

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Национальный исследовательский ядерный университет «МИФИ»
Обнинский институт атомной энергетики –
филиал федерального государственного автономного образовательного учреждения высшего образования
«Национальный исследовательский ядерный университет «МИФИ»
(ИАТЭ НИЯУ МИФИ)

Утверждено на заседании
УМС ИАТЭ НИЯУ МИФИ
Протокол №2-8/2024 От 30.08.2024

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Математика для анализа данных

Шифр, название дисциплины

01.04.02 «Прикладная математика и информатика»

Шифр, название специальности/направления подготовки

Математическое моделирование и прикладной анализ данных

Название программы магистратуры

магистр

(Квалификация (степень) выпускника)

Форма обучения: очная

г. Обнинск 2024 г.

Программа составлена в соответствии с требованиями образовательного стандарта высшего образования национального исследовательского ядерного университета «МИФИ» по направлению подготовки 01.04.02 – Прикладная математика и информатика. (квалификация (степень) магистр).

Программу составил:

_____ С.В. Ермаков, доцент, к.ф.-м.н, доцент

Рецензент:

_____ Г.Е. Деев, доцент, к.ф.-м.н, доцент

Программа рассмотрена на заседании ОИКС

(протокол № 5/7 от «30» июля от 2024 г.)

Руководитель направления подготовки 01.04.02
«Прикладная математика и информатика»

_____ Ермаков С.В.

« ____ » _____ 2024 г.

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения ООП магистратуры обучающийся должен овладеть следующими результатами обучения по дисциплине:

Коды компетенций	Результаты освоения ООП Содержание компетенций*	Перечень планируемых результатов обучения по дисциплине**
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	З-УК-1 Знать: методы системного и критического анализа; методики разработки стратегии действий для выявления и решения проблемной ситуации У-УК-1 Уметь: применять методы системного подхода и критического анализа проблемных ситуаций; разрабатывать стратегию действий, принимать конкретные решения для ее реализации В-УК-1 Владеть: методологией системного и критического анализа проблемных ситуаций; методиками постановки цели, определения способов ее достижения, разработки стратегий действий.
УКЦ-1	Способен решать исследовательские, научно-технические и производственные задачи в условиях неопределенности, в том числе выстраивать деловую коммуникацию и организовывать работу команды с использованием цифровых ресурсов и технологий в цифровой среде	З-УКЦ-1 Знать современные цифровые технологии, используемые для выстраивания деловой коммуникации и организации индивидуальной и командной работы У-УКЦ-1 Уметь подбирать наиболее релевантные цифровые решения для достижения поставленных целей и задач, в том числе в условиях неопределенности В-УКЦ-1 Владеть навыками решения исследовательских, научно-технических и производственных задач с использованием цифровых технологий
ОПК-1	Способен решать актуальные задачи фундаментальной и прикладной математики	З-ОПК-1 Знать актуальные задачи фундаментальной и прикладной математики, методы математического моделирования. У-ОПК-1 Уметь использовать методы математического моделирования для решения задач фундаментальной и прикладной математики. В-ОПК-1 Владеть методами математического моделирования и основами их использования.

2. Место дисциплины в структуре ООП магистратуры

Дисциплина «Математика для анализа данных» входит в учебный план подготовки по направлению **01.04.02. «Прикладная математика и информатика»** специализация **«Математическое моделирование и прикладной анализ данных»** и относится к дисциплинам общенаучного модуля.

Дисциплина изучается на 1 курсе в 1 семестре.

3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам занятий) и на самостоятельную работу обучающихся

Общая трудоемкость (объем) дисциплины составляет 5 зачетных единиц (з.е.), 180 академических часа.

3.1. Объем дисциплины по видам учебных занятий (в часах)

	Семестр		
	№ 1	№ 2	Всего
	Количество часов на вид работы:		
Контактная работа обучающихся с преподавателем			
Аудиторные занятия (всего)	64		64
В том числе:			
лекции	32		16
практические занятия	32		32
лабораторные занятия			-
Промежуточная аттестация			
В том числе:			
зачет	Зачет с оц.		-
экзамен	-		
Самостоятельная работа обучающихся (всего)	116		116
В том числе:			
проработка учебного (теоретического) материала	29		29
выполнение индивидуальных заданий	29		29
подготовка ко всем видам контрольных испытаний текущего контроля успеваемости (в течение семестра)	29		29
подготовка ко всем видам контрольных испытаний промежуточной аттестации (по окончании семестра)	29		29
Всего (часы):	180		180
Всего (зачетные единицы):	5		5

4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

**4.1. Разделы дисциплины и трудоемкость по видам учебных занятий
(в академических часах)**

№ п/п	Наименование раздела /темы дисциплины	Общая трудоём- кость всего (в часах)	Виды учебных занятий, включая самостоятельную работу обучающихся и трудоёмкость (в часах)				Формы текущего контроля успеваемости
			Аудиторные учебные занятия			СРО	
			Лек	Сем/Пр	Лаб		
1.1.	Линейная алгебра	9	4	4		14	Контрольная работа № 1
1.2.	Теория вероятностей	1	4	4		14	
1.3.	Математическая статистика	1	4	4		14	
1.4.	Методы оптимизации	1	4	4		14	Контрольная работа № 2
1.5.	Машинное обучение	2	4	4		14	
1.6.	Элементы теории графов	2	4	4		14	
1.7.	Нейронные сети и глубокое обучение	2	4	4		16	
1.8.	Машинное обучение и интеллектуальный анализ данных	4	4	4		16	

Прим.: Лек – лекции, Сем/Пр – семинары, практические занятия, Лаб – лабораторные занятия, СРО – самостоятельная работа обучающихся

4.2. Содержание дисциплины, структурированное по разделам (темам)

Лекционный курс

№	Наименование раздела /темы дисциплины	Содержание
1.1.	Линейная алгебра	Векторы, матрицы, преобразования матриц, линейные пространства, линейные операторы
1.2.	Теория вероятностей	Случайная величина, распределение случайных величин, условные распределения
1.3.	Математическая статистика	Эмпирическая функция распределения, метод максимального правдоподобия, метод моментов, доверительные интервалы, проверка статистических гипотез
1.4.	Методы оптимизации	Градиентный спуск, стохастический градиентный спуск, методы второго порядка, оптимизация с ограничениями
1.5.	Машинное обучение	Постановка задачи обучения, основные подходы: обучение с учителем, без учителя, с подкреплением
1.6.	Элементы теории графов	Алгоритмы поиска кратчайших путей в графе
1.7.	Нейронные сети и	Многослойные перцептроны, рекуррентные нейронные сети,

	глубокое обучение	глубокие свёрточные нейронные сети, обучение с подкреплением
1.8.	Машинное обучение и интеллектуальный анализ данных	Методы машинного обучения, кластеризация, классификация, регрессия.

Практические/семинарские занятия

№	Наименование раздела /темы дисциплины	Содержание
1.1.	Линейная алгебра	Векторы, матрицы, преобразования матриц, линейные пространства, линейные операторы
1.2.	Теория вероятностей	Случайная величина, распределение случайных величин, условные распределения
1.3.	Математическая статистика	Эмпирическая функция распределения, метод максимального правдоподобия, метод моментов, доверительные интервалы, проверка статистических гипотез
1.4.	Методы оптимизации	Градиентный спуск, стохастический градиентный спуск, методы второго порядка, оптимизация с ограничениями
1.5.	Машинное обучение	Постановка задачи обучения, основные подходы: обучение с учителем, без учителя, с подкреплением
1.6.	Элементы теории графов	Алгоритмы поиска кратчайших путей в графе
1.7.	Нейронные сети и глубокое обучение	Многослойные перцептроны, рекуррентные нейронные сети, глубокие свёрточные нейронные сети, обучение с подкреплением
1.8.	Машинное обучение и интеллектуальный анализ данных	Методы машинного обучения, кластеризация, классификация, регрессия.

Лабораторные занятия

Не предусмотрены.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

В качестве учебно-методических материалов используется рекомендованная литература.

6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

6.1. Паспорт фонда оценочных средств по дисциплине

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или её	Наименование оценочного средства
-------	--	--	----------------------------------

	(результаты по разделам)	части) / и ее формулировка	
1	Метод градиентного спуска	УК-1, УКЦ-1	Контрольная работа № 1
2	Дифференциал функции многих переменных, градиент.	УКЦ-1, ОПК-1	Контрольная работа № 2

6.2. Типовые контрольные задания или иные материалы

6.2.1. Зачет с оценкой

Вопросы к зачету:

- 1 Линейная алгебра. Векторы, матрицы, преобразования матриц, линейные пространства, линейные операторы.
- 2 Теория вероятностей. Случайная величина, распределение случайных величин, условные распределения.
- 3 Математическая статистика. Эмпирическая функция распределения, метод максимального правдоподобия, метод моментов, доверительные интервалы, проверка статистических гипотез.
- 4 Примеры задач регрессии и классификации.
- 5 Методы оптимизации. Градиентный спуск, стохастический градиентный спуск, методы второго порядка, оптимизация с ограничениями.
- 6 Машинное обучение. Постановка задачи обучения, основные подходы: обучение с учителем, без учителя, с подкреплением.
- 7 Глубокое обучение. Архитектура глубоких нейронных сетей, обучение свёрточных нейронных сетей, применение глубоких нейронных сетей в анализе данных.
- 8 Элементы теории графов. Алгоритмы поиска кратчайших путей в графе
- 9 Алгоритмы и структуры данных. Алгоритмы сортировки, хеш-таблицы, бинарные деревья поиска.
- 10 Численные методы. Численное интегрирование, решение систем линейных уравнений, численное решение обыкновенных дифференциальных уравнений.
- 11 Математическое программирование. Линейное и нелинейное программирование, целочисленное программирование, квадратичное программирование.
- 12 Стохастические методы. Марковские процессы, цепи Маркова,
- 13 Анализ временных рядов, прогнозирование временных рядов.
- 14 Статистическая теория обучения. Байесовский подход к обучению, оценка апостериорной вероятности.
- 15 Нейронные сети и глубокое обучение. Многослойные перцептроны, рекуррентные нейронные сети, глубокие свёрточные нейронные сети, обучение с подкреплением.
- 16 Обработка естественного языка и текстовые данные. Методы обработки текста, векторизация текстовых данных, классификация текстов, анализ тональности текстов.
- 17 Визуализация многомерных данных, графики рассеяния, гистограммы, тепловые карты, трёхмерные графики.
- 18 Машинное обучение и интеллектуальный анализ данных. Методы машинного обучения, кластеризация, классификация, регрессия.
- 19 Методы Монте-Карло.

Критерий оценки – правильность и полнота ответа на вопросы. Оценка выставляется по шкале от 0 до 40 баллов: теоретические вопросы – 30 баллов, 10 баллов – дополнительные вопросы. Зачет считается сданным при оценке не ниже 25 баллов.

6.2.2. Контрольная работа № 1

1. В алгоритме градиентного спуска мы надеемся, что для всех i будет выполнено $f(r_i) > f(r_{i+1})$. То есть, что на каждом шаге значение функции уменьшается. Обязательно ли это будет так? Другими словами, может ли так случиться, что мы делаем шаг градиентного спуска и попадаем в точку, в которой значение функции больше, чем в предыдущей?

Если может – объясните, почему, и приведите пример. Если не может – докажите, что не может.

2. Дана функция $f(x) = x^2$. Мы начинаем в точке $r_1 = 1024$.

В какой точке мы окажемся через 1 шаг градиентного спуска с learning rate $1/4$?

В какой точке мы окажемся через 2 шага градиентного спуска с learning rate $1/4$?

В какой точке мы окажемся через 13 шагов градиентного спуска с learning rate $1/4$?

6.2.2. Контрольная работа № 2

1. Дана функция $f(x_1, x_2) = (x_1 + x_2)^2$, мы находимся. Значение $f(x_1, x_2)$ в точке $(3, 4)$ равно $f(3, 4) = (3 + 4)^2 = 49$.

Мы делаем шаг длины $2\sqrt{2}$ в направлении, противоположном направлению градиента. Каково значение функции f в точке, в которой мы оказались?

2. а) Найдите дифференциал функции $g(x_1, x_2, x_3) = -x^4x_3 + x_1x^3x_3 + x_1x^2$ в точке $(2, 5, 4)$

б) Найдите дифференциал функции $h(x_1, x_2, \dots, x_n) = x^2x_2 + x^2x_3 + \dots + x^2x_n + x^2x_1$ в точке $(2, 2, \dots, 2)$

б) критерии оценивания компетенций (результатов) – правильная работа кода программы, понимание алгоритма метода оптимизации, умение вывести необходимые для алгоритма формулы.

в) описание шкалы оценивания:

Каждая задача оценивается по шкале от 0 до 15 баллов.

Контрольная работа считается выполненной успешно при суммарной оценке не ниже 18 баллов.

6.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Форма аттестации	Наименование оценочного средства	Баллы
Зачет (100 баллов)	Контрольная работа № 1	30
	Контрольная работа № 2	30
	Ответы на билет	40

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

а) основная учебная литература:

1. С. В. Куликов, "Математика для анализа данных" – 2020 – 348 страниц

2. В. М. Мельниченко, "Математические методы анализа данных" – 2021 – 426 страниц
3. А. В. Скуридин, "Математика для машинного обучения" – 2021 – 392 страницы
4. А. И. Гельфанд, "Математические методы для анализа данных и машинного обучения" – 2022 – 462 страницы

б) дополнительная учебная литература:

1. С. А. Лебедев, "Математические методы анализа больших данных" – 2023 – 365 страниц
2. В. С. Петров, "Основы математики для анализа данных и оптимизации" – 2021 – 444 страницы
3. А. В. Левин, "Математика и статистика в анализе данных" – 2020 – 390 страниц
4. Ю. М. Филипов, "Теория вероятностей и статистика для анализа данных" – 2022 – 402 страницы
5. Д. В. Чистяков, "Математическое моделирование и анализ данных" – 2021 – 318

8. Перечень ресурсов* информационно-телекоммуникационной сети «Интернет» (далее - сеть «Интернет»), необходимых для освоения дисциплины

-

9. Методические указания для обучающихся по освоению дисциплины

Вид учебного занятия	Организация деятельности студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; помечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначить вопросы, термины, материал, который вызывает трудности, пометить и попытаться найти ответ в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на консультации, на практическом занятии.
Практические занятия	Проработка рабочей программы, уделяя особое внимание целям и задачам, структуре и содержанию дисциплины. Работа с конспектом лекций, просмотр рекомендуемой литературы. Изучение выбранной предметной области на примерах решения задач семинарских занятий, индивидуальных домашних заданий.
Курсовая работа	Не предусмотрена
Контрольная работа	Ознакомиться с основной и дополнительной литературой, включая справочные издания, зарубежные источники, основополагающие термины. Попрактиковаться в решении аналогичных домашних задач по всем темам контрольных работ.
Лабораторная работа	Не предусмотрена.
Подготовка к зачету	При подготовке к зачету необходимо ориентироваться на конспекты лекций и рекомендуемую литературу.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Издательская система LaTeX для подготовки докладов, презентаций и учебного материала.

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Видеопроектор, компьютер, издательская система LaTeX для подготовки докладов, презентаций и учебного материала.

12. Иные сведения и (или) материалы

12.1. Перечень образовательных технологий, используемых при осуществлении образовательного процесса по дисциплине

Часов в интерактивной форме – 8.

В ходе практических занятий происходит публичное обсуждение каждой решаемой задачи. При этом студенты высказывают свои мнения по выбору наиболее простого способа поиска оптимального решения.

После решения домашних работ на консультациях проводится разбор допущенных студентами ошибок.

12.2. Формы организации самостоятельной работы обучающихся (темы, выносимые для самостоятельного изучения; вопросы для самоконтроля; типовые задания для самопроверки)

Некоторые темы изучаются студентами самостоятельно. Для изучения используется приведённая в списке основная и дополнительная литература. Контроль освоения материала осуществляется при проверке контрольных работ, домашнего задания и на зачете.

№	Тема и часть, изучаемая (осваиваемая) самостоятельно
1.1	Стохастические методы. Марковские процессы, цепи Маркова
1.2	Оптимизация и математическое программирование. Задачи оптимизации, линейное и нелинейное программирование, целочисленное программирование, квадратичное программирование
1.3	Обработка естественного языка и текстовые данные. Методы обработки текста, векторизация текстовых данных, классификация текстов, анализ тональности текстов

12.3. Краткий терминологический словарь

Регрессия	метод статистического анализа, используемый для моделирования зависимости между зависимой переменной (выходом) и одной или несколькими независимыми переменными (входами).
Метод главных компонент (РСА)	техника линейного уменьшения размерности, которая используется для упрощения данных с сохранением важной информации, путем проектирования данных на новое пространство с меньшим количеством измерений.