

**Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Национальный исследовательский университет «МЭИ»**

Направление подготовки/специальность: 01.04.02 Прикладная математика и информатика

Наименование образовательной программы: Искусственный интеллект

Уровень образования: высшее образование - магистратура

Форма обучения: Очная

**Оценочные материалы
по дисциплине
Технологии обработки больших данных**

**Москва
2022**

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ РАЗРАБОТАЛ:

Преподаватель

(должность)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Гданский Н.И.
	Идентификатор	Rb38b5a94-GdanskyNI-d2219177

(подпись)

Н.И.

Гданский

(расшифровка
подписи)

СОГЛАСОВАНО:

Руководитель
образовательной
программы

(должность, ученая степень,
ученое звание)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Маран М.М.
	Идентификатор	R7be141f2-MaranMM-804b01e2

(подпись)

М.М. Маран

(расшифровка
подписи)

Заведующий
выпускающей кафедры

(должность, ученая степень,
ученое звание)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Варшавский П.Р.
	Идентификатор	R9a563c96-VarshavskyPR-efb4bbd

(подпись)

П.Р.

Варшавский

(расшифровка
подписи)

ОБЩАЯ ЧАСТЬ

Оценочные материалы по дисциплине предназначены для оценки: достижения обучающимися запланированных результатов обучения по дисциплине, этапа формирования запланированных компетенций и уровня освоения дисциплины.

Оценочные материалы по дисциплине включают оценочные средства для проведения мероприятий текущего контроля успеваемости и промежуточной аттестации.

Формируемые у обучающегося компетенции:

1. ПК-3 Способен осуществлять оптимальный выбор и освоение программно-технической среды реализации программного обеспечения, выполнять разработки в ней
ИД-2 Демонстрирует умение выбирать и обосновывать выбор программно-технической среды реализации проектов по информационным технологиям
2. ПК-4 Способен выполнять все этапы жизненного цикла информационных систем в выбранной программно-технической среде
ИД-2 Демонстрирует умение выбора среды и реализовать ПО
3. ПК-8 Способен применять методы и технологии искусственного интеллекта для разработки интеллектуальных систем
ИД-2 Выбирает методы и средства обработки больших данных

и включает:

для текущего контроля успеваемости:

Форма реализации: Компьютерное задание

1. Деревья решений, регрессионные деревья, деревья классификации (Контрольная работа)

Форма реализации: Письменная работа

1. Обучение без учителя. Методы кластеризации (Контрольная работа)
2. Основные виды регрессий. Классификация регрессий (Контрольная работа)
3. Основные проблемы обработки больших данных (Контрольная работа)

БРС дисциплины

3 семестр

Раздел дисциплины	Веса контрольных мероприятий, %				
	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4
	Срок КМ:	4	8	12	16
Основные проблемы обработки больших данных					
Основные проблемы обработки больших данных	+				
Основные виды регрессий. Классификация регрессий					
Основные виды регрессий. Классификация регрессий			+		
Деревья решений, регрессионные деревья, деревья классификации					

Деревья решений, регрессионные деревья, деревья классификации			+	
Обучение без учителя. Методы кластеризации				
Обучение без учителя. Методы кластеризации				+
Вес КМ:	25	25	25	25

\$Общая часть/Для промежуточной аттестации\$

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ТЕКУЩЕГО КОНТРОЛЯ

I. Оценочные средства для оценки запланированных результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Индекс компетенции	Индикатор	Запланированные результаты обучения по дисциплине	Контрольная точка
ПК-3	ИД-2ПК-3 Демонстрирует умение выбирать и обосновывать выбор программно-технической среды реализации проектов по информационным технологиям	Знать: характеристики основных программно-технических сред для разработки проектов по информационным технологиям и методы работы в программно-технических средах для разработки проектов по информационным технологиям Уметь: осуществлять обоснованный выбор программно-технических сред для разработки проектов по информационным технологиям и практически применять программно-технические среды для разработки проектов по информационным	Основные проблемы обработки больших данных (Контрольная работа)

		технологиям	
ПК-4	ИД-2 _{ПК-4} Демонстрирует умение выбора среды и реализовать ПО	<p>Знать:</p> <p>основные методы контроля ПО</p> <p>основные этапы жизненного цикла информационных систем в выбранной программно-технической среде и методы реализации разработки ПО в выбранной программно-технической среде</p> <p>Уметь:</p> <p>применять методы контроля ПО на практике</p> <p>разрабатывать ПО в выбранной программно-технической среде</p>	<p>Основные виды регрессий. Классификация регрессий (Контрольная работа)</p> <p>Деревья решений, регрессионные деревья, деревья классификации (Контрольная работа)</p>
ПК-8	ИД-2 _{ПК-8} Выбирает методы и средства обработки больших данных	<p>Знать:</p> <p>методы и средства обработки больших данных</p> <p>Уметь:</p> <p>применять для решения практических задач</p> <p>методы и средства обработки больших данных</p>	<p>Обучение без учителя. Методы кластеризации (Контрольная работа)</p>

II. Содержание оценочных средств. Шкала и критерии оценивания

КМ-1. Основные проблемы обработки больших данных

Формы реализации: Письменная работа

Тип контрольного мероприятия: Контрольная работа

Вес контрольного мероприятия в БРС: 25

Процедура проведения контрольного мероприятия: Тестирование проводится на занятии

Краткое содержание задания:

Для заданных значений результата Y , соответствующих значениям регрессора $X=1,2,3,4,5$, требуется построить (найти формулы) следующих видов регрессий:

- линейной,
- степенной,
- показательной и
- равносторонней гиперболы.

Найти для них [среднюю ошибку аппроксимации](#) по каждой регрессии и выбрать наилучшую из них.

В.1. (2.3; 4,5; 9,1; 14.7; 20.5)

В.2. (2.8; 3,5; 4,3; 5.0; 5.3)

В.3. (20.3; 14,1; 9,7; 4.8; 2.5)

В.4. (22.8; 17,5; 11,3; 5.2; 2.0)

В.5. (12.1; 15,9; 19,7; 29.3; 40.2)

В.6. (23.5; 19,2; 14,3; 8.1; 0.4)

В.7. (44.8; 34,9; 29,7; 14.3; 2.9)

В.8. (33.6; 31,1; 26,5; 18.4; 8.2)

В.9. (12.7; 14,3; 19,2; 24.9; 32.2)

В.10. (13.8; 17,5; 24,3; 35.; 45.5)

В.11. (23.3; 34,5; 49,1; 54.7; 70.5)

В.12. (0.6; 3,7; 8,3; 15.0; 21.1)

В.13. (33.1; 44,1; 59,4; 64.8; 72.2)

В.14. (14.4; 11,1; 8,3; 5.8; 2.2)

В.15. (13.1; 17,9; 22,7; 29.3; 37.1)

В.16. (2.9; 25,5; 29,4; 34.1; 40.5)

- B.17.** (22.0; 23,8; 24,3; 25.2; 25.9)
- B.18.** (30.1; 24,1; 19,7; 14.5; 12.2)
- B.19.** (32.7; 27,2; 21,2; 15.4; 12.1)
- B.20.** (22.2; 35,3; 39,9; 49.2; 60.5)
- B.21.** (3.4; 9,4; 14,7; 19.1; 27.4)
- B.22.** (64.6; 54,8; 49,4; 44.1; 42.8)
- B.23.** (53.5; 51,8; 46,4; 38.2; 28.1)
- B.24.** (22.4; 25,6; 29,1; 34.3; 42.7)
- B.25.** (23.9; 27,2; 34,4; 45.0; 55.8)
- B.26.** (3.4; 4,7; 9,8; 14.1; 20.9)
- B.27.** (1.7; 4,6; 9,4; 16.6; 22.8)
- B.28.** (13.3; 24,8; 39,3; 54.8; 72.7)
- B.29.** (34.9; 31,0; 28,2; 25.2; 22.5)
- B.30.** (33.7; 37,7; 42,4; 49.8; 57.4)

Контрольные вопросы/задания:

<p>Знать: характеристики основных программно-технические сред для разработки проектов по информационным технологиям и методы работы в программно-технических средах для разработки проектов по информационным технологиям</p>	<ol style="list-style-type: none"> 1. Что представляет собой статистический вывод ? 2. В чем заключаются основные задачи регрессионного анализа ? 3. В чем заключается разнородность больших данных 4. В чем заключается накопление шума в больших данных ? · 5. В чем заключается преимущество использования разреженных моделей при анализе больших данных ? · 6. В чем заключается ложная коррекция при анализе больших данных ? 7. В чем заключается проблема случайной эндогенности при анализе больших данных ?
<p>Уметь: осуществлять обоснованный выбор программно-технических сред</p>	<ol style="list-style-type: none"> 1. Выполнить постановку задачи регрессионного анализа по заданным данным. 2. Выполнить оценку шума в заданных данных.

для разработки проектов по информационным технологиям и практически применять программно-технические среды для разработки проектов по информационным технологиям	3.Оценить разнородность заданных данных
--	---

Описание шкалы оценивания:

Оценка: 5

Нижний порог выполнения задания в процентах: 90

Описание характеристики выполнения знания: Все задания выполнены правильно

Оценка: 4

Нижний порог выполнения задания в процентах: 80

Описание характеристики выполнения знания: Задания выполнены правильно Есть незначительные погрешности

Оценка: 3

Нижний порог выполнения задания в процентах: 70

Описание характеристики выполнения знания: Есть существенные ошибки

КМ-2. Основные виды регрессий. Классификация регрессий

Формы реализации: Письменная работа

Тип контрольного мероприятия: Контрольная работа

Вес контрольного мероприятия в БРС: 25

Процедура проведения контрольного мероприятия: Тестирование производится на занятии

Краткое содержание задания:

Ответить на теоретический вопрос

Контрольные вопросы/задания:

Знать: основные методы контроля ПО	<p>1.1. Как производится линейная регрессия при анализе больших данных ?</p> <p>2. Как производится логистическая регрессия при анализе больших данных ?</p> <p>3. Как производится гребневая регрессия при анализе больших данных ?</p> <p>4.Как производится лассо-регрессия регрессия при анализе больших данных ?</p> <p>5. Как производится экологическая регрессия при анализе больших данных ?</p> <p>6.Как производится байесовская регрессия регрессия при анализе больших данных ?</p> <p>7. Как производится логическая регрессия регрессия при анализе больших данных ?</p>
------------------------------------	---

	<p>8. Как производится квантильная регрессия при анализе больших данных ?</p> <p>9. Как производится LAD-регрессия при анализе больших данных ?</p> <p>10.. Как производится джекнайф-регрессия при анализе больших данных ?</p> <p>11. Как представляет собой дискриминантный анализ больших данных ?</p> <p>12. Укажите основные команды языка статистических вычислений R.</p>
Уметь: применять методы контроля ПО на практике	<p>1. Построить линейную регрессию заданных данных</p> <p>2. Построить логистическую регрессию заданных данных</p> <p>3. Построить гребневую регрессию заданных данных</p>

Описание шкалы оценивания:

Оценка: 5

Нижний порог выполнения задания в процентах: 90

Описание характеристики выполнения знания: Дан полный правильный ответ

Оценка: 4

Нижний порог выполнения задания в процентах: 80

Описание характеристики выполнения знания: Ответ неполный или есть погрешности

Оценка: 3

Нижний порог выполнения задания в процентах: 70

Описание характеристики выполнения знания: В ответе есть существенные ошибки

КМ-3. Деревья решений, регрессионные деревья, деревья классификации

Формы реализации: Компьютерное задание

Тип контрольного мероприятия: Контрольная работа

Вес контрольного мероприятия в БРС: 25

Процедура проведения контрольного мероприятия: Выполняется во внеурочное время

Краткое содержание задания:

Написать программу, которая реализует требуемый метод обработки данных

Контрольные вопросы/задания:

Знать: основные этапы жизненного цикла информационных систем в выбранной программно-технической среде и методы реализации разработки ПО в выбранной программно-технической среде	<p>1. Как производится моделирование и статистическая обработка больших данных с использованием лассо-регрессии. ?</p> <p>2. Как производится моделирование и статистическая обработка больших данных с использованием экологической регрессии. ?</p> <p>3. Как производится моделирование и статистическая обработка больших данных с использованием Байесовской регрессии ?</p>
--	---

	<p>4. Как производится моделирование и статистическая обработка больших данных с использованием квантильной регрессии. ?</p> <p>5. Как производится моделирование и статистическая обработка больших данных с использованием LAD-регрессии ?</p>
<p>Уметь: разрабатывать ПО в выбранной программно-технической среде</p>	<ol style="list-style-type: none"> 1. Моделирование и статистическая обработка больших данных с использованием линейной регрессии. 2. Моделирование и статистическая обработка больших данных с использованием логистической регрессии. 3. Моделирование и статистическая обработка больших данных с использованием гребневой регрессии. 4. Моделирование и статистическая обработка больших данных с использованием лассо-регрессии. 5.. Моделирование и статистическая обработка больших данных с использованием экологической регрессии. 6. Моделирование и статистическая обработка больших данных с использованием Байесовской регрессии. 7.. Моделирование и статистическая обработка больших данных с использованием квантильной регрессии. 8.. Моделирование и статистическая обработка больших данных с использованием LAD-регрессии. 9. Моделирование и статистическая обработка больших данных с использованием Джекнайф-регрессии. 10. Моделирование и статистическая обработка больших данных с использованием деревьев на основе рекурсивного разбиения. 11. Моделирование и статистическая обработка больших данных с применением деревьев с использованием алгоритма условного вывода. 12.. Моделирование процедуры обучения с использованием графовых методов кластеризации. 13. Моделирование процедуры обучения с

	<p>использованием иерархической кластеризации (таксономии).</p> <p>14. Моделирование процедуры обучения с использованием статистических методов кластеризации.</p>
--	--

Описание шкалы оценивания:

Оценка: зачтено

Описание характеристики выполнения знания: Задание выполнено в полном объеме без существенных ошибок

Оценка: не зачтено

Описание характеристики выполнения знания: Задание выполнено частично, есть серьезные ошибки

КМ-4. Обучение без учителя. Методы кластеризации

Формы реализации: Письменная работа

Тип контрольного мероприятия: Контрольная работа

Вес контрольного мероприятия в БРС: 25

Процедура проведения контрольного мероприятия: после получения задания студент выполняет его в течение 60 мин и сдает на проверку преподавателю в письменном виде

Краткое содержание задания:

Выполнить кластеризацию заданных данных

Контрольные вопросы/задания:

Знать: методы и средства обработки больших данных	<ol style="list-style-type: none"> 1. В чем заключается процедура обучения ,без учителя ? 2. Укажите основные методы управления набором данных 3. В чем заключается процедура отсечения ребер ? 4. В чем заключается процедура гонгломерации и усиление ? 5. В чем заключается графовые методы кластеризации ? 6. В чем заключается иерархическая кластеризация ? 7. В чем заключаются статистические методы кластеризации ?
Уметь: применять для решения практических задач методы и средства обработки больших данных	<ol style="list-style-type: none"> 1. Применить процедуру отсечения ребер к заданным данным 2. Применить процедуру гонгломерации и усиление к заданным данным 3. Применить графовые методы кластеризации к заданным данным 4. Применить статистические методы кластеризации к заданным данным

Описание шкалы оценивания:

Оценка: зачтено

Описание характеристики выполнения знания:

Оценка: не зачтено

Описание характеристики выполнения знания:

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

3 семестр

Форма промежуточной аттестации: Экзамен

Пример билета

1. Основная цель регрессионного анализа. Основные задачи **регрессионного анализа**.
2. Оценка тесноты связи изучаемых явлений для линейных и нелинейных регрессий.

Процедура проведения

Вопросы выдаются студентам на 60 мин. После чего ответы в письменном виде сдаются преподавателю

1. Перечень компетенций/индикаторов и контрольных вопросов проверки результатов освоения дисциплины

1. Компетенция/Индикатор: ИД-2_{ПК-3} Демонстрирует умение выбирать и обосновывать выбор программно-технической среды реализации проектов по информационным технологиям

Вопросы, задания

- 1.
1. Основная цель регрессионного анализа. Основные задачи **регрессионного анализа**.
- 2.2. Виды регрессий по виду функциональных зависимостей.
- 3.3. Нелинейные регрессии.

Материалы для проверки остаточных знаний

1. Что называют большими данными ?

Ответы:

1. Структурированные и [неструктурированные данные](#) огромных объёмов и значительного многообразия, которые эффективно обрабатываются [горизонтально масштабируемыми программными](#) инструментами
2. Структурированные [данные](#) огромных объёмов, которые эффективно обрабатываются [программными](#) инструментами
3. [Неструктурированные данные](#) значительного многообразия, которые могут быть эффективно обработаны [программными](#) инструментами

Верный ответ: 1.

2. Компетенция/Индикатор: ИД-2_{ПК-4} Демонстрирует умение выбора среды и реализовать ПО

Вопросы, задания

- 1.4. Основной метод построения регрессий. Задача оптимизации, решаемая при использовании метода наименьших квадратов.
- 2.5. Методика построения парных линейных регрессий при использовании метода наименьших квадратов.
- 3.6. Оценка тесноты связи изучаемых явлений для линейных и нелинейных регрессий.
- 4.7. Оценка [средней ошибки аппроксимации](#) в регрессии..

Материалы для проверки остаточных знаний

1. Что называют регрессией ?

Ответы:

1. Многосторонняя стохастическая зависимость случайной зависимой переменной от независимых случайных переменных.
2. Односторонняя стохастическая зависимость, устанавливающая соответствие между случайной зависимой переменной и независимыми переменными при условии, что это выражение будет иметь статистическую значимость.
3. Математическая зависимость случайной зависимой переменной от независимых случайных переменных, определяющих ее.

Верный ответ: 2.

3. Компетенция/Индикатор: ИД-2ПК-8 Выбирает методы и средства обработки больших данных

Вопросы, задания

- 1.8. Коэффициент эластичности в регрессии.
- 2.9. Коэффициент (индекс) детерминации признака в регрессии.
- 3.10. Проверка - гипотезы Но о случайной природе оцениваемых характеристик .
- 4.11. Оценка статистической значимости коэффициентов регрессии и корреляции.

Материалы для проверки остаточных знаний

1. Какой способ [машинного обучения](#) задает обучение без учителя ?

Ответы:

1. [Обучение](#) испытуемой системы путем предъявления ей заранее подготовленного набора задач
2. [Обучение](#) испытуемой системы путем предъявления ей спонтанно генерируемого без вмешательства со стороны экспериментатора набора задач
3. [Обучение](#), при котором испытуемая система спонтанно обучается выполнять поставленную задачу без вмешательства со стороны экспериментатора.

Верный ответ: 3.

II. Описание шкалы оценивания

Оценка: 5

Нижний порог выполнения задания в процентах: 90

Описание характеристики выполнения знания: Работа выполнена в рамках "продвинутого" уровня. Ответы даны верно, четко сформулированные особенности практических решений

Оценка: 4

Нижний порог выполнения задания в процентах: 80

Описание характеристики выполнения знания: Работа выполнена в рамках "базового" уровня. Большинство ответов даны верно. В части материала есть незначительные недостатки

Оценка: 3

Нижний порог выполнения задания в процентах: 60

Описание характеристики выполнения знания: Работа выполнена в рамках "порогового" уровня. Основная часть задания выполнена верно. на вопросы углубленного уровня

III. Правила выставления итоговой оценки по курсу

Оценка определяется в соответствии с Положением о балльно-рейтинговой системе для студентов НИУ «МЭИ» на основании семестровой и экзаменационной составляющих.