

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Астраханский государственный университет имени В. Н. Татищева»
(Астраханский государственный университет им. В. Н. Татищева)

СОГЛАСОВАНО
Руководитель ОПОП

_____ М.В. Коломина

«8» сентября 2022 г.

УТВЕРЖДАЮ
Зав. кафедрой ПМИ

_____ М.В. Коломина

«8» сентября 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

«Автоматическое машинное обучение»

Составитель(и)	Муравьев Сергей Борисович, к.т.н., доцент ИТМО
Направление подготовки / специальность	01.03.02 Прикладная математика и информатика
Направленность (профиль) ОПОП	Программирование и искусственный интеллект
Квалификация (степень)	бакалавр
Форма обучения	очная
Год приёма	2023
Курс	4
Семестр(ы)	8

Астрахань – 2022

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

1.1. Целями освоения дисциплины «Автоматическое машинное обучение» является формирование у студентов компетенций в области автоматического машинного обучения, углубление у студентов умений и навыков в области машинного обучения.

1.2. Задачи освоения дисциплины (модуля):

- изучение ключевых понятий, целей и задач использования автоматического машинного обучения;
- приобретение навыков применения алгоритмов автоматического машинного обучения;
- формирование умений работать с файлами, памятью, графикой, структурами данных, отладкой разработанных программ, настраивать гиперпараметры и выбирать алгоритмы, выбирать и настраивать архитектуры нейронных сетей

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина «Автоматическое машинное обучение» относится к части, формируемой участниками образовательных отношений (элективным дисциплинам) и осваивается в 8 семестре.

2.2. Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые предшествующими учебными дисциплинами:

- Анализ данных
- Машинное обучение

Знания: базовые основы и методы математики, принципы работы компьютеров и программного обеспечения.

Умения: решать типовые теоретические и вычислительные задачи, формулировать задачи машинного обучения и находить оптимальные решения, собирать, обрабатывать и анализировать данные, умение работать с большими объемами данных и использовать инструменты для их обработки.

Навыки: решения типовых математических задач, навыки программирования

2.3. Последующие учебные дисциплины, для которых необходимы знания, умения, навыки, формируемые данной учебной дисциплиной:

- Компьютерное зрение
- Производственная практика
- ВКР

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Процесс освоения дисциплины (модуля) направлен на формирование элементов следующей(их) компетенции(ий) в соответствии с ФГОС ВО и ОПОП ВО по данному направлению подготовки / специальности:

а) профессиональных (ПК):

ПК-18. Способен разрабатывать и применять методы машинного обучения для решения задач

ПК-19. Способен использовать инструментальные средства для решения задач машинного обучения

Таблица 1 – Декомпозиция результатов обучения

Код и наименование компетенции	Планируемые результаты обучения по дисциплине (модулю)		
	Знать (1)	Уметь (2)	Владеть (3)
ПК-18. ПК-18.1. Проводит ана-	ПК-18.1.1. принципы и методы машинного обу-	ПК-18.1.2 сопоставить задачам предметной области классы за-	ПК-18.1.3 методами машинного

Код и наименование компетенции	Планируемые результаты обучения по дисциплине (модулю)		
	Знать (1)	Уметь (2)	Владеть (3)
лиз требований и определяет необходимые классы задач машинного обучения	чения, типы и классы задач машинного обучения, методологию ML Ops; статистические методы анализа данных	дач машинного обучения	обучения и статистическими методами анализа данных
ПК-18. ПК-18.2. Принимает участие в оценке, выборе и при необходимости разработке методов машинного обучения	ПК-18.2.1 классические методы и алгоритмы машинного обучения: предиктивные - обучение с учителем, обучение без учителя	ПК-18.2.2 проводить сравнительный анализ и осуществлять выбор, настройку при необходимости разработку методов и алгоритмов для решения задач машинного обучения	ПК-18.1.3 методами обучения с учителем, обучение без учителя
ПК-19. ПК-19.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи ПК-19.2. Разрабатывает модели машинного обучения для решения задач	ПК-19.1.1. возможности современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения	ПК-19.1.2. проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения	ПК-19.1.3. инструментальными средствами и системами программирования в области создания моделей и методов машинного обучения

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Объём дисциплины (модуля) составляет 2 зачётные единицы, в том числе 22 часа, выделенных на контактную работу обучающихся с преподавателем (из них 22 часа – лабораторные работы), и 50 часов – на самостоятельную работу обучающихся.

Таблица 2 – Структура и содержание дисциплины (модуля)

Раздел, тема дисциплины (модуля)	Семестр	Контактная работа (в часах)			Самост. работа		Форма текущего контроля успеваемости, форма промежуточной аттестации
		Л	ПЗ	ЛР	КР	СР	
Автоматическое машинное обучение для задач обучения с учителем	8	-	-	7	-	16	практическая работа
Автоматическое машинное обучение для задач обучения без учителя	8	-	-	8	-	16	практическая работа
Автоматическое машинное обучение для глубокого обучения	8	-	-	7	-	18	практическая работа
Итого		-	-	22	-	50	Зачёт

Примечание: Л – лекция; ПЗ – практическое занятие, семинар; ЛР – лабораторная работа; КР – курсовая работа; СР – самостоятельная работа.

Таблица 3 – Матрица соотношения разделов, тем учебной дисциплины (модуля) и формируемых компетенций

Раздел, тема дисциплины (модуля)	Кол-во часов	Код компетенции		Общее количество компетенций
		ПК-18	ПК-19	
Автоматическое машинное обучение для задач обучения с учителем	7	+	+	2
Автоматическое машинное обучение для задач обучения без учителя	8	+	+	2
Автоматическое машинное обучение для глубокого обу-	7	+	+	2

Раздел, тема дисциплины (модуля)	Кол-во часов	Код компетенции		Общее количество компетенций
		ПК-18	ПК-19	
чения				
Итого	22			2

Краткое содержание каждой темы дисциплины (модуля)

Раздел 1. Автоматическое машинное обучение для задач обучения с учителем

Предмет и задачи автоматического машинного обучения. Основные принципы, задачи и подходы, использование в различных областях науки и индустрии. Задача одновременного выбора модели и гиперпараметров. Мета-обучение, системы активного тестирования. Подходы на основе SMBO, эволюционные методы в задаче подбора гиперпараметров.

Раздел 2. Автоматическое машинное обучение для задач обучения без учителя

Проблема выбора критерия качества. Подходы автоматического машинного обучения для задач кластеризации: методы на основе обучения с подкреплением, методы на основе эволюционных алгоритмов. Подходы автоматического машинного обучения в задаче выбора признаков, алгоритм MeLiF и его разновидности.

Раздел 3. Автоматическое машинное обучение для глубокого обучения

Задача выбора архитектуры нейронной сети. Эволюционные подходы. Подходы на основе байесовской оптимизации. One-shot модели поиска архитектуры нейронной сети.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРЕПОДАВАНИЮ И ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1. Указания для преподавателей по организации и проведению учебных занятий по дисциплине (модулю)

Лабораторные занятия

Лабораторное занятие – целенаправленная форма организации педагогического процесса, направленная на углубление научно-теоретических знаний и овладение определенными методами работы, в процессе которых вырабатываются умения и навыки выполнения тех или иных учебных действий в данной сфере науки. Они развивают научное мышление и речь, позволяют проверить знания студентов и выступают как средства оперативной обратной связи.

Правильно организованные лабораторные занятия ориентированы на решение следующих задач:

- обобщение, систематизация, углубление, закрепление полученных на лекциях и в процессе
- самостоятельной работы теоретических знаний по дисциплине (предмету);
- формирование практических умений и навыков, необходимых в будущей профессиональной деятельности, реализация единства интеллектуальной и практической деятельности;
- выработка при решении поставленных задач таких профессионально значимых качеств, как самостоятельность, ответственность, точность, творческая инициатива.

Состав заданий для лабораторного занятия должен быть спланирован с расчетом, чтобы за отведенное время они могли быть качественно выполнены большинством учащихся.

Лабораторные занятия должны так быть организованы, чтобы студенты ощущали нарастание сложности выполнения заданий, испытывали бы положительные эмоции от переживания собственного успеха в учении, поисками правильных и точных решений.

Самостоятельная работа

Самостоятельная работа – это вид учебной деятельности, которую студент совершает в установленное время и в установленном объеме индивидуально или в группе, без непосредственной помощи преподавателя (но при его контроле), руководствуясь сформированными ранее представлениями о порядке и правильности выполнения действий.

В учебном процессе образовательного учреждения выделяются два вида самостоятельной работы:

аудиторная – выполняется на учебных занятиях, под непосредственным руководством преподавателя и по его заданию (выполнение самостоятельных работ; выполнение контрольных и практических работ; решение задач);

внеаудиторная – выполняется по заданию преподавателя, но без его непосредственного участия (подготовка к аудиторным занятиям; изучение учебного материала, вынесенного на самостоятельную проработку; выполнение домашних заданий разнообразного характера; выполнение индивидуальных заданий, направленных на развитие у студентов самостоятельности и инициативы; подготовка к контрольной работе). Внеаудиторные самостоятельные работы представляют собой логическое продолжение аудиторных занятий, проводятся по заданию преподавателя, который инструктирует студентов и устанавливает сроки выполнения задания.

5.2. Указания для обучающихся по освоению дисциплины (модулю)

Лабораторное занятие

- Наиболее активный вид учебных занятий в вузе. Он предполагает самостоятельную работу над лекциями и учебными пособиями.
- К каждому занятию нужно готовиться. Подготовку следует начинать с повторения теории (по записям лекций или по учебному пособию). После этого нужно решать задачи из предложенного домашнего задания

Организация самостоятельной работы

Самостоятельность в учебной работе способствует развитию заинтересованности студента в изучаемом материале, вырабатывает у него умение и потребность самостоятельно получать знания, что весьма важно для специалиста с высшим образованием. Самостоятельная работа студентов представлена в следующих формах:

- работа с учебной литературой и конспектом лекций с целью подготовки к лабораторным занятиям, составление конспектов тем, выносимых на самостоятельную проработку;
- систематическое выполнение домашних работ.

Таблица 4 – Содержание самостоятельной работы обучающихся

Вопросы, выносимые на самостоятельное изучение	Кол-во часов	Форма работы
Автоматическое машинное обучение для задач обучения с учителем	16	Изучение теоретического материала, просмотр видеолекций, подготовка к лабораторной работе
Автоматическое машинное обучение для задач обучения без учителя	16	Изучение теоретического материала, просмотр видеолекций, подготовка к лабораторной работе
Автоматическое машинное обучение для глубокого обучения	18	Изучение теоретического материала, просмотр видеолекций, подготовка к лабораторной работе

5.3. Виды и формы письменных работ, предусмотренных при освоении дисциплины (модуля), выполняемые обучающимися самостоятельно

Отчет по лабораторной работе – оформляется и отчитывается в электронном виде: формат листа А4, книжная ориентация страницы. Отчеты по всем лабораторным работам имеют единый титульный лист, на котором указывается наименование дисциплины, ФИО и группа исполнителя, ФИО преподавателя, принимающего отчеты. В отчете по каждой лабораторной работе должно быть представлено наименование работы, цель, ход выполнения работы (скриншоты, краткое текстовое описание), выводы по результатам работы.

6. ОБРАЗОВАТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

6.1. Образовательные технологии

При реализации различных видов учебной работы по дисциплине могут использоваться электронное обучение и дистанционные образовательные технологии.

Учебные занятия по дисциплине могут проводиться с применением информационно телекоммуникационных сетей при опосредованном (на расстоянии) интерактивном взаимодействии обучающихся и преподавателя в режимах on-line или off-line в формах.

Таблица 5 – Образовательные технологии, используемые при реализации учебных занятий

Раздел, тема дисциплины (модуля)	Форма учебного занятия		
	Лекция	Практическое занятие, семинар	Лабораторная работа
Автоматическое машинное обучение для задач обучения с учителем	Не предусмотрено	Не предусмотрено	выполнение лабораторных заданий
Автоматическое машинное обучение для задач обучения без учителя машинное обучение для задач обучения без учителя	Не предусмотрено	Не предусмотрено	выполнение лабораторных заданий
Автоматическое машинное обучение для глубокого обучения	Не предусмотрено	Не предусмотрено	выполнение лабораторных заданий

6.2. Информационные технологии

При реализации различных видов учебной и внеучебной работы используются следующие информационные технологии:

- система управления обучением LMS Moodle;
- использование возможностей Интернета в учебном процессе (рассылка заданий, предоставление выполненных работ, ответы на вопросы, ознакомление обучающихся с оценками и т.д.);
- использование электронных учебников и различных сайтов (например, электронные библиотеки, журналы и т.д.) как источник информации;
- использование возможностей электронной почты;
- использование средств представления учебной информации (электронных учебных пособий, применение новых технологий для проведения занятий с использованием презентаций и т.д.);
- использование интерактивных средств взаимодействия участников образовательного процесса (технологии дистанционного или открытого обучения в глобальной сети);
- использование интегрированных образовательных сред, где главной составляющей являются не только применяемые технологии, но и содержательная часть, т.е. информационные ресурсы (доступ к мировым информационным ресурсам, на базе которых строится учебный процесс).

6.3. Современные профессиональные базы данных и информационные справочные системы

6.3.1. Программное обеспечение

Наименование программного обеспечения	Назначение
Adobe Reader	Программа для просмотра электронных документов
Платформа дистанционного обучения LMS Moodle	Виртуальная обучающая среда
Microsoft Office 2013, Microsoft Office Project 2013, Microsoft Office Visio 2013	Пакет офисных программ
7-zip	Архиватор
Microsoft Windows 7 Professional	Операционная система
Kaspersky Endpoint Security	Средство антивирусной защиты
Google Chrome	Браузер
OpenOffice	Пакет офисных программ

6.3.2. Современные профессиональные базы данных и информационные справочные системы

1. Электронная библиотека «Астраханский государственный университет» собственной генерации на платформе ЭБС «Электронный Читальный зал – БиблиоТех». <https://biblio.asu.edu.ru>
2. Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента». www.studentlibrary.ru.
3. Электронная библиотечная система издательства ЮРАЙТ, раздел «Легендарные книги». www.biblio-online.ru
4. Электронный каталог Научной библиотеки АГУ на базе MARK SQL НПО «Информ-систем». <https://library.asu.edu.ru>

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

7.1. Паспорт фонда оценочных средств

При проведении текущего контроля и промежуточной аттестации по дисциплине «Автоматическое машинное обучение» проверяется сформированность у обучающихся компетенций, указанных в разделе 3 настоящей программы. Этапность формирования данных компетенций в процессе освоения образовательной программы определяется последовательным освоением дисциплин (модулей) и прохождением практик, а в процессе освоения дисциплины (модуля) – последовательным достижением результатов освоения содержательно связанных между собой разделов, тем.

Таблица 6 – Соответствие разделов, тем дисциплины (модуля), результатов обучения по дисциплине (модулю) и оценочных средств

Контролируемый раздел, тема дисциплины (модуля)	Код контролируемой компетенции	Наименование оценочного средства
Автоматическое машинное обучение для задач обучения с учителем	ПК-18, ПК-19	Лабораторная работа
Автоматическое машинное обучение для задач обучения без учителя	ПК-18, ПК-19	Лабораторная работа
Автоматическое машинное обучение для глубокого обучения	ПК-18, ПК-19	Лабораторная работа

7.2. Описание показателей и критериев оценивания компетенций, описание шкал оценивания

Таблица 7 – Показатели оценивания результатов обучения в виде знаний

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует глубокое знание теоретического материала, умение обоснованно излагать свои мысли по обсуждаемым вопросам, способность полно, правильно и аргументированно отвечать на вопросы, приводить примеры
4 «хорошо»	демонстрирует знание теоретического материала, его последовательное изложение, способность приводить примеры, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует неполное, фрагментарное знание теоретического материала, требующее наводящих вопросов преподавателя, допускает существенные ошибки в его изложении, затрудняется в приведении примеров и формулировке выводов
2 «неудовлетворительно»	демонстрирует существенные пробелы в знании теоретического материала, не способен его изложить и ответить на наводящие вопросы преподавателя, не может привести примеры

Таблица 8 – Показатели оценивания результатов обучения в виде умений и владений

Шкала оценивания	Критерии оценивания
5 «отлично»	демонстрирует способность применять знание теоретического материала при выполнении заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы
4	демонстрирует способность применять знание теоретического материала при выполнении

Шкала оценивания	Критерии оценивания
«хорошо»	заданий, последовательно и правильно выполняет задания, умеет обоснованно излагать свои мысли и делать необходимые выводы, допускает единичные ошибки, исправляемые после замечания преподавателя
3 «удовлетворительно»	демонстрирует отдельные, несистематизированные навыки, испытывает затруднения и допускает ошибки при выполнении заданий, выполняет задание по подсказке преподавателя, затрудняется в формулировке выводов
2 «неудовлетворительно»	не способен правильно выполнить задания

7.3. Контрольные задания и иные материалы, необходимые для оценки результатов обучения по дисциплине (модулю)

Лабораторная работа

Пример задания лабораторной работы

Лабораторная работа №3 «Поиск оптимального мета-признакового описания наборов данных для auto-sklearn 1.0»

Порядок выполнения:

1. Рассмотрите 100 наборов данных для задач классификации, например из репозитория OpenML.

2. Рассмотрите принцип работы алгоритма, реализованного в библиотеке auto-sklearn 1.0, которая воплотила в себя два подхода: мета-обучение и байесовскую оптимизацию.

3. Требуется выяснить, для чего используется мета-обучение, каким образом оно влияет на работу алгоритма.

4. Постройте подмножество исходного множества мета-признаков в auto-sklearn 1.0 и вычислите, насколько Ваш алгоритм превосходит имеющийся исходный по времени и по качеству.

5. Дополните исходное мета-признаковое описание auto-sklearn 1.0 и проведите сравнительный анализ с исходным алгоритмом по времени и качеству.

Перечень вопросов и заданий, выносимых на зачёт

Описание технологии проведения:

- формат проведения экзамена: устный зачет в формате ответов на вопросы
- порядок выбора вопросов из общего перечня, их количество для каждого обучающегося: обучающемуся предлагается ответить на два вопроса, по одному вопросу из первой и второй части перечня вопросов к дифференцированному зачету.

Примерный перечень вопросов к зачету:

1. Понятие автоматического машинного обучения. Система Райса.
2. Наивные алгоритмы настройки гиперпараметров. Случайный поиск. Поиск по сетке.
3. Понятие мета-обучения. Мета-обучение для рекомендации алгоритмов.
4. Древовидная оценка Парзена (TPE). Примеры реализаций. TPOT.
5. Байесовская оптимизация. SMBO, SMAC.
6. Эволюционные методы для настройки гиперпараметров.
7. Задача одновременного выбора модели и её гиперпараметров. Общая постановка. Auto-sklearn.
8. Проблема выбора меры в задаче кластеризации.
9. Алгоритм рекомендации меры качества на основе мета-обучения.
10. Метод одновременного выбора и настройки эвристического алгоритма кластеризации на основе обучения с подкреплением.
11. Настройка эволюционного алгоритма кластеризации.
12. Автоматическое машинное обучение в задаче выбора признаков: MeLiF, MeLif+.
13. Поиск архитектуры нейронной сети (NAS). Auto-PyTorch.

Пример билета № 1

1. Байесовская оптимизация. SMBO, SMAC.
2. Автоматическое машинное обучение в задаче выбора признаков: MeLiF, MeLif+.

Таблица 9 – Примеры оценочных средств с ключами правильных ответов

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
<i>ПК-18. Способен разрабатывать и применять методы машинного обучения для решения задач</i>				
1.	Задание закрытого типа	<i>Выберите верный ответ.</i> Какой метод машинного обучения используется для классификации? а. линейная регрессия б. логистическая регрессия в. градиентный спуск г. метод ближайших соседей	б	1-3
2.		<i>Выберите верный ответ.</i> Что такое неконтролируемое обучение? а. обучение на размеченных данных б. обучение без использования данных в. обучение на неразмеченных данных г. обучение с использованием внешнего контроллера	в	1-3
3.		<i>Выберите верный ответ.</i> Какой алгоритм используется для кластеризации данных? а. линейная регрессия б. К-средних в. логистическая регрессия г. метод опорных векторов (SVM)	б	1-3
4.		<i>Выберите верный ответ.</i> Что такое обучение с подкреплением? а. обучение на размеченных данных б. обучение на данных, где каждому действию присваивается вознаграждение или наказание в. обучение на текстовых данных г. обучение на изображениях	б	1-3
5.		<i>Выберите верный ответ.</i>	б	1-3

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		<p>Какой метод используется для уменьшения размерности данных?</p> <p>а. линейная регрессия б. метод главных компонент (РСА) в. логистическая регрессия г. метод ближайших соседей</p>		
6.	Задание открытого типа	Что такое поиск по сетке?	<p>Поиск по сетке используется для точной настройки модели машинного обучения. Предположим, у вас есть список перспективных моделей. И вам нужно их настроить. Итак, как вы это сделаете?</p> <p>Один из вариантов – повозиться с гиперпараметрами вручную, пока вы не найдете отличную комбинацию значений гиперпараметров. Это будет очень утомительная работа, и у вас может не хватить времени на изучение многих комбинаций.</p> <p>Вместо этого вы можете использовать алгоритм поиска по сетке, чтобы найти нужные данные. Все, что вам нужно сделать, это указать ему, с какими гиперпараметрами вы хотите поэкспериментировать и какие значения необходимо попробовать, и он будет использовать перекрестную проверку для оценки всех возможных комбинаций значений гиперпараметров.</p>	8-10
7.		Дайте определение мета-обучению	<p>Мета-обучение — подход, позволяющий определять наиболее подходящий алгоритм (иногда, вместе с параметрами к нему) для конкретной задачи из портфолио алгоритмов. Основная идея мета-обучения — свести задачу выбора алгоритма к задаче обучения с учителем: задачи описываются мета-признаками.</p>	3-5
8.		Дайте определение термину <i>автоматическое машинное обучение</i> .	<p>Автоматическое машинное обучение (англ. Automated Machine Learning, AutoML) — процесс создания динамической комбинации различных методов для формирования простой в использовании сквозной конвейерной системы машинного обучения. AutoML использует хорошо зарекомендовавшие себя методы, которые мы классифицируем в следующие категории на основе конвейера машинного обучения.</p>	3-5
9.		Известно, что конструирование признаков состоит из трёх подэтапов: выбор признаков (англ. feature selection), извлечение признаков (англ. feature extraction) и построение признаков (англ. feature	<p>Конструирование признаков состоит из трёх подэтапов: выбор признаков (англ. feature selection), извлечение признаков (англ. feature extraction) и построение признаков (англ. feature construction). Извлечение и построение признаков — это варианты преобразования, с помощью которых создается новый набор признаков. Во многих случаях,</p>	5-8

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		construction). Опишите эти подэтапы.	целью извлечения признаков является уменьшение исходной размерности путём применения некоторых функций отображения, в то время как построение признаков используется для расширения исходного пространства признаков. Цель выбора признаков состоит в том, чтобы уменьшить избыточность признаков путем выбора наиболее важных из них. В итоге, суть автоматического конструирования признаков в некоторой степени заключается в динамическом сочетании этих трех принципов.	
10.		Опишите способ выбора модели TPOT.	TPOT (Tree-base Pipeline Optimization Tool) В основе TPOT лежит эволюционный алгоритм поиска для нахождения лучшей модели и одновременной оптимизации её гиперпараметров. Представляет собой надстройку над scikit-learn, при этом в данную библиотеку также включены собственные алгоритмы регрессии и классификации. В мае 2020 года вышла версия, в которую был добавлен модуль для работы с нейронными сетями на основе PyTorch. Время работы TPOT сильно зависит от размера входных данных. Не поддерживает обработку естественного языка и категориальных данных.	5-8
11.	Задание комбинированного типа	<i>Верно ли утверждение:</i> Байесовский лэндмарк $P(\theta_{NB}, t_j)$ является не наивным. Поясните ответ.	Утверждение неверно, поскольку Байесовский лэндмарк называется наивным потому что предполагается, что все атрибуты независимы друг от друга.	3-5
ПК-19. Способен использовать инструментальные средства для решения задач машинного обучения				
12.	Задание закрытого типа	<i>Выберите верный ответ.</i> Что такое точность (ассигасу) в контексте оценки модели? а. доля верных предсказаний модели из всех предсказаний б. средняя ошибка модели в. разница между предсказанными и реальными значениями г. скорость обучения модели	а	1-3
13.		<i>Выберите верный ответ.</i> Какой метод используется для уменьшения дисперсии модели? а. увеличение сложности модели б. кросс-валидация в. увеличение размера тест-	б	1-3

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
		вой выборки г. снижение сложности модели		
14.		<p><i>Выберите верный ответ.</i></p> <p>Что такое ансамбль методов в машинном обучении?</p> <p>а. использование одного алгоритма для решения всех задач б. комбинирование нескольких моделей для улучшения предсказаний в. применение методов машинного обучения только к текстовым данным г. использование неконтролируемого обучения</p>	б	1-3
15.		<p><i>Выберите верный ответ.</i></p> <p>Какой алгоритм используется для бинарной классификации?</p> <p>а. линейная регрессия б. логистическая регрессия в. метод главных компонент (РСА) г. К-средних</p>	б	1-3
16.		<p><i>Выберите верный ответ.</i></p> <p>Какой метод используется для борьбы с переобучением?</p> <p>а. увеличение числа слоев в нейронной сети б. регуляризация в. увеличение размера тренировочной выборки г. снижение числа итераций обучения</p>	б	1-3
17.	Задание открытого типа	Процесс тестирования ML включает разделение данных, оценки метрики, проверка на переобучение и анализ ошибок. Опишите эти процессы.	<p>Процесс тестирования</p> <p>Разделение данных: исходные данные делятся на обучающую, валидационную и тестовую выборки. Тестовая выборка используется только в финальном этапе, чтобы обеспечить непредвзятую оценку.</p> <p>Оценка метрик: после обучения модели производится оценка её производительности с использованием различных метрик, таких как точность, полнота и другие.</p> <p>Проверка на переобучение: проверяется, как модель работает на тестовой выборке по</p>	5-8

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
			сравнению с обучающей. Если модель значительно хуже работает на тестовом наборе, возможно, она переобучена. Анализ ошибок: исследование неправильных результатов помогает выявить слабые места модели и возможные улучшения.	
18.		Дайте определение понятия гиперпараметров модели.	Гиперпараметры модели — параметры, значения которых задается до начала обучения модели и не изменяется в процессе обучения. У модели может не быть гиперпараметров. Параметры модели — параметры, которые изменяются и оптимизируются в процессе обучения модели и итоговые значения этих параметров являются результатом обучения модели. Примерами гиперпараметров могут служить количество слоев нейронной сети, а также количество нейронов на каждом слое. Примерами параметров могут служить веса ребер нейронной сети. Для нахождения оптимальных гиперпараметров модели могут применяться различные алгоритмы настройки гиперпараметров	8-10
19.		Сформулируйте задачу выбора модели.	<small>Пусть A — модель алгоритма, характеризующаяся гиперпараметрами $\lambda = \{\lambda_1, \dots, \lambda_m\}$, $\lambda_i \in A_{i1}, \dots, \lambda_m \in A_{im}$. Тогда с ней связано пространство гиперпараметров $\Lambda = A_{11} \times \dots \times A_{im}$. За A_λ обозначим алгоритм, то есть модель алгоритма, для которой задан вектор гиперпараметров $\lambda \in \Lambda$. Для выбора наилучшего алгоритма необходимо зафиксировать меру качества работы алгоритма. Назовем эту меру $Q(A_\lambda, D)$. Задачу выбора наилучшего алгоритма можно разбить на две подзадачи: подзадачу выбора лучшего алгоритма из портфеля и подзадачу настройки гиперпараметров.</small>	8-10
20.		Сформулируйте подзадачу оптимизации гиперпараметров.	Подзадача оптимизации гиперпараметров заключается в подборе таких $\lambda \in \Lambda$, $\lambda \in \Lambda$, при которых заданная модель алгоритма A_λ будет наиболее эффективна. Гиперпараметры могут выбираться из ограниченного множества или с помощью перебора из неограниченного множества гиперпараметров, это зависит от непосредственной задачи. Во втором случае актуален вопрос максимального времени, которое можно потратить на поиск наилучших гиперпараметров, так как чем больше времени происходит перебор, тем лучше гиперпараметры можно найти, но при этом может быть ограничен временной бюджет, из-за чего перебор придется прервать.	8-10
21.		Опишите основную идею алгоритма кросс-валидации.	Основная идея алгоритма кросс-валидации — разбить обучающую выборку на обучающую и тестовую. Таким образом, будет возможным эмулировать наличие тестовой выборки, не участвующей в обучении, но для которой известны правильные ответы. Достоинства и недостатки кросс-валидации: 1. Ошибка в процедуре кросс-валидации является достаточно точной оценкой ошибки на генеральной совокупности; 2. Проведение кросс-валидации требует значительного времени на многократное	8-10

№ п/п	Тип задания	Формулировка задания	Правильный ответ	Время выполнения (в минутах)
			повторное обучение алгоритмов и применимо лишь для «быстрых» алгоритмов машинного обучения; 3. Кросс-валидация плохо применима в задачах кластерного анализа и прогнозирования временных рядов.	
22.	Задание комбинированного типа	<i>Верно ли утверждение</i> Тестирование устойчивости: оценка скорости работы модели и её способности обрабатывать большие объёмы данных в реальном времени. Ответ обоснуйте.	Тестирование устойчивости: проверка того, как модель реагирует на изменения в данных или условиях. Это может включать стресстестирование и анализ, как модель ведёт себя при наличии аномалий в данных.	3-5

7.4. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине (модулю)

Таблица 10 – Технологическая карта рейтинговых баллов по дисциплине

№ п/п	Контролируемые мероприятия	Количество мероприятий/ баллы	Максимальное количество баллов	Срок представления
Основной блок				
1.	<i>Лабораторные работы</i>	6/15	90	
Всего			90	-
Блок бонусов				
2.	<i>Посещение занятий</i>		5	
3.	<i>Своевременное выполнение всех заданий</i>		5	
Всего			10	-
ИТОГО			100	-

Таблица 11 – Шкала перевода рейтинговых баллов в итоговую оценку за семестр по дисциплине

Сумма баллов	Оценка по 4-балльной шкале	
90–100	5 (отлично)	Зачтено
85–89	4 (хорошо)	
75–84		
70–74		
65–69	3 (удовлетворительно)	
60–64		
Ниже 60	2 (неудовлетворительно)	Не зачтено

При реализации дисциплины в зависимости от уровня подготовленности обучающихся могут быть использованы иные формы, методы контроля и оценочные средства, исходя из конкретной ситуации.

8. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

8.1. Основная литература

1. Brazdil Pavel, van Rijn Jan N., Soares Carlos, Vanschoren Joaquin. *Metalearning Applications to Automated Machine Learning and Data Mining*. — Cognitive Technologies Series. 2021 URL: <https://link.springer.com/content/pdf/10.1007/978-3-030-67024-5.pdf>

2. Tanay Agrawal. Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient. — Apress. 2020. ISBN: 9781484265796 DOI <https://doi.org/10.1007/978-1-4842-6579-6>
3. Frank Hutter, Lars Kotthoff, Joaquin Vanschoren. Automated Machine Learning: Methods, Systems, Challenges. — Springer, Cham. 2019. DOI <https://doi.org/10.1007/978-3-030-05318-5>
4. Митяков, Е. С. Искусственный интеллект и машинное обучение : учебное пособие для вузов / Е. С. Митяков, А. Г. Шмелева, А. И. Ладынин. — Санкт-Петербург : Лань, 2025. — 252 с. — ISBN 978-5-507-51465-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/450827>

8.2. Дополнительная литература

1. Лаврищева, Е. М. Программная инженерия и технологии программирования сложных систем : учебник для вузов / Е. М. Лаврищева. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 432 с. — (Высшее образование). — ISBN 978-5-534-07604-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://www.bibliononline.ru/bcode/513067> (Электронная библиотечная система издательства ЮРАЙТ)

8.3. Интернет-ресурсы, необходимые для освоения дисциплины (модуля)

1. Электронно-библиотечная система (ЭБС) ООО «Политехресурс» «Консультант студента». Многопрофильный образовательный ресурс «Консультант студента» является электронной библиотечной системой, предоставляющей доступ через сеть Интернет к учебной литературе и дополнительным материалам, приобретенным на основании прямых договоров с правообладателями. Каталог содержит более 15 000 наименований изданий. www.studentlibrary.ru.
2. Электронная библиотечная система издательства ЮРАЙТ, раздел «Легендарные книги». <https://urait.ru/>

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для проведения лекционных занятий:

1. Используется аудитория, оборудованная необходимым количеством столов, стульев, доской маркерной и электронной.
2. Аудитория должна иметь следующие нормы освещенности
 - СНиП 23-05-95 «Естественное и искусственное освещение» норма освещенности аудиторий ВУЗов 400 Лк.
 - СанПиН 2.2.1/2.1.1.1278-03 «Гигиенические требования к естественному, искусственному и совмещенному освещению жилых и общественных зданий» пункт 3.3.3. «Общее освещение в помещениях общественных зданий должно быть равномерным».
3. Электронная доска должна быть подключена к сети Интернет.

Для проведения лабораторных занятий:

1. Лабораторные занятия проводятся с группами или подгруппами не более 15 человек.
2. Аудитория должна быть оснащена необходимым количеством столов, стульев, доской маркерной и электронной.
4. Аудитория должна иметь следующие нормы освещенности
 - СНиП 23-05-95 «Естественное и искусственное освещение» норма освещенности аудиторий ВУЗов 400 Лк.
 - СанПиН 2.2.1/2.1.1.1278-03 «Гигиенические требования к естественному, искусственному и совмещенному освещению жилых и общественных зданий» пункт 3.3.3. «Общее освещение в помещениях общественных зданий должно быть равномерным».
5. В аудитории должно быть не менее 15 компьютеров, находящихся в исправном состоянии.
6. Расположение компьютеров в аудитории должно позволять преподавателю подойти к рабочему месту студента.

7. Компьютеры должны быть соединены локальной сетью со скоростью не менее 1 Гбит/с и подключены к сети Интернет.
8. Видеокамеры.
9. Компьютеры должны обладать минимальными характеристиками:
 - Объем оперативной памяти 16 Гб
 - Накопитель SDD 500 Гб
 - Процессор 12th Gen Intel(R) Core(TM) i3-12100
 - Видеоадаптер Intel(R) UHD Graphics 730

Рабочая программа дисциплины (модуля) при необходимости может быть адаптирована для обучения (в том числе с применением дистанционных образовательных технологий) лиц с ограниченными возможностями здоровья, инвалидов. Для этого требуется заявление обучающихся, являющихся лицами с ограниченными возможностями здоровья, инвалидами, или их законных представителей и рекомендации психолого-медико-педагогической комиссии. Для инвалидов содержание рабочей программы дисциплины (модуля) может определяться также в соответствии с индивидуальной программой реабилитации инвалида (при наличии).