

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ
Заведующий кафедрой
математического анализа
Шабров С.А.
подпись, расшифровка подписи
17.04.2024 г.



РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
Б1.В.ДВ.01.02.04 Машинное обучение на языке Python
Код и наименование дисциплины в соответствии с учебным планом

1. Код и наименование подготовки:

02.03.01 Математика и компьютерные науки

2. Профиль подготовки / специализация :

Математическое и компьютерное моделирование

3. Квалификация выпускника: Бакалавр

4. Форма обучения: Очная

5. Кафедра, отвечающая за реализацию дисциплины: Кафедра математического анализа

6. Составители программы:

(ФИО, ученая степень, ученое звание)

Зверева Маргарита Борисовна, канд. физ-мат. наук, доцент

7. Рекомендована: Научно-методическим Советом математического факультета, протокол от 28.03.2024 № 0500-03

8. Учебный год: 2027-2028

Семестр(ы): 8

9. Цели и задачи учебной дисциплины:

Целями освоения учебной дисциплины являются:

формирование базовых знаний о математическом аппарате машинного обучения;
формирование навыков владения информационными и компьютерными технологиями, необходимыми для решения прикладных задач;
получение знаний о процессах, алгоритмах и инструментах машинного обучения.

Задачи учебной дисциплины:

формирование знаний по основам машинного обучения;
формирование умений использовать математический аппарат для решения практических задач машинного обучения;
выработка умений и навыков использования библиотек языка Python для разработки прикладного программного обеспечения на основе существующих алгоритмов машинного обучения;
выработка умений анализировать полученные результаты;
приобретение навыков работы со специальной литературой.

10. Место учебной дисциплины в структуре ООП:

Дисциплина «Машинное обучение на языке Python» относится к учебным дисциплинам части программы бакалавриата, формируемой участниками образовательных отношений, модуль по выбору.

Дисциплина «Машинное обучение на языке Python» базируется на знаниях, полученных в рамках курсов «Технология программирования и работа на ЭВМ», «Объектно-ориентированное программирование». Приобретенные в результате обучения знания, умения и навыки используются во всех естественнонаучных дисциплинах, модулях и практиках. Полученные знания могут быть использованы при продолжении образования в аспирантуре и в дальнейшей трудовой деятельности выпускников.

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

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
ПК-2	Способен использовать	ПК-2.1	ПК-2.1 Способен	Знать основные определения, понятия и идеи изучаемых разделов курса.

	современные информационные технологии при решении задач математического и компьютерного моделирования		строить алгоритмы и реализовывать их программными методами, в том числе на базе пакетов прикладных программ	Уметь собирать, обрабатывать, анализировать и обобщать результаты, реализовывать алгоритмы программными методами. Владеть методами построения алгоритмов.
		ПК-2.2	ПК-2.2 Способен использовать современные методы математического и компьютерного моделирования при решении теоретических и прикладных задач	Знать математический аппарат машинного обучения. Уметь использовать методы математического и компьютерного моделирования при решении задач . Владеть методами математического и компьютерного моделирования .

12. Объем дисциплины в зачетных единицах/час.(в соответствии с учебным планом) 2/72

Форма промежуточной аттестации(зачет/экзамен) *зачет*

13. Трудоемкость по видам учебной работы

Вид учебной работы		Трудоемкость	
		Всего	По семестрам
			8 семестр
Контактная работа		38	38
в том числе:	лекции	12	12
	практические	26	26
	лабораторные	0	0
Самостоятельная работа		34	34
в том числе: курсовая работа (проект)			
Форма промежуточной аттестации (зачет)			
Итого:		72	72

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью он-лайн курса, ЭУМК
-----	---------------------------------	-------------------------------	---

1. Лекции			
1.1	Основы языка программирования Python	Базовые конструкции языка Python	Курс: Машинное обучение на языке Python (edu.vsu.ru)
1.2	Элементы искусственного интеллекта	Основные понятия и определения искусственного интеллекта. Классификация перцептронов. Функция единичного скачка. Сигмоидная функция активации. Гиперболический тангенс. Однослойные и многослойные нейронные сети. Обучающая и тестовая выборка. Обучение с учителем и без учителя.	
1.3	Решение задачи регрессии	Метод наименьших квадратов. Измерение ошибки в задачах Регрессии. Многомерная регрессия, проблема мультиколлинеарности. Линейная и полиномиальная регрессия. Решение проблемы переобучения.	
1.4	Обзор классификаторов с использованием библиотеки scikit-learn	Линейная модель классификации. Логистическая регрессия как линейный классификатор. Функция потерь (ошибок классификации). Метрики качества классификации.	
1.5	Ансамбли моделей бэггинг, бустинг, градиентный бустинг	Бэггинг, случайный лес как пример бэггинга. Бэггинг линейных классификаторов. Бустинг. Adaboost для ансамбля из простых деревьев. Градиентный бустинг. Градиентный бустинг в задаче регрессии. Градиентный бустинг в задаче классификации. Градиентный бустинг над деревьями.	
1.6	Обработка изображений с помощью библиотеки Open CV	Обученные классификаторы Хаара для распознавания объектов в изображениях.	
2. Практические занятия			
2.1	Основы языка программирования Python	Переменные, функции, массивы, условия и циклы, классы и объекты	Курс: Машинное обучение на языке Python (edu.vsu.ru)
2.2	Элементы искусственного интеллекта	Линейная разделимость объектов. Решение задач классификации на основе логистических функций. Линейная аппроксимация. Дельта-правило. Подбор весов связей.	
2.3	Решение задачи регрессии	Решение примеров построения моделей регрессии	
2.4	Обзор классификаторов с использованием библиотеки scikit-learn	Тренировка перцептрона в scikit-learn. Тренировка логистической регрессионной модели в scikit-learn. Построение дерева решений. Применение метода k-ближайших соседей.	
2.5	Ансамбли моделей бэггинг, бустинг, градиентный бустинг	Сравнение результатов бустинга для слабых и сильных моделей.	
2.6	Обработка изображений с помощью библиотеки Open CV	Программы поиска лиц, автомобильных номеров	

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Лекции	Практические	Лабораторные	Самостоятельная работа	Всего
1	Основы языка программирования Python	2	2	0	4	8
2	Элементы искусственного интеллекта	2	5	0	6	13
3	Решение задачи регрессии	2	4	0	6	12
4	Обзор классификаторов с использованием библиотеки scikit-learn	2	5	0	6	13
5	Ансамбли моделей бэггинг, бустинг, градиентный бустинг	2	5	0	6	13
6	Обработка изображений с помощью библиотеки Open CV	2	5	0	6	13
	Итого:	12	26	0	34	72

14. Методические указания для обучающихся по освоению дисциплины

(рекомендации обучающимся по освоению дисциплины: указание наиболее сложных разделов, работа с конспектами лекций, презентационным материалом, рекомендации по выполнению курсовой работы, по организации самостоятельной работы по дисциплине и др)

Для обеспечения систематической и регулярной работы по изучению дисциплины и успешного прохождения аттестаций студентам рекомендуется придерживаться следующего порядка обучения:

1. Самостоятельно определить объем времени, необходимого для проработки каждой темы.
2. Регулярно изучать каждую тему дисциплины как по конспектам, так и по рекомендованной литературе, видео урокам, используя различные формы индивидуальной работы.
3. Согласовывать с преподавателем виды работы по изучению дисциплины.
4. По завершении отдельных тем передавать выполненные работы (домашние задания) преподавателю.
5. При успешном прохождении рубежных контрольных испытаний студент может претендовать на сокращение программы промежуточной (итоговой) аттестации по дисциплине.

Методические указания для обучающихся при самостоятельной работе.

1. Самостоятельная работа обучающихся направлена на самостоятельное освоение всех тем и вопросов учебной дисциплины, предусмотренных программой. Самостоятельная работа является обязательным видом деятельности для каждого обучающегося, ее

объем по учебному курсу определяется учебным планом. При самостоятельной работе обучающийся взаимодействует с рекомендованными материалами при минимальном участии преподавателя.

2. Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и ресурсами сети Internet является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у обучающихся заинтересованное отношение к конкретной проблеме.
3. Вопросы, которые вызывают у обучающихся затруднения при подготовке, должны быть заранее сформулированы и озвучены во время занятий в аудитории для дополнительного разъяснения преподавателем.
4. Для успешного и плодотворного обеспечения итогов самостоятельной работы разработаны учебно-методические указания к самостоятельной работе студентов над различными разделами дисциплины.
5. Виды самостоятельной работы: конспектирование учебной и научной литературы; проработка учебного материала (по конспектам лекций, учебной и научной литературе); работа в электронной библиотечной системе; работа с информационными справочными системами, выполнение домашних заданий (практических и теоретических); выполнение контрольных работ; подготовка к практическим занятиям; работа с вопросами для самопроверки.
6. Все задания, выполняемые студентами самостоятельно, подлежат последующей проверке преподавателем.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины (список литературы оформляется в соответствии с требованиями ГОСТ и используется общая сквозная нумерация для всех видов источников)

а) основная литература:

№ п/п	Источник
1	Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения [Электронный ресурс] : руководство / С. Рашка ; пер. с англ. Логунова А.В.. — Электрон. дан. — Москва : ДМК Пресс, 2017. — 418 с. — Режим доступа: https://e.lanbook.com/book/100905
2	Коэльо, Л.П. Построение систем машинного обучения на языке Python [Электронный ресурс] / Л.П. Коэльо, В. Ричарт ; пер. с англ. Слинкин А. А.. — Электрон. дан. — Москва : ДМК Пресс, 2016. — 302 с. — Режим доступа: https://e.lanbook.com/book/82818
3	Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Электронный ресурс] / П. Флах. — Электрон. дан. — Москва : ДМК Пресс, 2015. — 400 с. — Режим доступа: https://e.lanbook.com/book/69955

б) дополнительная литература:

№ п/п	Источник
4	Постолит, А. Основы искусственного интеллекта в примерах на Python / А. Постолит.— Санкт-Петербург: БХВ-Петербург, 2024. — 447 с.
5	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с
6	Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. -СПб.: Питер, 2017. -336 с.
7	Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. -СПб.: Питер, 2017. -336 с
8	Шарден, Б. Крупномасштабное машинное обучение вместе с Python [Электронный ресурс] : учебное пособие / Б. Шарден, Л. Массарон, А. Боскетти ; пер. с англ. А. В. Логунова. — Электрон. дан. — Москва : ДМК Пресс, 2018. — 358 с. — Режим доступа: https://e.lanbook.com/book/10583
9	Вьюгин, В.В. Математические основы машинного обучения и прогнозирования [Электронный ресурс] : учебное пособие / В.В. Вьюгин. — Электрон. дан. — Москва : МЦНМО, 2013. — 304 с. — Режим доступа: https://e.lanbook.com/book/56397

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)*:

№ п/п	Ресурс
10	Электронный каталог Научной библиотеки Воронежского государственного университета. – (http // www.lib.vsu.ru/)
11.	Электронный курс Машинное обучение на языке Python (edu.vsu.ru)

* Вначале указываются ЭБС, с которыми имеются договора у ВГУ, затем открытые электронно-образовательные ресурсы, онлайн курсы, ЭУМК.

16. Перечень учебно-методического обеспечения для самостоятельной работы (учебно-методические рекомендации, пособия, задачки, методические указания по выполнению практических (контрольных) работ и др.)

№ п/п	Источник
1	Курс «Машинное обучение» на портале «Электронный университет ВГУ», автор Каширина И.Л. https://edu.vsu.ru/course/view.php?id=3579
2	Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. -СПб.: Питер, 2017. -336 с.: Материалы к книге: https://github.com/brinkar/real-world-machine-learning
3	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с. Материалы к книге: https://github.com/jakevdp/PythonDataScienceHandbook

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

Дисциплина может реализовываться с применением дистанционных образовательных технологий, например, на платформе «Электронный университет ВГУ» (Курс Машинное обучение на языке Python (edu.vsu.ru)).

Перечень необходимого программного обеспечения: Mozilla Firefox (Mozilla Public License (MPL), бесплатное и/или свободное ПО, лицензия: <https://www.mozilla.org/en-US/MPL/>)

18. Материально-техническое обеспечение дисциплины:

Компьютерный класс: специализированная мебель, маркерная доска, персональные компьютеры

Mozilla Firefox (Mozilla Public License (MPL), бесплатное и/или свободное ПО, лицензия: <https://www.mozilla.org/en-US/MPL/>)

19. Оценочные средства для проведения текущей и промежуточной аттестаций

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

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1	Основы языка программирования Python	ПК-2	ПК – 2.1, ПК – 2.2	Промежуточная аттестация – зачет, контрольно-измерительные материалы к зачету, диагностическая работа, лабораторная работа.
2	Элементы искусственного интеллекта	ПК-2	ПК – 2.1, ПК – 2.2	Промежуточная аттестация – зачет, контрольно-измерительные материалы к зачету, диагностическая работа, лабораторная работа.
3	Решение задачи регрессии	ПК-2	ПК – 2.1, ПК – 2.2	Промежуточная аттестация – зачет, контрольно-измерительные материалы к зачету, диагностическая работа, лабораторная работа.
4	Обзор классификаторов с использованием библиотеки scikit-learn	ПК-2	ПК – 2.1, ПК – 2.2	Промежуточная аттестация – зачет, контрольно-измерительные материалы к зачету, диагностическая работа, лабораторная работа.
5	Ансамбли моделей бэггинг, бустинг, градиентный бустинг	ПК-2	ПК – 2.1, ПК – 2.2	Промежуточная аттестация – зачет, контрольно-измерительные материалы к зачету, диагностическая работа, лабораторная работа.
6	Обработка изображений с помощью библиотеки Open CV	ПК-2	ПК – 2.1, ПК – 2.2	Промежуточная аттестация – зачет, контрольно-измерительные материалы к зачету, диагностическая работа, лабораторная работа.
Промежуточная аттестация Форма контроля - зачет				Перечень вопросов к зачету.

20. Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1 Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств: лабораторная работа

Примерный комплект заданий для лабораторных работ

Вариант 1.

- 1) Разбейте предоставленный Вам преподавателем набор данных на обучающую и тестовую части в соотношении 8:2.
- 2) Обучите, а затем провалидируйте на тестовых данных следующие модели, используя в качестве метрики качества R^2 , предварительно отмасштабировав данные: LinearRegression; Lasso с коэффициентом регуляризации, равным 0.01.
- 3) Проанализируйте качество получившихся моделей и сравните количество строго нулевых весов в них.

Составитель _____  _____ М.Б. Зверева
(подпись)

Вариант 2.

- 1) Разбейте предоставленный Вам преподавателем набор данных на обучающую и тестовую части в соотношении 8:2.
- 2) Проведите предобработку данных: заполнение пропусков, кодирование, масштабирование
- 3) Обучите, а затем провалидируйте на тестовых данных модель логистической регрессии
- 4) Вычислите значения метрик: recall, precision, F1-мера, AUC-ROC. Постройте ROC-кривую.

Составитель _____  _____ М.Б. Зверева
(подпись)

Текущий контроль представляет собой проверку усвоения учебного материала теоретического и практического характера, регулярно осуществляемую на занятиях.

К основным формам текущего контроля можно отнести устный опрос, проверку домашних заданий, лабораторной работы.

Задание для текущего контроля и проведения промежуточной аттестации должны быть направлены *на оценивание*:

1. уровня освоения теоретических и практических понятий, научных основ профессиональной деятельности;
2. степени готовности обучающегося применять теоретические и практические знания и профессионально значимую информацию, сформированности когнитивных умений.
3. приобретенных умений, профессионально значимых для профессиональной деятельности.

Текущий контроль предназначен для проверки хода и качества формирования компетенций, стимулирования учебной работы обучаемых и совершенствования методики освоения новых знаний. Он обеспечивается проведением лабораторных работ и домашних работ, проверкой конспектов лекций, периодическим опросом слушателей на занятиях.

Формы, методы и периодичность текущего контроля определяет преподаватель. При текущем контроле уровень освоения учебной дисциплины и степень сформированности компетенции определяются оценками «зачтено» и «не зачтено».

Описание технологии проведения

Тестирование проводится письменно, лабораторные работы выполняются на компьютерах.

Требование к выполнению заданий

Лабораторная работа

За лабораторную работу ставится оценка «зачтено», в случае, если обучающийся выполнил:

- правильно в полном объеме все задания лабораторной работы, показал отличные владения навыками применения полученных знаний и умений при решении профессиональных задач в рамках усвоенного материала;
- обучающийся выполнил все задания с небольшими неточностями и показал хорошие владения навыками применения полученных знаний и умений при решении профессиональных задач в рамках усвоенного материала;
- обучающийся выполнил половину из предложенных заданий правильно, остальные с существенными неточностями и показал удовлетворительное владение навыками полученных знаний и умений при решении профессиональных задач в рамках усвоенного материала.

В остальных случаях обучающемуся ставится за лабораторную работу «не зачтено».

20.2 Промежуточная аттестация

Промежуточная аттестация предназначена для определения уровня освоения всего объема учебной дисциплины. Промежуточная аттестация по дисциплине «Машинное обучение на языке Python» проводится в форме зачета.

Промежуточная аттестация, как правило, осуществляется в конце семестра и завершает изучение дисциплины. Промежуточная аттестация помогает оценить более крупные совокупности знаний и умений, в некоторых случаях – даже формирование определенных профессиональных компетенций. На зачете оценивается уровень освоения учебной дисциплины и степень сформированности компетенции определяются оценками «зачтено», «не зачтено».

Описание технологии проведения

На зачете студент вытягивает билет, который содержит один теоретический вопрос и один практический, выполняемый на компьютере. Все вопросы и задачи, входящие в билеты, охватывают весь материал, изучаемый за весь семестр.

Перечень вопросов к зачету:

1. Основные понятия машинного обучения. Основные постановки задач. Примеры прикладных задач.
2. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
3. Метрики качества алгоритмов регрессии и классификации.
4. Линейная регрессия. Простая многомерная регрессия. Регрессия с

полиномиальными признаками. Методы регуляризации: Ridge, Lasso, ElasticNet.

5. Логистическая регрессия.

6. Деревья решений. Методы построения деревьев. Их регуляризация.

7. Композиции алгоритмов. Разложение ошибки на смещение и разброс.

8. Случайный лес, его особенности.

9. Градиентный бустинг, его особенности при использовании деревьев в качестве базовых алгоритмов.

10. Обработка изображений с помощью библиотеки Open CV

Примерный комплект билетов для зачета

Контрольно-измерительный материал № 1

1. Линейная регрессия

2. Считайте выборку из файла и ответьте на следующие вопросы:

Имеются ли в данных пропущенные значения? Сколько всего пропущенных элементов в таблице "объект-признак"? Сколько объектов имеют хотя бы один пропуск? Сколько признаков имеют хотя бы одно пропущенное значение? Приведите данные к виду, пригодному для обучения линейных моделей. Разбейте выборку и вектор целевой переменной на обучение и контроль в отношении 70/30. Обучите линейную регрессию на 1000 объектах из обучающей выборки и выведите значения MSE и R^2 и на этой подвыборке и контрольной выборке. Проинтерпретируйте полученный результат.

Составитель _____



(подпись)

_____ М.Б. Зверева

Контрольно-измерительный материал № 2

1. Обработка изображений с помощью библиотеки Open CV.

2. Загрузите изображение и напишите программу для поиска лиц на фотографии, сформировав прямоугольник вокруг каждого обнаруженного лица.

Составитель _____



(подпись)

_____ М.Б. Зверева

Критерии выставления оценок:

Оценки	Критерии
Зачтено	обучающийся показывает свой интеллектуальный и общекультурный уровень, знает предмет учебной дисциплины, логично излагает изученный материал, умеет применять теоретические знания для решения практических заданий.
Не зачтено	обучающийся демонстрирует фрагментарные знания и умения или отсутствие их.

20.3 Фонд оценочных средств сформированности компетенций студентов, рекомендуемый для проведения диагностических работ

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

Задание 1

Какова максимальная точность, которую можно достичь с помощью метода k ближайших соседей?

- А 50%
- Б 90%
- В 70%
- Г Точность не ограничена

Задание 2

Что такое евклидово расстояние в методе k ближайших соседей?

- А Математическое ожидание расстояния
- Б Мера сходства между двумя примерами
- В Расстояние между двумя линиями
- Г Расстояние между точками на карте

Задание 3

Каким образом можно использовать метод k ближайших соседей для задачи классификации?

- А Выбрать k на основе максимального расстояния до k ближайших соседей и отнести объект к классу, который наиболее часто встречается среди k ближайших соседей.
- Б Использовать метод k ближайших соседей только для задач регрессии.
- В Нельзя использовать метод k ближайших соседей для задачи классификации.
- Г Выбрать k на основе минимального расстояния до k ближайших соседей и отнести объект к классу, который наиболее часто встречается среди k ближайших соседей.

Задание 4

Каким образом выбираются веса для взвешенного метода k ближайших соседей?

- А Задается вручную
- Б Используется расстояние до каждого ближайшего соседа и их порядковый номер
- В Используется расстояние до каждого ближайшего соседа
- Г Используется расстояние до каждого ближайшего соседа и их классы

Задание 5

Что представляет собой метод k ближайших соседей?

- А Метод градиентного спуска
- Б Метод вычисления статистических показателей
- В Метод оптимизации функций
- Г Метод машинного обучения для классификации и регрессии данных

Задание 6

Как предпочтительно выбирать значение k в методе k ближайших соседей?

- А Значение k выбирается случайным образом
- Б Выбор k не имеет значения
- В Значение k выбирается на основе кросс-валидации
- Г Значение k выбирается на основе экспертного мнения

Задание 7

Какой алгоритм используется для поиска k ближайших соседей?

- А Алгоритм Беллмана-Форда
- Б Алгоритм поиска ближайших соседей
- В Алгоритм Дейкстры
- Г Жадный алгоритм

Задание 8

Каким образом можно уменьшить влияние выбросов непосредственно при обучении модели машинного обучения методом k ближайших соседей?

- А Использовать алгоритмы кластеризации для определения выбросов и исключения их из обучающего набора
- Б Уменьшить значение k
- В Использовать метрики расстояния, устойчивые к выбросам, например, манхэттенское расстояние
- Г Использовать взвешенный метод k ближайших соседей, где вес каждого ближайшего соседа зависит от расстояния до нового примера

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

Задание 1

Пусть задана зависимость между выработкой продукции на одного работника и удельного веса рабочих высокой квалификации:

x	10	12	15	17	18	19	19	20	20	21
y	6	6	7	7	7	8	8	9	9	10

Выписать коэффициент возле x в уравнении парной линейной регрессии y от x , оставляя в коэффициентах 4 знака после запятой

Задание 2

Имеются следующие данные разных стран об индексе розничных цен на продукты питания x и индексе промышленного производства y :

x	100	105	108	113	118	118	110	115	119	118	120	124	129	132
y	70	79	85	84	85	85	96	99	100	98	99	102	105	112

В какой степени войдет x в уравнение степенной регрессии y от x ? Оставить 4 знака после запятой.

Задание 3

Установите соответствие между видами нелинейных регрессий

А степенная; Б экспоненциальная; В гиперболическая; Г обратная;

Д квадратичная; Е логарифмическая.

1. $\hat{y} = a + b_1 x + b_2 x^2$

2. $\hat{y} = \frac{1}{a + bx}$

3. $\hat{y} = ax^b$

1. $\hat{y} = a + b \ln x$

2. $\hat{y} = ae^{bx}$

3. $\hat{y} = a + \frac{b}{x}$

Задание 4

В таблице отражена зависимость расходов на питание y (тыс. руб.) от личного дохода населения x (тыс. руб.).

№	1	2	3	4	5	6	7
---	---	---	---	---	---	---	---

y	12	14	17	19	20	26	28
x	25	31	35	40	42	51	64

Оставив 2 знака после запятой, найти остаточную дисперсию

Задание 5

В таблице отражена зависимость расходов на питание y (тыс. руб.) от личного дохода населения x (тыс. руб.).

№	1	2	3	4	5	6	7
y	12	14	17	19	20	26	28
x	25	31	35	40	42	51	64

Найти значение F–статистики по модели линейной регрессии, оставив 2 знака после запятой

Задание 6

В таблице отражена зависимость расходов на питание y (тыс. руб.) от личного дохода населения x (тыс. руб.).

№	1	2	3	4	5	6	7
y	12	14	17	19	20	26	28
x	25	31	35	40	42	51	64

Найти значение коэффициента корреляции R , оставив 2 знака после запятой

Задание 7

Установите соответствие между мерой качества модели и ее характеристикой.

1. Критерий Фишера.
2. Коэффициент корреляции
3. t-критерий Стьюдента
4. Коэффициент детерминации

А. Характеризует долю дисперсии (вариации) исследуемого процесса, объясняемую регрессией (влиянием фактора), в общей дисперсии переменной.

Б. Используется для оценки значимости и адекватности модели в целом.

В. Используется для оценки степени тесноты связи между данными

Г. Используется для оценки значимости каждого параметра построенной модели.

Задание 8

Установите соответствие между понятиями и определениями.

1. Несмещенная оценка параметра.
2. Состоятельная оценка параметра.
3. Эффективная оценка параметра.

А. Математическое ожидание равно оцениваемому параметру.

Б. Оценка имеет наименьшую дисперсию в классе всех несмещенных оценок, вычисленных по выборкам одного и того же объема n .

В. При возрастании количества наблюдений оценка сходится по вероятности к оцениваемому параметру.

Задание 9

В таблице представлены расходы на питание y (тыс. руб.) в зависимости от личного дохода x (тыс. руб.).

№	1	2	3	4	5	6	7
y	12	14	17	19	20	26	28
x	25	31	35	40	42	51	64

Построить модель логарифмической регрессии, выписав в ответ коэффициент возле $\ln x$ с одним знаком после запятой.

Задание 10

В таблице представлены расходы на питание y (тыс. руб.) в зависимости от личного дохода x (тыс. руб.).

№	1	2	3	4	5	6	7
y	12	14	17	19	20	26	28
x	25	31	35	40	42	51	64

Выписать факторную дисперсию по логарифмической модели с 2 знаками после запятой.

Задание 11

В таблице представлены расходы на питание y (тыс. руб.) в зависимости от личного дохода x (тыс. руб.).

№	1	2	3	4	5	6	7
---	---	---	---	---	---	---	---

y	12	14	17	19	20	26	28
x	25	31	35	40	42	51	64

Выписать остаточную дисперсию по логарифмической модели с двумя знаками после запятой.

Задание 12

В таблице представлены расходы на питание y (тыс. руб.) в зависимости от личного дохода x (тыс. руб.).

№	1	2	3	4	5	6	7
y	12	14	17	19	20	26	28
x	25	31	35	40	42	51	64

Найти значение F–статистики по модели логарифмической регрессии, оставив 2 знака после запятой

Критерии и шкалы оценивания заданий ФОС:

1) Задания закрытого типа (выбор одного варианта ответа, верно/неверно):

- 1 балл – указан верный ответ;
- 0 баллов – указан неверный ответ.

2) Задания закрытого типа (множественный выбор):

- 2 балла – указаны все верные ответы;
- 0 баллов — указан хотя бы один неверный ответ.

3) Задания закрытого типа (на соответствие):

- 2 балла – все соответствия определены верно;
- 0 баллов – хотя бы одно сопоставление определено неверно.

4) Задания открытого типа (короткий текст):

- 2 балла – указан верный ответ;
- 0 баллов – указан неверный ответ.

5) Задания открытого типа (число):

- 2 балла – указан верный ответ;
- 0 баллов – указан неверный ответ.

Задания раздела 20.3 рекомендуются к использованию при проведении диагностических работ с целью оценки остаточных результатов освоения данной дисциплины (знаний, умений, навыков).