебель.

Адреса (местоположения) помещений

Наименование помещения (номер аудитории)	Адрес (местоположение) помещения
479	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 479
380	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 380
505п	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 505
477	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 477
292	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 292
297	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 297
290	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 290
291	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 291
293	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 293
295	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 295
305п	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 305
307п	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 307
303п	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 303
314п	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 314
316п	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 316
381	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 381
382	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 382
383	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 383
384	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 384
385	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 385
387	394018, г. Воронеж, площадь Университетская, д. 1, корп.1а, ауд. 387
308пп	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 308
309п	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 309
301п	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 301
190а	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 190а
403п	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 403
420	394018, г. Воронеж, площадь Университетская, д. 1, корп.1б, ауд. 420
425	394018, г. Воронеж, площадь Университетская, д. 1, корп.1, ауд. 425

Перечень программного обеспечения, используемого в образовательном процессе

Наименование ПО	Производитель ПО (или торговая марка, Или правообладатель) при наличии
ОС Windows v.7, 8, 10	Microsoft (прим. 1)
Платформа электронного обучения LMS-Moodle, основа Образовательного портала «Электронный университет ВГУ»	Moodle Pty Ltd, GNU General Public License
Foxit PDF Reader	корпорация FOXIT SOFTWARE INC., проприетарная бесплатная лицензия

## 19. Оценочные средства для проведения текущей и промежуточной аттестаций

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

№ п/п	Разделы дисциплины (модули)	Код компетенции	Код индикатора	Оценочные средства для текущей аттестации
1	Разделы 1-9	ПК-10	ПК-10.1	Лабораторные работы 1-7
2	Разделы 1-9	ПК-10	ПК-10.2	Лабораторные работы 1-7
3	Разделы 1-9	ПК-14	ПК-14.2	Лабораторные работы 1-7
4	Разделы 1-9	ПК-5	ПК-5.1	Лабораторные работы 1-7

5	Разделы 1-9	ПК-5	ПК-5.2	Лабораторные работы 1-7
6	Разделы 1-9	ПК-15	ПК-15.1	Лабораторные работы 1-7

Промежуточная аттестация

Форма контроля - Зачет с оценкой, Контрольная работа

Оценочные средства для промежуточной аттестации

Тест

## **20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания**

### **20.1 Текущий контроль успеваемости**

1. Загрузите данные из файла *advertising.csv* в объект pandas DataFrame.

Создайте массивы NumPy  $X$  из столбцов TV, Radio и Newspaper и  $y$  - из столбца Sales. Используйте атрибут *values* объекта pandas DataFrame.

Отмасштабируйте столбцы матрицы  $X$ , вычтя из каждого значения среднее по соответствующему столбцу и поделив результат на стандартное отклонение. Для определенности, используйте методы *mean* и *std* векторов NumPy (реализация *std* в Pandas может отличаться). Обратите внимание, что в numpy вызов функции *.mean()* без параметров возвращает среднее по всем элементам массива, а не по столбцам, как в pandas. Чтобы произвести вычисление по столбцам, необходимо указать параметр *axis*.

Добавьте к матрице  $X$  столбец из единиц, используя методы *hstack*, *ones* и *reshape* библиотеки NumPy.

2. Реализуйте функцию *mserror* - среднеквадратичную ошибку прогноза. Она принимает два аргумента - объекты Series  $y$  (значения целевого признака) и  $y_{pred}$  (предсказанные значения). Не используйте в этой функции циклы - тогда она будет вычислительно неэффективной.

Какова среднеквадратичная ошибка прогноза значений Sales, если всегда предсказывать медианное значение Sales по исходной выборке?

3. Реализуйте функцию *normal\_equation*, которая по заданным матрицам (массивам NumPy)  $X$  и  $y$  вычисляет вектор весов  $w$  согласно нормальному уравнению линейной регрессии.

Какие продажи предсказываются линейной моделью с весами, найденными с помощью нормального уравнения, в случае средних инвестиций в рекламу по ТВ, радио и в газетах? (то есть при нулевых значениях масштабированных признаков TV, Radio и Newspaper).

4. Напишите функцию *linear\_prediction*, которая принимает на вход матрицу  $X$  и вектор весов линейной модели  $w$ , а возвращает вектор прогнозов в виде линейной комбинации столбцов матрицы  $X$  с весами  $w$ .

Какова среднеквадратичная ошибка прогноза значений Sales в виде линейной модели с весами, найденными с помощью нормального уравнения?

5. Напишите функцию *stochastic\_gradient\_step*, реализующую шаг стохастического градиентного спуска для линейной регрессии. Функция должна принимать матрицу  $X$ , вектора  $y$  и  $w$ , число *train\_ind* - индекс объекта обучающей выборки (строки матрицы  $X$ ), по которому считается изменение весов, а также число *eta* - шаг градиентного спуска (по умолчанию *eta*=0.01). Результатом будет вектор обновленных весов.

6. Напишите функцию *stochastic\_gradient\_descent*, реализующую стохастический градиентный спуск для линейной регрессии. Функция принимает на вход следующие аргументы:

- $X$  - матрица, соответствующая обучающей выборке  $y$  - вектор значений
- целевого признака  $w\_init$  - вектор начальных весов модели  $\eta$  - шаг
- градиентного спуска (по умолчанию 0.01)  $max\_iter$  - максимальное число
- итераций градиентного спуска (по умолчанию 10000)  $min\_weight\_dist$  -
- максимальное евклидово расстояние между векторами весов на соседних итерациях градиентного спуска, при котором алгоритм прекращает работу (по умолчанию  $1e-8$ )
- $seed$  - число, используемое для воспроизводимости сгенерированных псевдослучайных чисел (по умолчанию 42)
- $verbose$  - флаг печати информации (например, для отладки, по умолчанию False)

На каждой итерации в вектор (список) должно записываться текущее значение среднеквадратичной ошибки. Функция должна возвращать вектор весов  $w$ , а также вектор (список) ошибок.

Какова среднеквадратичная ошибка прогноза значений Sales в виде линейной модели с весами, найденными с помощью градиентного спуска?

## 20.2 Промежуточная аттестация

**В чём заключается недостаток среднеквадратичной ошибки как метрики качества в задачах регрессии?**

+: Данная метрика неустойчива к выбросам.

-: Имеются проблемы при оптимизации функционала из-за отсутствия производной.

**В каком пространстве градиентный бустинг осуществляет градиентный спуск?**

-: В пространстве коэффициентов при базовых алгоритмах.

+: В пространстве прогнозов алгоритма на объектах обучающей выборки.

-: В пространстве признаков.

-: В пространстве весов при признаках.

## Оценка остаточных знаний

ПК-5. Способен организационно и технологически обеспечивать проектирование и дизайн ИС

Период окончания формирования компетенции: \_\_ семестр (см. УП)

Перечень дисциплин (модулей), практик, участвующих в формировании компетенции:

- Дисциплины (модули): (см УП)
  - Алгоритмы и методы машинного обучения (1 семестр);

Перечень заданий для проверки сформированности компетенции:

1) закрытые задания (тестовые, средний уровень сложности):

1. В каком пространстве градиентный бустинг осуществляет градиентный спуск?

- a) В пространстве коэффициентов при базовых алгоритмах.
- b) В пространстве прогнозов алгоритма на объектах обучающей выборки.**
- c) В пространстве признаков.
- d) В пространстве весов при признаках.

2. В чём заключается недостаток среднеквадратичной ошибки как метрики качества в задачах регрессии?

- a) Данная метрика неустойчива к выбросам.**

- b) Имеются проблемы при оптимизации функционала из-за отсутствия производной.
3. Вам дан набор из 10000 писем, отправленных одним и тем же человеком, и требуется сгруппировать их так, чтобы в одной группе оказались письма на схожие темы — например, личная переписка, письма с авиабилетами и т.д. Что это за задача?
- a) Регрессия.  
b) Классификация.  
c) Кластеризация.
4. Вам нужно предсказать, каким завтра будет курс доллара. Какая это задача?
- a) Регрессия.  
b) Классификация.  
c) Кластеризация.
5. Вам нужно предсказать, повысится или понизится завтра курс доллара. Какая это задача?
- a) Регрессия.  
b) Классификация.  
c) Кластеризация.
6. Градиент какой функции/функционала и по какому аргументу используется в градиентном спуске при обучении линейной регрессии?
- a) Функционала ошибки – например, среднеквадратичной ошибки – по прогнозам алгоритма.  
b) Функционала ошибки – например, среднеквадратичной ошибки – по вектору весов.  
c) Алгоритма – то есть скалярного произведения вектора признаков – по вектору весов.
7. Как в общем устроен процесс построения решающего дерева?
- a) Жадно — начинаем с одной вершины, разбиваем её на две, после чего рекурсивно повторяем процедуру для новых дочерних вершин.  
b) Жадно — начинаем с дерева, у которого в каждом листе находится по одному объекту, и удаляем из него вершины, пока улучшается качество.  
c) Полным перебором — вычисляем качество каждого возможного дерева, выбираем лучшее.  
d) Аналитически — можно в явном виде выписать формулы, задающие структуру оптимального дерева.
8. Как вычисляется предсказание в линейной модели?
- a) Значения всех признаков перемножаются между собой, после чего добавляется свободный член.  
b) Выбирается самый важный признак, после чего восстанавливается линейная зависимость ответа от данного признака.  
c) Значения всех признаков домножаются на некоторые веса и суммируются, после чего добавляется свободный член.
9. Переобучение — это явление, при котором полученный при обучении алгоритм...
- a) показывает на новых данных более низкое качество, чем на обучающей выборке.  
b) показывает сопоставимое качество на обучающей выборке и новых данных.  
c) показывает на обучающей выборке более низкое качество, чем на новых данных.
10. Почему в градиентном спуске на каждой итерации делается шаг в сторону антиградиента?
- a) Антиградиент функционала ошибки зависит только от одного объекта.  
b) Антиградиент легко найти, в отличие от других направлений.  
c) Антиградиент совпадает с направлением наискорейшего убывания.

2). открытые задания (тестовые, повышенный уровень сложности)

1. Каково значение полноты для алгоритма со следующей матрицей ошибок?

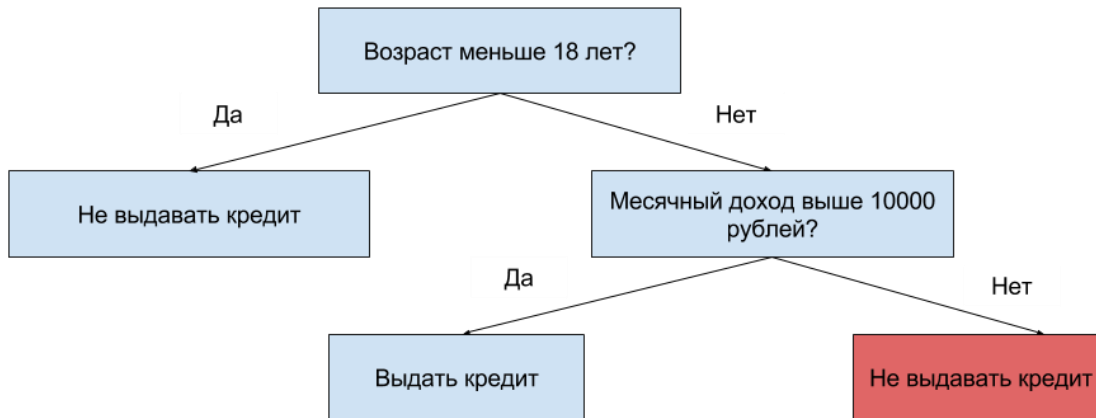
	$y = 1$	$y = -1$
$a(x) = 1$	20	20
$a(x) = -1$	5	80

Ответ 0,8

2. Пусть в выборке 4 признака, и мы решили попробовать добавить квадратичные признаки — то есть все квадраты и попарные произведения исходных признаков. Сколько признаков мы получим после такого расширения?

**Ответ 14**

3. Рассмотрим решающее дерево для решения задачи кредитного скоринга, которое основано на двух признаках: возрасте и месячном доходе.



Пусть дана следующая выборка из пяти объектов (первый признак — возраст, второй — месячный доход):

[20, 8000]

[15, 15000]

[28, 9500]

[24, 30000]

[30, 20000]

Сколько из них попадёт в лист, подсвеченный красным?

**Ответ 2**

3) открытые задания (мини-кейсы, средний уровень сложности):

1. Что такое ансамбли моделей машинного обучения? Какие вы знаете стратегии построения ансамблей?

2. Случайный лес основан на принципе бэггинга — усреднения предсказания нескольких независимых моделей. Но вот проблема: обучающая выборка у нас одна. Как мы можем получить независимые модели? Почему это не приводит к проблемам?

### Ответы на вопросы

Номер вопроса	Ответ
---------------	-------

1.	<p>Классические модели машинного обучения несовершенны. Например, с помощью одного дерева решений вряд ли можно решить сложную задачу. Но если собрать много базовых моделей, то из них может родиться что-то гораздо более интересное. Есть несколько стратегий построения ансамблей:</p> <p>Бэггинг. Его идея в том, чтобы просто агрегировать предсказания, выдаваемые базовыми моделями (например, усреднять или брать самое популярное предсказание). Чтобы результат получился качественным, нужно, чтобы каждая модель была достаточно сильной (основной пример бэггинга — случайный лес — использует глубокие, переобученные решающие деревья).</p> <p>Бустинг. Эта стратегия предполагает, что базовые модели строятся последовательно, и каждая следующая уточняет предсказание всех предыдущих. Обычно в качестве базовых берутся слабые модели, чаще всего деревья небольшой глубины.</p> <p>Стекинг. Предлагается использовать предсказания базовых моделей в качестве новых признаков, а поверх них обучить какую-нибудь ещё модель.</p>
2.	<p>На практике оказывается, что строгое выполнение предположения о независимости не обязательно. Достаточно, чтобы алгоритмы были в некоторой степени не похожи друг на друга. Достаточная непохожесть обеспечивается тем, что при обучении каждого дерева мы:</p> <p>берём случайное подмножество обучающей выборки; берём случайное подмножество признаков.</p>

Критерии оценивания	Шкала оценок (в баллах)
Обучающийся приводит полные и безошибочные ответы.	3 балла
Обучающийся приводит полные и безошибочные ответы. Допускаются незначительные неточности.	2 балла
Ответ представлен частично, есть неверные суждения.	1 балл
Представлены неверные ответы. Присутствуют грубые ошибки или неточности.	0 баллов

ПК-10. Способен определять варианты структур программного обеспечения информационных систем (программного средства), необходимые информационные потоки и исследовать варианты структур с использованием моделей различного уровня

Период окончания формирования компетенции: \_\_\_ семестр (см. УП)

Перечень дисциплин (модулей), практик, участвующих в формировании компетенции:

- Дисциплины (модули): (см УП)
  - Алгоритмы и методы машинного обучения (1 семестр);

Перечень заданий для проверки сформированности компетенции:

1) закрытые задания (тестовые, средний уровень сложности):

1. В каком пространстве градиентный бустинг осуществляет градиентный спуск?

- a) В пространстве коэффициентов при базовых алгоритмах.
- b) В пространстве прогнозов алгоритма на объектах обучающей выборки.**
- c) В пространстве признаков.
- d) В пространстве весов при признаках.



2. В чём заключается недостаток среднеквадратичной ошибки как метрики качества в задачах регрессии?
- Данная метрика неустойчива к выбросам.
  - Имеются проблемы при оптимизации функционала из-за отсутствия производной.
3. Вам дан набор из 10000 писем, отправленных одним и тем же человеком, и требуется сгруппировать их так, чтобы в одной группе оказались письма на схожие темы — например, личная переписка, письма с авиабилетами и т.д. Что это за задача?
- Регрессия.
  - Классификация.
  - Кластеризация.
4. Вам нужно предсказать, каким завтра будет курс доллара. Какая это задача?
- Регрессия.
  - Классификация.
  - Кластеризация.
5. Вам нужно предсказать, повысится или понизится завтра курс доллара. Какая это задача?
- Регрессия.
  - Классификация.
  - Кластеризация.
6. Градиент какой функции/функционала и по какому аргументу используется в градиентном спуске при обучении линейной регрессии?
- Функционала ошибки — например, среднеквадратичной ошибки — по прогнозам алгоритма.
  - Функционала ошибки — например, среднеквадратичной ошибки — по вектору весов.
  - Алгоритма — то есть скалярного произведения вектора признаков — по вектору весов.
7. Как в общем устроен процесс построения решающего дерева?
- Жадно — начинаем с одной вершины, разбиваем её на две, после чего рекурсивно повторяем процедуру для новых дочерних вершин.
  - Жадно — начинаем с дерева, у которого в каждом листе находится по одному объекту, и удаляем из него вершины, пока улучшается качество.
  - Полным перебором — вычисляем качество каждого возможного дерева, выбираем лучшее.
  - Аналитически — можно в явном виде выписать формулы, задающие структуру оптимального дерева.
8. Как вычисляется предсказание в линейной модели?
- Значения всех признаков перемножаются между собой, после чего добавляется свободный член.
  - Выбирается самый важный признак, после чего восстанавливается линейная зависимость ответа от данного признака.
  - Значения всех признаков домножаются на некоторые веса и суммируются, после чего добавляется свободный член.
9. Переобучение — это явление, при котором полученный при обучении алгоритм...
- показывает на новых данных более низкое качество, чем на обучающей выборке.
  - показывает сопоставимое качество на обучающей выборке и новых данных.
  - показывает на обучающей выборке более низкое качество, чем на новых данных.
10. Почему в градиентном спуске на каждой итерации делается шаг в сторону антиградиента?
- Антиградиент функционала ошибки зависит только от одного объекта.
  - Антиградиент легко найти, в отличие от других направлений.
  - Антиградиент совпадает с направлением наискорейшего убывания.

2). открытые задания (тестовые, повышенный уровень сложности)

1. Каково значение полноты для алгоритма со следующей матрицей ошибок?

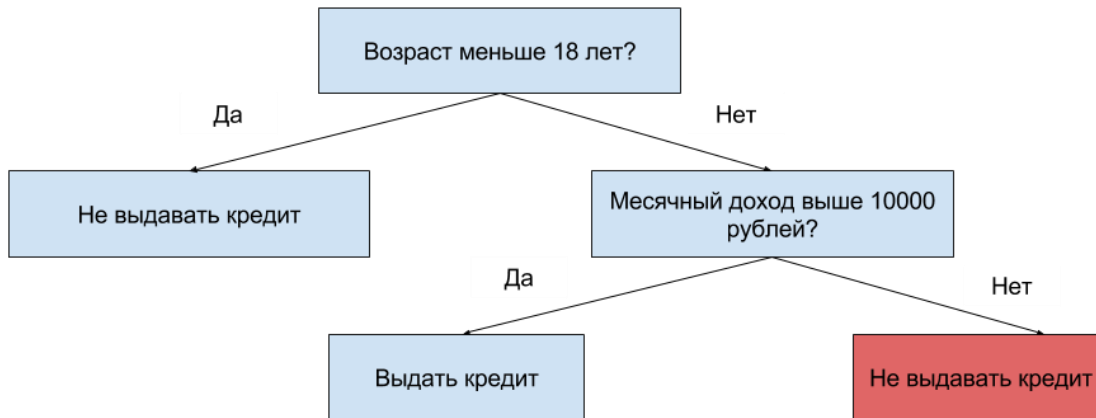
	$y = 1$	$y = -1$
$a(x) = 1$	20	20
$a(x) = -1$	5	80

Ответ 0,8

2. Пусть в выборке 4 признака, и мы решили попробовать добавить квадратичные признаки — то есть все квадраты и попарные произведения исходных признаков. Сколько признаков мы получим после такого расширения?

**Ответ 14**

3. Рассмотрим решающее дерево для решения задачи кредитного скоринга, которое основано на двух признаках: возрасте и месячном доходе.



Пусть дана следующая выборка из пяти объектов (первый признак — возраст, второй — месячный доход):

[20, 8000]

[15, 15000]

[28, 9500]

[24, 30000]

[30, 20000]

Сколько из них попадёт в лист, подсвеченный красным?

**Ответ 2**

3) открытые задания (мини-кейсы, средний уровень сложности):

1. Что такое ансамбли моделей машинного обучения? Какие вы знаете стратегии построения ансамблей?

2. Случайный лес основан на принципе бэггинга — усреднения предсказания нескольких независимых моделей. Но вот проблема: обучающая выборка у нас одна. Как мы можем получить независимые модели? Почему это не приводит к проблемам?

### Ответы на вопросы

Номер вопроса	Ответ
---------------	-------

1.	<p>Классические модели машинного обучения несовершенны. Например, с помощью одного дерева решений вряд ли можно решить сложную задачу. Но если собрать много базовых моделей, то из них может родиться что-то гораздо более интересное. Есть несколько стратегий построения ансамблей:</p> <p>Бэггинг. Его идея в том, чтобы просто агрегировать предсказания, выдаваемые базовыми моделями (например, усреднять или брать самое популярное предсказание). Чтобы результат получился качественным, нужно, чтобы каждая модель была достаточно сильной (основной пример бэггинга — случайный лес — использует глубокие, переобученные решающие деревья).</p> <p>Бустинг. Эта стратегия предполагает, что базовые модели строятся последовательно, и каждая следующая уточняет предсказание всех предыдущих. Обычно в качестве базовых берутся слабые модели, чаще всего деревья небольшой глубины.</p> <p>Стекинг. Предлагается использовать предсказания базовых моделей в качестве новых признаков, а поверх них обучить какую-нибудь ещё модель.</p>
2.	<p>На практике оказывается, что строгое выполнение предположения о независимости не обязательно. Достаточно, чтобы алгоритмы были в некоторой степени не похожи друг на друга. Достаточная непохожесть обеспечивается тем, что при обучении каждого дерева мы:</p> <p>берём случайное подмножество обучающей выборки; берём случайное подмножество признаков.</p>

Критерии оценивания	Шкала оценок (в баллах)
Обучающийся приводит полные и безошибочные ответы.	3 балла
Обучающийся приводит полные и безошибочные ответы. Допускаются незначительные неточности.	2 балла
Ответ представлен частично, есть неверные суждения.	1 балл
Представлены неверные ответы. Присутствуют грубые ошибки или неточности.	0 баллов

#### ПК-14. Способен проектировать архитектуру программного средства

Период окончания формирования компетенции: \_\_ семестр (см. УП)

Перечень дисциплин (модулей), практик, участвующих в формировании компетенции:

- Дисциплины (модули):(см УП)
  - Алгоритмы и методы машинного обучения (1 семестр);

Перечень заданий для проверки сформированности компетенции:

1) закрытые задания (тестовые, средний уровень сложности):

1. В каком пространстве градиентный бустинг осуществляет градиентный спуск?

- a) В пространстве коэффициентов при базовых алгоритмах.
- b) В пространстве прогнозов алгоритма на объектах обучающей выборки.**
- c) В пространстве признаков.
- d) В пространстве весов при признаках.

2. В чём заключается недостаток среднеквадратичной ошибки как метрики качества в задачах регрессии?

- a) Данная метрика неустойчива к выбросам.
- b) Имеются проблемы при оптимизации функционала из-за отсутствия производной.
3. Вам дан набор из 10000 писем, отправленных одним и тем же человеком, и требуется сгруппировать их так, чтобы в одной группе оказались письма на схожие темы — например, личная переписка, письма с авиабилетами и т.д. Что это за задача?
- a) Регрессия.
- b) Классификация.
- c) Кластеризация.
4. Вам нужно предсказать, каким завтра будет курс доллара. Какая это задача?
- a) Регрессия.
- b) Классификация.
- c) Кластеризация.
5. Вам нужно предсказать, повысится или понизится завтра курс доллара. Какая это задача?
- a) Регрессия.
- b) Классификация.
- c) Кластеризация.
6. Градиент какой функции/функционала и по какому аргументу используется в градиентном спуске при обучении линейной регрессии?
- a) Функционала ошибки — например, среднеквадратичной ошибки — по прогнозам алгоритма.
- b) Функционала ошибки — например, среднеквадратичной ошибки — по вектору весов.
- c) Алгоритма — то есть скалярного произведения вектора признаков — по вектору весов.
7. Как в общем устроен процесс построения решающего дерева?
- a) Жадно — начинаем с одной вершины, разбиваем её на две, после чего рекурсивно повторяем процедуру для новых дочерних вершин.
- b) Жадно — начинаем с дерева, у которого в каждом листе находится по одному объекту, и удаляем из него вершины, пока улучшается качество.
- c) Полным перебором — вычисляем качество каждого возможного дерева, выбираем лучшее.
- d) Аналитически — можно в явном виде выписать формулы, задающие структуру оптимального дерева.
8. Как вычисляется предсказание в линейной модели?
- a) Значения всех признаков перемножаются между собой, после чего добавляется свободный член.
- b) Выбирается самый важный признак, после чего восстанавливается линейная зависимость ответа от данного признака.
- c) Значения всех признаков домножаются на некоторые веса и суммируются, после чего добавляется свободный член.
9. Переобучение — это явление, при котором полученный при обучении алгоритм...
- a) показывает на новых данных более низкое качество, чем на обучающей выборке.
- b) показывает сопоставимое качество на обучающей выборке и новых данных.
- c) показывает на обучающей выборке более низкое качество, чем на новых данных.
10. Почему в градиентном спуске на каждой итерации делается шаг в сторону антиградиента?
- a) Антиградиент функционала ошибки зависит только от одного объекта.
- b) Антиградиент легко найти, в отличие от других направлений.
- c) Антиградиент совпадает с направлением наискорейшего убывания.

2). открытые задания (тестовые, повышенный уровень сложности)

1. Каково значение полноты для алгоритма со следующей матрицей ошибок?

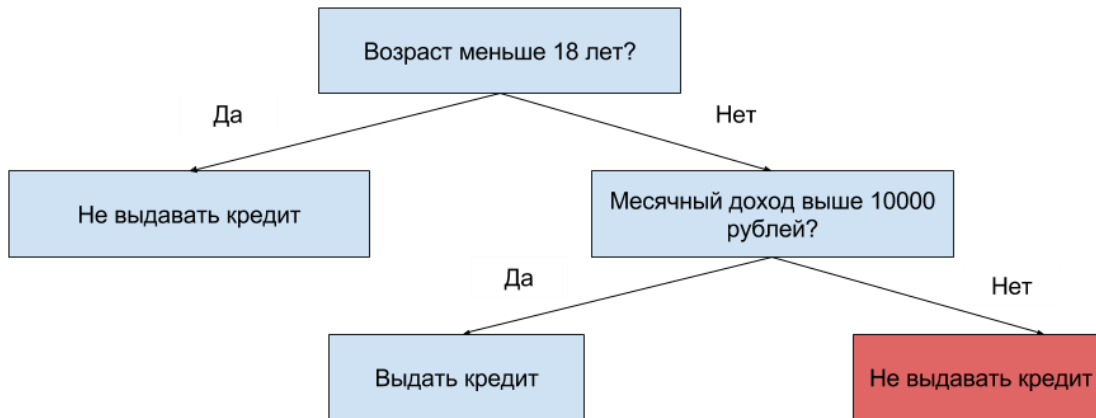
	$y = 1$	$y = -1$
$a(x) = 1$	20	20
$a(x) = -1$	5	80

Ответ 0,8

2. Пусть в выборке 4 признака, и мы решили попробовать добавить квадратичные признаки — то есть все квадраты и попарные произведения исходных признаков. Сколько признаков мы получим после такого расширения?

**Ответ 14**

3. Рассмотрим решающее дерево для решения задачи кредитного скоринга, которое основано на двух признаках: возрасте и месячном доходе.



Пусть дана следующая выборка из пяти объектов (первый признак — возраст, второй — месячный доход):

[20, 8000]

[15, 15000]

[28, 9500]

[24, 30000]

[30, 20000]

Сколько из них попадёт в лист, подсвеченный красным?

**Ответ 2**

3) открытые задания (мини-кейсы, средний уровень сложности):

1. Что такое ансамбли моделей машинного обучения? Какие вы знаете стратегии построения ансамблей?

2. Случайный лес основан на принципе бэггинга — усреднения предсказания нескольких независимых моделей. Но вот проблема: обучающая выборка у нас одна. Как мы можем получить независимые модели? Почему это не приводит к проблемам?

### Ответы на вопросы

Номер вопроса	Ответ
---------------	-------

1.	<p>Классические модели машинного обучения несовершенны. Например, с помощью одного дерева решений вряд ли можно решить сложную задачу. Но если собрать много базовых моделей, то из них может родиться что-то гораздо более интересное. Есть несколько стратегий построения ансамблей:</p> <p>Бэггинг. Его идея в том, чтобы просто агрегировать предсказания, выдаваемые базовыми моделями (например, усреднять или брать самое популярное предсказание). Чтобы результат получился качественным, нужно, чтобы каждая модель была достаточно сильной (основной пример бэггинга — случайный лес — использует глубокие, переобученные решающие деревья).</p> <p>Бустинг. Эта стратегия предполагает, что базовые модели строятся последовательно, и каждая следующая уточняет предсказание всех предыдущих. Обычно в качестве базовых берутся слабые модели, чаще всего деревья небольшой глубины.</p> <p>Стекинг. Предлагается использовать предсказания базовых моделей в качестве новых признаков, а поверх них обучить какую-нибудь ещё модель.</p>
2.	<p>На практике оказывается, что строгое выполнение предположения о независимости не обязательно. Достаточно, чтобы алгоритмы были в некоторой степени не похожи друг на друга. Достаточная непохожесть обеспечивается тем, что при обучении каждого дерева мы:</p> <p>берём случайное подмножество обучающей выборки; берём случайное подмножество признаков.</p>

Критерии оценивания	Шкала оценок (в баллах)
Обучающийся приводит полные и безошибочные ответы.	3 балла
Обучающийся приводит полные и безошибочные ответы. Допускаются незначительные неточности.	2 балла
Ответ представлен частично, есть неверные суждения.	1 балл
Представлены неверные ответы. Присутствуют грубые ошибки или неточности.	0 баллов