

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение высшего
образования
"САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ"

Кафедра № 32

УТВЕРЖДАЮ

Руководитель направления

ДОЦ., К.Т.Н., ДОЦ.

(должность, уч. степень, звание)

С.В. Солёный

(инициалы, фамилия)



(подпись)

«08» апреля 2024 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Основы машинного обучения и анализа данных»
(Наименование дисциплины)

Код направления подготовки	15.03.06
Наименование направления подготовки	Мехатроника и робототехника
Наименование направленности	Цифровой инжиниринг робототехнических комплексов
Форма обучения	очная

Лист согласования рабочей программы дисциплины

Программу составил (а)

проф., д.т.н., доц.
(должность, уч. степень, звание)



(подпись, дата)

С.А. Сериков
(инициалы, фамилия)

Программа одобрена на заседании кафедры № 32
«08» апреля 2024 г., протокол №8

Заведующий кафедрой № 32


к.т.н., доц.
(уч. степень, звание)


(подпись, дата)

С.В. Солёный
(инициалы, фамилия)

Ответственный за ОП ВО 15.03.06(02)


доц., к.т.н., доц.
(должность, уч. степень, звание)


(подпись, дата)

О.Я. Солёная
(инициалы, фамилия)

Заместитель директора института №3 по методической работе

ст.преп.
(должность, уч. степень, звание)


(подпись, дата)

Н.В. Решетникова
(инициалы, фамилия)

Аннотация

Дисциплина «Основы машинного обучения и анализа данных» входит в образовательную программу высшего образования – программу бакалавриата по направлению подготовки 15.03.06 «Мехатроника и робототехника» направленности «Цифровой инжиниринг робототехнических комплексов». Дисциплина реализуется кафедрой «№32».

Дисциплина не является обязательной при освоении обучающимся образовательной программы и направлена на углубленное формирование следующих компетенций:

ОПК-11 «Способен разрабатывать и применять алгоритмы и современные цифровые программные методы расчетов и проектирования отдельных устройств и подсистем мехатронных и робототехнических систем с использованием стандартных исполнительных и управляющих устройств, средств автоматизации, измерительной и вычислительной техники в соответствии с техническим заданием, разрабатывать цифровые алгоритмы и программы управления робототехнических систем»

ПК-1 «Способен участвовать в научных исследованиях объектов профессиональной деятельности»

ПК-4 «Способен к выполнению работ по отладке, регулированию, настройке и тестированию мехатронных и робототехнических систем и комплексов»

Содержание дисциплины охватывает круг вопросов, связанных с применением методов машинного обучения и интеллектуального анализа данных при решении прикладных задач в области мехатроники и робототехники.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, лабораторные работы, самостоятельная работа обучающегося.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме дифференцированного зачета.

Общая трудоемкость освоения дисциплины составляет 2 зачетных единицы, 72 часа.

Язык обучения по дисциплине «русский».

1. Перечень планируемых результатов обучения по дисциплине

1.1. Цели преподавания дисциплины

Целью преподавания дисциплины является ознакомление студентов с принципами, методами и инструментальными средствами машинного обучения, как одного из самых популярных и продуктивных направлений искусственного интеллекта, связанного с исследованием и применением самообучающихся алгоритмов, извлекающих знания из данных. Обучающиеся должны изучить особенности основных алгоритмов машинного обучения, получить необходимые навыки для применения систем машинного обучения при решении задач классификации, регрессии, кластеризации, ассоциации, уменьшения размерности.

1.2. Дисциплина является факультативной дисциплиной по направлению образовательной программы высшего образования (далее – ОП ВО).

1.3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОП ВО.

В результате изучения дисциплины обучающийся должен обладать следующими компетенциями или их частями. Компетенции и индикаторы их достижения приведены в таблице 1.

Таблица 1 – Перечень компетенций и индикаторов их достижения

Категория (группа) компетенции	Код и наименование компетенции	Код и наименование индикатора достижения компетенции
Общепрофессиональные компетенции	ОПК-11 Способен разрабатывать и применять алгоритмы и современные цифровые программные методы расчетов и проектирования отдельных устройств и подсистем мехатронных и робототехнических систем с использованием стандартных исполнительных и управляющих устройств, средств автоматизации, измерительной и вычислительной техники в соответствии с техническим заданием, разрабатывать цифровые алгоритмы и программы	ОПК-11.3.2 знает методы разработки алгоритмов и программного обеспечения в рамках систем искусственного интеллекта

	управления робототехнических систем	
Профессиональные компетенции	ПК-1 Способен участвовать в научных исследованиях объектов профессиональной деятельности	ПК-1.3.1 знать методы сбора и анализа научно-технической информации ПК-1.У.3 уметь применять физико-математический аппарат, компьютерные технологии, вычислительные методы и технологии искусственного интеллекта для решения научно-технических задач
Профессиональные компетенции	ПК-4 Способен к выполнению работ по отладке, регулированию, настройке и тестированию мехатронных и робототехнических систем и комплексов	ПК-4.У.1 умеет организовать сбор и обработку данных, необходимых для функционирования робототехнической системы

2. Место дисциплины в структуре ОП

Дисциплина может базироваться на знаниях, ранее приобретенных обучающимися при изучении следующих дисциплин:

Дисциплина может базироваться на знаниях, ранее приобретенных обучающимися при изучении следующих дисциплин:

- «Микропроцессорная техника в мехатронике и робототехнике»;
- «Программное обеспечение мехатронных и робототехнических систем»;
- «Информационные технологии».

Знания, полученные при изучении материала данной дисциплины, имеют как самостоятельное значение, так и используются при изучении других дисциплин:

- «Моделирование и исследование роботов и робототехнических систем»;
- «Управление роботами и робототехническими системами».

3. Объем и трудоемкость дисциплины

Данные об общем объеме дисциплины, трудоемкости отдельных видов учебной работы по дисциплине (и распределение этой трудоемкости по семестрам) представлены в таблице 2.

Таблица 2 – Объем и трудоемкость дисциплины

Вид учебной работы	Всего	Трудоемкость по семестрам
		№6
1	2	3
Общая трудоемкость дисциплины, ЗЕ/ (час)	2/ 72	2/ 72
Из них часов практической подготовки	11	11
Аудиторные занятия, всего час.	34	34
в том числе:		
лекции (Л), (час)	17	17
практические/семинарские занятия (ПЗ),		

(час)		
лабораторные работы (ЛР), (час)	17	17
курсовой проект (работа) (КП, КР), (час)		
экзамен, (час)		
Самостоятельная работа , всего (час)	38	38
Вид промежуточной аттестации: зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.**)	Дифф. Зач.	Дифф. Зач.

Примечание: ** кандидатский экзамен

4. Содержание дисциплины

4.1. Распределение трудоемкости дисциплины по разделам и видам занятий.

Разделы, темы дисциплины и их трудоемкость приведены в таблице 3.

Таблица 3 – Разделы, темы дисциплины, их трудоемкость

Разделы, темы дисциплины	Лекции (час)	ПЗ (СЗ) (час)	ЛР (час)	КП (час)	СРС (час)
Семестр 6					
Введение	1		1		4
Раздел 1. Основные понятия машинного обучения	3		3		6
Раздел 2. Основные модели машинного обучения.	4		4		6
Раздел 3. Реализация алгоритмов машинного обучения.	3		3		6
Раздел 4. Предобработка данных в задачах машинного обучения.	3		3		6
Раздел 5. Оптимизация потоков операций при помощи конвейеров.	2		2		6
Заключение	1		1		4
Итого в семестре:	17		17		38
Итого	17	0	17	0	38

Практическая подготовка заключается в непосредственном выполнении обучающимися определенных трудовых функций, связанных с будущей профессиональной деятельностью.

4.2. Содержание разделов и тем лекционных занятий.

Содержание разделов и тем лекционных занятий приведено в таблице 4.

Таблица 4 – Содержание разделов и тем лекционного цикла

Номер раздела	Название и содержание разделов и тем лекционных занятий
	Введение Предмет дисциплины и ее задачи. Структура и содержание дисциплины. Взаимосвязь изучаемой дисциплины с другими дисциплинами учебного плана. Характеристика основной и дополнительной литературы
1	Основные понятия машинного обучения Общая характеристика методов и разделов, входящих в предметную область –

	<p>машинное обучение. Область применения и задачи, решаемые технологиями машинного обучения, основные понятия и обозначения. Основные библиотеки языка Python для работы с данными.</p> <p>Три вида машинного обучения:</p> <ul style="list-style-type: none"> – обучение с учителем. Основные решаемые задачи. Области применения. <p>Прогнозирование меток классов в задачах классификации. Прогнозирование значений непрерывной целевой переменной на основе регрессионного анализа;</p> <ul style="list-style-type: none"> – обучение с подкреплением. Основные решаемые задачи. Области применения. <p>Основные модели обучения с подкреплением;</p> <ul style="list-style-type: none"> – обучение без учителя. Основные решаемые задачи. Области применения. Работа с немаркированными данными – кластерный анализ. Обнаружение скрытых структур в данных и иерархии кластеров. Снижение размерности для «сжатия данных». <p>Выявление ассоциативных связей.</p>
2	<p>Основные модели машинного обучения.</p> <p>Основные этапы машинного обучения, их назначение и основные особенности:</p> <ul style="list-style-type: none"> – предварительная обработка данных в системах машинного обучения. <p>Назначение. Основные особенности;</p> <ul style="list-style-type: none"> – этап «Обучение» в системах машинного обучения. Принципы формирования тренировочного набора данных. Использование перекрёстной проверки. <p>Гиперпараметрическая оптимизация. Основные метрики качества машинного обучения;</p> <ul style="list-style-type: none"> – этап оценки обобщающей способности в системах машинного обучения. <p>Понятия точности (precision) и верности (accuracy) функционирования моделей.</p> <p>Особенности выбора оптимальной модели машинного обучения. Типовая зависимость между сложностью модели машинного обучения и оценкой её эффективности. Типовая зависимость оценки эффективности моделей машинного обучения от размера обучающей последовательности.</p> <p>Формальная постановка задачи машинного обучения</p>
3	<p>Реализация алгоритмов машинного обучения.</p> <p>Модули NumPy, Pandas, Matplotlib, Scikit-Learn.</p> <p>Строение и функции биологического нейрона. Реализация перцептрона на Python с использованием scikit-learn. Тренировка перцептрона.</p> <p>Адаптивный линейный нейрон (ADALINE). Тренировка ADALINE методами пакетного и стохастического градиентного спуска. Динамическое обучение.</p> <p>Модель логистической регрессии. Тренировка логистической регрессионной модели в scikit-learn. Решение проблемы переобучения при помощи регуляризации.</p> <p>Метод опорных векторов (SVM). Классификация с максимальным зазором и с мягким зазором на основе SVM. Решение нелинейных задач ядерным методом SVM.</p> <p>Модели машинного обучения на основе деревьев решений. Объяснение результатов обучения. Случайный лес.</p> <p>Модель машинного обучения на основе k ближайших соседей.</p> <p>Наивный Байесовский классификатор.</p> <p>Композиции алгоритмов машинного обучения (ансамблевое обучение). Комитет большинства, бэггинг, бустинг, стеккинг.</p>
4	<p>Предобработка данных в задачах машинного обучения.</p> <p>Решение проблемы пропущенных данных. Обработка категориальных данных. Разбивка набора данных на тренировочное и тестовое подмножества. Приведение признаков к одинаковой шкале. Отбор содержательных признаков. Определение важности признаков.</p> <p>Сжатие данных путем снижения размерности. Снижение размерности без учителя на основе анализа главных компонент. Сжатие данных с учителем путем линейного дискриминантного анализа. Ядерный метод анализа главных компонент для нелинейных отображений.</p>

5	Оптимизация потоков операций при помощи конвейеров. Тонкая настройка машинно-обучаемых моделей методом сеточного поиска. Тренировка искусственных нейронных сетей для распознавания изображений.
	Заключение Основные тенденции развития технологий машинного обучения

4.3. Практические (семинарские) занятия

Темы практических занятий и их трудоемкость приведены в таблице 5.

Таблица 5 – Практические занятия и их трудоемкость

№ п/п	Темы практических занятий	Формы практических занятий	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Учебным планом не предусмотрено					
Всего					

4.4. Лабораторные занятия

Темы лабораторных занятий и их трудоемкость приведены в таблице 6.

Таблица 6 – Лабораторные занятия и их трудоемкость

№ п/п	Наименование лабораторных работ	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Семестр 6				
1	Разработка модели линейного адаптивного нейрона (ADALINE) на основе базового класса персептрона	2	2	1
2	Классификация на основе персептронной модели с использованием библиотеки Scikit-learn	3	3	2
3	Классификация на основе модели логистической регрессии с использованием библиотеки Scikit-learn	3	3	3
4	Классификация методом опорных векторов с использованием библиотеки Scikit-learn	3	3	4
5	Классификация линейно разделимых объектов (три признака, четыре класса) с использованием библиотеки Scikit-learn	3	3	5
6	Классификация линейно неразделимых объектов с использованием библиотеки Scikit-learn	3	3	6
Всего		17	17	

4.5. Курсовое проектирование/ выполнение курсовой работы

Учебным планом не предусмотрено

4.6. Самостоятельная работа обучающихся

Виды самостоятельной работы и ее трудоемкость приведены в таблице 7.

Таблица 7 – Виды самостоятельной работы и ее трудоемкость

Вид самостоятельной работы	Всего, час	Семестр 6, час
1	2	3
Изучение теоретического материала дисциплины (ТО)	18	18
Курсовое проектирование (КП, КР)		
Расчетно-графические задания (РГЗ)		
Выполнение реферата (Р)		
Подготовка к текущему контролю успеваемости (ТКУ)	10	10
Домашнее задание (ДЗ)		
Контрольные работы заочников (КРЗ)		
Подготовка к промежуточной аттестации (ПА)	10	10
Всего:	38	38

5. Перечень учебно-методического обеспечения

для самостоятельной работы обучающихся по дисциплине (модулю)

Учебно-методические материалы для самостоятельной работы обучающихся указаны в п.п. 7-11.

6. Перечень печатных и электронных учебных изданий

Перечень печатных и электронных учебных изданий приведен в таблице 8.

Таблица 8– Перечень печатных и электронных учебных изданий

Шифр/ URL адрес	Библиографическая ссылка	Количество экземпляров в библиотеке (кроме электронных экземпляров)
УДК 681.3.07	Рашка Себастьян, Мирджалили Вахид. Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и TensorFlow 2, 3-е изд. Пер. с англ. – СПб.: «Диалектика», 2020. – 848 с.	
УДК: 681.3.07	Жерон Орельен. Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем, 2-е изд.: Пер. с англ.– СПб.: ООО «Диалектика», 2020.–1040 с.	
УДК 004.43 П37	Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. – СПб.: Питер, 2018. – 576 с.	

УДК 681.3.07	Джоши Прадик. Искусственный интеллект с примерами на Python.: Пер. с англ. - СПб.: ООО "Диалектика", 2019. – 448 с.	
ISBN 978-5-9908910-8-1	Андреас Мюллер, Сара Гвидо. Введение в машинное обучение с помощью python. Руководство для специалистов по работе с данными. – М.: ИЦ «Гевисста», 2017. – 393 с.	
УДК 0004.438	Луис Педро Коэльо, Вилли Ричарт. Построение систем машинного обучения на языке Python. 2-е издание/ пер. с англ. Слинкин А.А. – М.: ДМК Пресс, 2016. – 302 с.	
УДК 681.3.07	Элбон Крис. Машинное обучение с использованием Python. Сборник рецептов: Пер. с англ. — СПб.: БХВ-Петербург, 2019. — 384 с.: ил.	
УДК 004.62	Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и науки о данных. – СПб.: Питер, 2017. – 336 с.	

7. Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины приведен в таблице 9.

Таблица 9 – Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

URL адрес	Наименование
http://www.machinelearning.ru/wiki/index.php?title=Заглавная_страница	Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных
http://python.org	Официальный сайт языка Python
https://www.anaconda.com/download/	Anaconda - Дистрибутив Python,

8. Перечень информационных технологий

8.1. Перечень программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине.

Перечень используемого программного обеспечения представлен в таблице 10.

Таблица 10– Перечень программного обеспечения

№ п/п	Наименование
1	Anaconda 2020.11 For Windows Installer
2	Python 3.7 version
3	Spyder 4.2.3

8.2. Перечень информационно-справочных систем,используемых при осуществлении образовательного процесса по дисциплине

Перечень используемых информационно-справочных систем представлен в таблице 11.

Таблица 11– Перечень информационно-справочных систем

№ п/п	Наименование
1	Онлайн-книга «Справочник по науке о данных Python» https://jakevdp.github.io/PythonDataScienceHandbook/05.06-linear-regression.html
2	Датасеты по машинному обучению http://archive.ics.uci.edu/ml/datasets.php

9. Материально-техническая база

Состав материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине, представлен в таблице 12.

Таблица 12 – Состав материально-технической базы

№ п/п	Наименование составной части материально-технической базы	Номер аудитории (при необходимости)
1	Мультимедийная лекционная аудитория	Б.М. 21-21
2	Компьютерный класс	Б.М. 31-04

10. Оценочные средства для проведения промежуточной аттестации

10.1. Состав оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведен в таблице 13.

Таблица 13 – Состав оценочных средств для проведения промежуточной аттестации

Вид промежуточной аттестации	Перечень оценочных средств
Дифференцированный зачёт	Список вопросов; Тесты; Задачи.

10.2. В качестве критериев оценки уровня сформированности (освоения) компетенций обучающимися применяется 5-балльная шкала оценки сформированности компетенций, которая приведена в таблице 14. В течение семестра может использоваться 100-балльная шкала модульно-рейтинговой системы Университета, правила использования которой, установлены соответствующим локальным нормативным актом ГУАП.

Таблица 14 – Критерии оценки уровня сформированности компетенций

Оценка компетенции 5-балльная шкала	Характеристика сформированных компетенций
«отлично» «зачтено»	<ul style="list-style-type: none"> – обучающийся глубоко и всесторонне усвоил программный материал; – уверенно, логично, последовательно и грамотно его излагает; – опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий.
«хорошо» «зачтено»	<ul style="list-style-type: none"> – обучающийся твердо усвоил программный материал, грамотно и по существу излагает его, опираясь на знания основной литературы; – не допускает существенных неточностей; – увязывает усвоенные знания с практической деятельностью направления; – аргументирует научные положения; – делает выводы и обобщения; – владеет системой специализированных понятий.
«удовлетворительно» «зачтено»	<ul style="list-style-type: none"> – обучающийся усвоил только основной программный материал, по существу излагает его, опираясь на знания только основной литературы; – допускает несущественные ошибки и неточности; – испытывает затруднения в практическом применении знаний направления; – слабо аргументирует научные положения; – затрудняется в формулировании выводов и обобщений; – частично владеет системой специализированных понятий.
«неудовлетворительно» «не зачтено»	<ul style="list-style-type: none"> – обучающийся не усвоил значительной части программного материала; – допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении; – испытывает трудности в практическом применении знаний; – не может аргументировать научные положения; – не формулирует выводов и обобщений.

10.3. Типовые контрольные задания или иные материалы.

Вопросы (задачи) для экзамена представлены в таблице 15.

Таблица 15 – Вопросы (задачи) для экзамена

№ п/п	Перечень вопросов (задач) для экзамена	Код индикатора
	Учебным планом не предусмотрено	

Вопросы (задачи) для зачета / дифф. зачета представлены в таблице 16.

Таблица 16 – Вопросы (задачи) для зачета / дифф. зачета

№ п/п	Перечень вопросов (задач) для зачета / дифф. зачета	Код индикатора
1	1. Дайте определение понятию «Искусственный интеллект». Поясните термины «Слабый искусственный интеллект» и «Сильный искусственный интеллект». Перечислите особенности слабоформализованных задач как основной области применения технологий искусственного интеллекта. Приведите примеры таких задач.	ОПК-11.3.2

	<p>2. По какому признаку можно выделить интеллектуальные задачи среди других задач обработки информации? Приведите примеры интеллектуальных задач. Поясните, что называют интеллектуальными информационными технологиями.</p> <p>3. Дайте определение понятию «Машинное обучение». С чем связано свойство обучения в таких моделях? Перечислите основные факторы, влияющие на эффективность использования методов машинного обучения, и дайте им краткую характеристику.</p> <p>4. Приведите классификацию систем машинного обучения по объёму и типу контроля во время обучения, в зависимости от способности обучаться на основе потока входящих данных, по способу обобщения знаний. Дайте краткую характеристику перечисленным классам систем.</p> <p>5. Поясните особенности обучения с учителем как вида машинного обучения. Приведите постановку задачи обучения с учителем. Перечислите типы задач обучения с учителем, в зависимости от вида множества допустимых ответов. Поясните, что называется признаковым описанием объекта из множества исследуемых объектов. Какие типы признаков существуют?</p> <p>6. Перечислите основные этапы решения задачи машинного обучения с учителем. Поясните содержание каждого этапа.</p>	
2	<p>7. Поясните особенности обучения без учителя как вида машинного обучения. Перечислите наиболее характерные области использования обучения без учителя. Какие типы задач решаются методами обучения без учителя? Дайте краткую характеристику этим задачам.</p> <p>8. Поясните особенности обучения с подкреплением как вида машинного обучения. Приведите постановку задачи обучения с подкреплением. Опишите схему взаимодействия агента со средой при реализации обучения с подкреплением. Перечислите характерные области применения обучения с подкреплением.</p> <p>9. Поясните особенности оценки качества работы классификатора. Почему коэффициент ошибочных классификаций и правильность классификации не всегда являются эффективными показателями качества классификации? Что такое матрица неточностей? Как определить точность и полноту классификации с использованием матрицы неточностей? Какие особенности классификатора характеризуют точность и полнота?</p> <p>10. Как выглядят типовые зависимости точности от полноты классификации, а также точности и полноты от порога принятия решения? Поясните такое поведение данных зависимостей. Поясните, что такое «мера F_1», как она определяется и для чего используется.</p> <p>11. Поясните, как используется кривая рабочей характеристики приёмника (ROC) для оценки качества работы классификатора. Приведите характерный вид данной кривой. Как выглядит кривая рабочей характеристики приёмника при чисто случайной классификации? Для каких целей обычно используется измерение площади под кривой рабочей характеристики приёмника (AUC)?</p> <p>12. Поясните, как строится матрица неточностей при многоклассовой классификации. Как можно оценить точность и полноту для каждого класса, а также результирующие значения</p>	ПК-1.3.1

	<p>точности и полноты классификации.</p> <p>13. Поясните особенности оценки качества работы моделей регрессии. В каких случаях отдают предпочтение критерию качества RMSE, а в каких случаях критерию MAE?</p>	
3	<p>14. В чём состоит свойство обобщающей способности модели машинного обучения. Какие приёмы используются для теоретической и эмпирической оценки обобщающей способности.</p> <p>15. Назовите основные причины возникновения переобучения и недообучения в моделях машинного обучения. Перечислите основные способы минимизации переобучения.</p> <p>16. Поясните понятия смещения, дисперсии и неустранимой погрешности как трёх составляющих ошибки обобщения моделей машинного обучения.</p> <p>17. Поясните характер изменения оценок эффективности моделей машинного обучения при возрастании их сложности и размера обучающей выборки.</p> <p>18. Персептрон Розенблатта. Структурная схема и правило обучения персептрона.</p> <p>19. Модель адаптивного линейного нейрона. Обучение ADALINE при помощи метода градиентного спуска. Пакетный и стохастический градиентный спуск.</p> <p>20. Модель логистической регрессии. Покажите, что выход сигмоидальной функции можно интерпретировать как вероятность принадлежности отдельно взятого образца первому классу.</p> <p>21. Обучение модели логистической регрессии. Решение проблемы переобучения при помощи регуляризации. L2-регуляризация</p>	ПК-1.У.3
4	<p>22. Метод опорных векторов. Классификация с мягким зазором для обеспечения сходимости оптимизации при допустимости ошибочной классификации.</p> <p>23. Решение нелинейных задач при помощи метода опорных векторов. Добавление к набору данных дополнительных признаков. Использование признаков близости. RBF-функция как функция подобия.</p> <p>24. Гиперпараметрическая оптимизация моделей машинного обучения. Использование решётчатого поиска для выбора наилучших гиперпараметров модели.</p> <p>25. Модель на основе дерева принятия решений. Основные особенности модели. Построение дерева решений на основе принципа наибольшего прироста информации. Гиперпараметры, обеспечивающие подстройку модели.</p> <p>26. Модель на основе случайного леса. Особенности модели. Алгоритм «выращивания» случайного леса. Гиперпараметры, обеспечивающие подстройку модели.</p> <p>27. Классификатор на основе k ближайших соседей и его особенности. Гиперпараметры, обеспечивающие подстройку модели.</p> <p>28. Наивный байесовский классификатор. Основные особенности модели и области её применения.</p>	ПК-4.У.1

Перечень тем для курсового проектирования/выполнения курсовой работы представлены в таблице 17.

Таблица 17 – Перечень тем для курсового проектирования/выполнения курсовой работы

№ п/п	Примерный перечень тем для курсового проектирования/выполнения курсовой работы
	Учебным планом не предусмотрено

Вопросы для проведения промежуточной аттестации в виде тестирования представлены в таблице 18.

Таблица 18 – Примерный перечень вопросов для тестов

№ п/п	Примерный перечень вопросов для тестов	Код индикатора
1.1	Системы, свойства которых не сводятся к сумме свойств составляющих их элементов, называют: а) простыми б) сложными в) комплексные Ответ: б	ОПК-11.3.2 ПК-1.3.1 ПК-1.У.3
1.2	Информация фактического характера, описывающая объекты, процессы и явления конкретной предметной области это а) знания б) данные в) эвристика Ответ: б	ПК-4.У.1
1.3	Под алгоритмом понимают точное предписание о выполнении в определенном порядке системы: а) операций б) функций в) задач Ответ: а	
1.4	Задачи, связанные с отысканием алгоритма решения класса задач определенного типа это: а) прикладная задача б) интеллектуальная задача в) прагматическая задача Ответ: б	
1.5	Что понимается под обработкой знаний: а) понимание естественного языка б) поддержка принятия решения в сложных ситуациях в) анализ визуальной информации г) все ответы Ответ: г	
1.6	Знания можно разделить на следующие категории: а) глубинные знания, поверхностные знания б) процедурные знания, декларативные знания в) все ответы верны Ответ: в	

1.7 Какое из направлений исследований искусственного интеллекта обеспечивает решение задач классификации, регрессии, кластеризации, ассоциации, уменьшения размерности:

- а) системы нечёткого вывода
- б) классическое машинное обучение
- в) генетические алгоритмы

Ответ: б

1.8 Обучение, при котором параметры модели обновляются на основе каждого поступившего тренировочного образца:

- а) динамическое
- б) пакетное
- в) мини-пакетное

Ответ: а

1.9 По способу обобщения знаний различают:

- а) обучение на основе образцов
- б) обучение на основе моделей
- в) оба ответа верны

Ответ: в

1.10 Прогнозирование метки класса для новых экземпляров на основе предыдущих наблюдений является задачей:

- а) кластеризации
- б) классификации
- в) регрессии

Ответ: в

2.1 Сложные системы характеризуются ...

- а) большим количеством параметров
- б) разнородностью элементов
- в) иерархичность структуры
- г) все ответы верны

Ответ: г

2.2 В математике и кибернетике класс задач определенного типа считается решенным, когда для его решения:

- а) найдены алгоритмы
- б) получены результаты
- в) определены цели

Ответ: а

2.3 Какое из направлений исследований искусственного интеллекта использует случайный подбор, комбинирование и вариации искомых параметров с применением механизмов, напоминающих биологическую эволюцию:

- а) искусственные нейронные сети
- б) классическое машинное обучение
- в) генетические алгоритмы

Ответ: в

2.4 Подход, который используется для сжатия данных в подпространство

меньшей размерности при сохранении большей части релевантной информации:

- а) таксономия
- б) снижение размерности
- в) поиск ассоциаций

Ответ: б

2.5 Метрика, определяющая долю истинно положительных классификаций, которую иногда называют “чувствительностью”:

- а) полнота
- б) объём
- в) точность

Ответ: а

2.6 Цель применения эвристик:

- а) найти рациональное, приемлемое решение
- б) найти оптимальное решение
- в) найти наиболее полное решение

Ответ: а

2.7 Точное предписание о выполнении в определенном порядке системы операций для решения любой задачи из некоторого заданного класса задач это:

- а) данные
- б) системы
- в) алгоритмы

Ответ: в

2.8 Какое из направлений исследований искусственного интеллекта использует принцип организации и функционирования биологических сетей нервных клеток живого организма для обработки информации:

- а) искусственные нейронные сети
- б) классическое машинное обучение
- в) генетические алгоритмы

Ответ: а

2.9 Обучение, предполагающее обновление параметров модели на основе суммы накопленных ошибок по всем образцам обучающей выборки:

- а) динамическое
- б) пакетное
- в) мини-пакетное

Ответ: б

3.1 Через какую точку должна пройти ROC-кривая в случае идеального классификатора?

- а) (0,0)
- б) **(0,1)**
- с) (1,1)

Ответ: б

3.2 Какое дерево называется бинарным?

- а) Если при его построении используются только бинарные признаки

- б) Если оно имеет только два листа
- с) Если из любой его внутренней вершины выходит ровно два ребра**

Ответ: с

3.3 Можно ли при построении бинарного дерева использовать небинарные признаки?

- а) Да**
- б) Нет

Ответ: а

3.4 Что такое бутстрап?

- а) один из подходов к построению подвыборок**
- б) метод обучения
- с) минимизируемый функционал

Ответ: а

3.5 Что такое бутстрапированная выборка?

- а) это некоторый сегмент обучающей выборки произвольной длины
- б) это выборка, полученная из обучающей выборки с возможным повторением объектов**
- с) это выборка, из которой удалены неизвестные значения признаков

Ответ: б

3.6 Какой длины будет бутстрапированная выборка?

- а) больше длины исходной выборки
- б) меньше длины исходной выборки
- с) равной длине исходной выборки**

Ответ: с

3.7 Какова зависимость корреляции между деревьями в случайном лесе от параметра q (размер случайного подмножества признаков)?

- а) Чем меньше q , тем больше корреляция
- б) Чем больше q , тем больше корреляция**
- с) Данный параметр не влияет на корреляцию

Ответ: б

3.8 Градиентный спуск – это * при использовании в машинном обучении

- а) алгоритм оптимизации параметров модели**
- б) алгоритм поиска минимального значения гиперпараметров
- с) алгоритм оптимизации архитектуры модели
- д) алгоритм поиска максимального значения функции потерь

Ответ: а

3.9 Какой "девиз" у парадигмы ансамблирования?

- а) Много слабых моделей даёт результат лучше, чем одна сильная**
- б) Много сильных моделей вместе даёт во всем лучший результат
- с) Много моделей даёт хуже результат, чем одна

Ответ: а

3.10 Оценка корреляции по Пирсону оказала значение 0 между

переменными. Что это означает?

- a) **Переменные не имеют линейной взаимосвязи**
- б) Переменные имеют нелинейную взаимосвязь
- с) Переменные не имеют нелинейную взаимосвязь
- д) Переменные имеют максимальную линейную взаимосвязь (так как значение оценки 1 - отсутствие взаимосвязи)

Ответ: а

3.11 Для чего применяется кросс-валидация:

- a) **Поиск оптимальных гиперпараметров при малом количестве данных для обучения**
- б) Поиск оптимальных гиперпараметров при большом количестве данных для обучения
- с) Поиск оптимальных гиперпараметров при отсутствии данных для обучения
- д) Чтобы не использовать тестовую выборку для оценки метрик

Ответ: а

3.12 В чем заключается метод случайных подпространств?

- а) Генерация случайного подмножества обучающей выборки для дальнейшего обучения базового алгоритма
- б) **Получение случайного подмножества из исходного количества признаков для дальнейшего обучения базового алгоритма**
- с) Генерация случайного подмножества базовых алгоритмов для получения агрегированного результата

Ответ: б

3.13 Какая из метрик лучше подходит для поиска сильных (больших) ошибок модели в задаче регрессии:

- a) **RMSE**
- б) MAE
- с) R^2
- д) F1

Ответ: а

3.14 Допустим, мы предсказываем сорт цветков (один из трех: F, G, E) и оцениваем метрики класса F.

Какой из показателей классификации отражает количество правильно предсказанных ответов оцениваемого класс.

- a) **True Positive**
- б) False Positive
- с) True Negative
- д) False Negative

Ответ: а

3.15 Какая из метрик может принимать отрицательные значения?

- a) **R^2**
- б) MAE
- с) RMSE
- д) Accuracy

Ответ: а

3.16 При наличии сильного перекоса классов по целевой переменной (1000 записей класса А и 40 записей класса Б) какая из метрик может исказить оценку работы модели?

- а) **accuracy**
- б) recall
- с) precision
- д) F1

Ответ: а

3.17 Как бороться с эффектом переобучения?

- а) **Сделать модель проще (менее комплексной)**
- б) Взять меньше данных
- с) Сделать модель сложнее (более комплексной)
- д) Проверять метрики только по обучающей выборке

Ответ: а

3.18 Как связаны между собой плотность распределения и функция распределения?

- а) **Плотность распределения - это производная функции распределения**
- б) Функция распределения - это производная плотности распределения
- с) Эти величины никак не связаны

Ответ: а

3.19 В чем заключается правило трёх сигм?

- а) Три среднеквадратических отклонения - это доверительный интервал для математического ожидания
- б) **Вероятность того, что случайная величина примет значение, отклоняющееся от математического ожидания больше чем на три среднеквадратических отклонения пренебрежимо мала**
- с) Три среднеквадратических отклонения - это максимально возможное значение случайной величины

Ответ: б

3.20 К какому типу признаков Вы бы отнесли рост человека (в см)?

- а) Бинарный
- б) **Вещественный**
- с) Категориальный
- д) Порядковый

Ответ: б

4.1 К какому типу признака Вы бы отнести цвет глаз человека?

- а) Бинарный
- б) Вещественный
- с) **Категориальный**
- д) Порядковый

Ответ: с

4.2 Какой из подходов регуляризации позволяет обнулить веса линейной модели, тем самым показать ненужность признака?

- а) **Лассо**

б) L2 с) Ридж д) Тихонова Ответ: а 4.3. Может ли решающее дерево восстанавливать нелинейные зависимости? а) Да б) Нет	
---	--

Перечень тем контрольных работ по дисциплине обучающихся заочной формы обучения, представлены в таблице 19.

Таблица 19 – Перечень контрольных работ

№ п/п	Перечень контрольных работ
	Не предусмотрено

10.4. Методические материалы, определяющие процедуры оценивания индикаторов, характеризующих этапы формирования компетенций, содержатся в локальных нормативных актах ГУАП, регламентирующих порядок и процедуру проведения текущего контроля успеваемости и промежуточной аттестации обучающихся ГУАП.

11. Методические указания для обучающихся по освоению дисциплины

11.1. Методические указания для обучающихся по освоению лекционного материала.

Основное назначение лекционного материала – логически стройное, системное, глубокое и ясное изложение учебного материала. Назначение современной лекции в рамках дисциплины не в том, чтобы получить всю информацию по теме, а в освоении фундаментальных проблем дисциплины, методов научного познания, новейших достижений научной мысли. В учебном процессе лекция выполняет методологическую, организационную и информационную функции. Лекция раскрывает понятийный аппарат конкретной области знания, её проблемы, дает цельное представление о дисциплине, показывает взаимосвязь с другими дисциплинами.

Планируемые результаты при освоении обучающимися лекционного материала:

- получение современных, целостных, взаимосвязанных знаний, уровень которых определяется целевой установкой к каждой конкретной теме;
- получение опыта творческой работы совместно с преподавателем;
- развитие профессионально-деловых качеств, любви к предмету и самостоятельного творческого мышления.
- появление необходимого интереса, необходимого для самостоятельной работы;
- получение знаний о современном уровне развития науки и техники и о прогнозе их развития на ближайшие годы;
- научиться методически обрабатывать материал (выделять главные мысли и положения, приходить к конкретным выводам, повторять их в различных формулировках);
- получение точного понимания всех необходимых терминов и понятий.

Лекционный материал может сопровождаться демонстрацией слайдов и использованием раздаточного материала при проведении коротких дискуссий об особенностях применения отдельных тематик по дисциплине.

Структура предоставления лекционного материала:

Лекция состоит из трёх основных частей: вступительной, основной и заключительной.

Вступительная часть определяет название темы, план и цель лекции. Она призвана заинтересовать и настроить аудиторию. В этой части лекции излагается актуальность, основная идея, связь данной лекции с предыдущими занятиями, ее основные вопросы. Введение должно быть кратким и целенаправленным.

В основной части лекции реализуется научное содержание темы, все главные узловые вопросы, проводится вся система доказательств с использованием наиболее целесообразных методических приёмов. Каждый учебный вопрос заканчивается краткими выводами, логически подводящими студентов к следующему вопросу лекции.

Заключительная часть имеет целью обобщать в кратких формулировках основные идеи лекции, логически завершая её как целостное творение.

Отдельные виды лекций могут иметь свои особенности как по содержанию, так и по структуре.

11.2. Методические указания для обучающихся по выполнению лабораторных работ.

В ходе выполнения лабораторных работ обучающийся должен углубить и закрепить знания, практические навыки, овладеть современной методикой и техникой эксперимента в соответствии с квалификационной характеристикой обучающегося. Выполнение лабораторных работ состоит из экспериментально-практической, расчетно-аналитической частей и контрольных мероприятий.

Выполнение лабораторных работ обучающимся является неотъемлемой частью изучения дисциплины, определяемой учебным планом, и относится к средствам, обеспечивающим решение следующих основных задач обучающегося:

- приобретение навыков исследования процессов, явлений и объектов, изучаемых в рамках данной дисциплины;
- закрепление, развитие и детализация теоретических знаний, полученных на лекциях;
- получение новой информации по изучаемой дисциплине;
- приобретение навыков самостоятельной работы с лабораторным оборудованием и приборами.

Задание и требования к проведению лабораторных работ

Лабораторные работы, предусмотренные в данном курсе, выполняются в компьютерном классе на персональном компьютере с использованием пакетов поддержки разработки и отладки программного обеспечения на языке Python.

В процессе подготовки к лабораторной работе необходимо изучить соответствующие методические указания и повторить лекционный материал, который относится к теме работы.

Перед началом выполнения работы необходимо создать отдельную папку для создаваемых в работе файлов и установить её в системе качестве текущей директории. Путь к данной папке не должен содержать имён, написанных кириллицей.

В процессе выполнения работы полученные результаты расчётов, листинги разрабатываемых программ, схемы и другие рабочие материалы должны сохраняться на диске для их дальнейшего использования при оформлении отчёта.

По окончании работы необходимо составить отчёт и подготовиться к его защите на следующем занятии.

Структура и форма отчета о лабораторной работе

Отчёт о лабораторной работе должен содержать:

- фамилию, имя и отчество студента, выполнившего работу;
- номер учебной группы;
- дату выполнения работы;
- название работы;
- цель работы;
- краткую формулировку задания на лабораторную работу;
- основные теоретические сведения и формулы, использовавшиеся в процессе выполнения работы;
- листинги программ, разработанные и отлаженные в процессе выполнения работы;
- схемы, разработанные в процессе выполнения работы;
- при использовании дополнительной литературы указать ссылки и привести список литературы;
- Выводы по работе.

Требования к оформлению отчета о лабораторной работе

Пример оформления титульного листа отчёта по лабораторной работе приведен на сайте университета ww.guap.ru.

При оформлении отчёта о лабораторной работе необходимо придерживаться требований ГОСТ 7.32-2017 «Отчёт о научно-исследовательской работе. Структура и правила оформления».

По каждой лабораторной работе должен быть подготовлен отчёт в бумажном и в электронном виде. После защиты лабораторных работ отчёты в бумажном виде с проставленными оценками хранятся на кафедре, а отчёты в электронной форме должны быть выложены в личном кабинете учащегося на сайте университета.

11.3. Методические указания для обучающихся по прохождению самостоятельной работы

В ходе выполнения самостоятельной работы, обучающийся выполняет работу по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Для обучающихся по заочной форме обучения, самостоятельная работа может включать в себя контрольную работу.

В процессе выполнения самостоятельной работы, у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет им развивать умения и навыки в усвоении и систематизации приобретаемых знаний, обеспечивает высокий уровень успеваемости в период обучения, помогает получить навыки повышения профессионального уровня.

Методическими материалами, направляющими самостоятельную работу обучающихся являются:

- учебно-методический материал по дисциплине;
- методические указания по выполнению контрольных работ (для обучающихся по заочной форме обучения).

11.4. Методические указания для обучающихся по прохождению текущего контроля успеваемости.

Текущий контроль успеваемости предусматривает контроль качества знаний обучающихся, осуществляемого в течение семестра с целью оценивания хода освоения дисциплины.

Контроль качества знаний проводится в форме индивидуального собеседования по материалу отдельных разделов дисциплины, а также проверки отчётов о выполнении практических заданий.

Результаты текущего контроля могут учитываться при проведении промежуточной аттестации.

11.5. Методические указания для обучающихся по прохождению промежуточной аттестации.

Промежуточная аттестация обучающихся предусматривает оценивание промежуточных и окончательных результатов обучения по дисциплине. Она проводится в форме дифференцированного зачёта.

Дифференцированный зачет – это форма оценки знаний, полученных обучающимся при изучении дисциплины, при выполнении курсовых проектов, курсовых работ, научно-исследовательских работ и прохождении практик с аттестационной оценкой «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Система оценок при проведении промежуточной аттестации осуществляется в соответствии с требованиями Положений «О текущем контроле успеваемости и промежуточной аттестации студентов ГУАП, обучающихся по программам высшего образования» и «О модульно-рейтинговой системе оценки качества учебной работы студентов в ГУАП».

Лист внесения изменений в рабочую программу дисциплины

Дата внесения изменений и дополнений. Подпись внесшего изменения	Содержание изменений и дополнений	Дата и № протокола заседания кафедры	Подпись зав. кафедрой