

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Костромской государственный университет»
(КГУ)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Математика анализа данных

Направление подготовки – **01.04.02 «Прикладная математика и информатика»**

Направленность «**Математическое моделирование и программирование**»

Квалификация (степень) выпускника: магистр

**Кострома
2024**

1. Цели и задачи освоения дисциплины

Цель дисциплины: сформировать математический фундамент для освоения методов анализа больших данных.

Задачи дисциплины:

- повторить основные концепции математического анализе, линейной алгебры, методов оптимизации, теории вероятностей и математической статистики;
- овладеть принципами и понятийным аппаратом, описывающими современные методы анализа данных;
- усвоить теоретические основы современных технологий и методов анализа данных.

2. Перечень планируемых результатов обучения по дисциплине

В результате освоения дисциплины обучающийся должен:

освоить компетенции:

ОПК1: Способен решать актуальные задачи фундаментальной и прикладной математики;

Код и содержание индикаторов компетенции

Индикаторы ОПК1:

ОПК-1.1. Знать: актуальные задачи фундаментальной и прикладной математики;

ОПК-1.2. Уметь: решать профессиональные задачи, используя методы фундаментальной и прикладной математики;

ОПК-1.3. Иметь навыки: теоретического и экспериментального исследования объектов профессиональной деятельности с применением фундаментальной и прикладной математики.

Знать:

– основные концепции математического анализа, линейной алгебры, математической статистики;

– основные литературные источники, в том числе интернет-ресурсы, отражающие современный уровень развития методов анализа данных;

Уметь:

– составить математическую модель анализа данных;

– подбирать необходимые программные инструменты для решения вычислительных задач.

Владеть:

– методами и навыками решения задач анализа данных.

3. Место дисциплины в структуре ОП ВО

Дисциплина относится к обязательной части учебного плана. Изучается в первом семестре(ах) обучения.

Изучение дисциплины основывается на ранее освоенных дисциплинах/практиках, полученных в рамках бакалавриата: математический анализ, линейная алгебра, методы оптимизации, теория вероятностей и математическая статистика.

Изучение дисциплины является основой для освоения последующих дисциплин/практик: Управление разработкой программных продуктов, организационно-управленческая практика.

Вместе с математикой анализа данных компетенции ОПК-1 формируют также дисциплины современные проблемы прикладной математики, научно-исследовательская работа, защита выпускной квалификационной работы.

4. Объем дисциплины

4.1. Объем дисциплины в зачетных единицах с указанием академических часов и виды учебной работы

Виды учебной работы,	Очная форма	Очно-заочная	Заочная
Общая трудоемкость в зачетных единицах	4		
Общая трудоемкость в часах	144		
Аудиторные занятия в часах, в том числе:	36		
Лекции	14		
Практические занятия	22		
Лабораторные занятия	-		
Самостоятельная работа в часах	108		
Форма промежуточной аттестации	Экзамен		

4.2. Объем контактной работы на 1 обучающегося

Виды учебных занятий	Очная форма	Очно-заочная	Заочная
Лекции	14		
Практические занятия	22		
Лабораторные занятий			
Консультации	2		
Зачет/зачеты	-		
Экзамен/экзамены	0,35		
Курсовые работы			
Курсовые проекты			
Всего	38,35		

5 Содержание дисциплины, структурированное по темам (разделам), с указанием количества часов и видов занятий

5.1 Тематический план учебной дисциплины

№	Название раздела, темы	Всего з.е/час	Аудиторные занятия			Самостоятельная работа
			Лекц.	Практ.	Лаб.	
1	Машинное обучение и анализ данных	0,6/22	2	2		18
2	Язык Python	0,72/26	4	4		18
3	Библиотеки Python и линейная алгебра	0,67/24	2	4		18
4	Оптимизация и матричные разложения	0,67/24	2	4		18
5	Теория вероятностей	0,67/24	2	4		18
6	Статистика	0,67/24	2	4		18
	Итого:	4/144	14	22	0	108

5.2. Содержание:

Тема 1. Машинное обучение и анализ данных. Анализ данных и машинное обучение опираются на результаты из математического анализа, линейной алгебры, методов оптимизации, теории вероятностей. Без фундаментальных знаний по этим наукам невозможно понять, как строены методы анализа данных. Для успешного применения методов анализа данных нужно уметь программировать. Наиболее удобным языком анализа данных является Python.

Тема 2. Язык Python. Python – один из главных инструментов специалиста в науке о данных. Установка Анаконды. Ноутбуки. Типы данных. Циклы, функции, генераторы, list compression. Чтение данных из файлов. Запись данных, изменение файлов. Функции и их свойства. Предел и производная. Геометрический смысл производной. Производная сложной функции. Задача нахождения экстремума. Вторая производная и выпуклость.

Тема 3. Библиотека Python и линейная алгебра. Реализация операций с многомерными массивами данных. Pandas. Data Frame. Индексация и селекция. Знакомство с NumPy, SciPy и Matplotlib. Решение оптимизационных задач в SciPy. Векторные пространства. Матричные операции.

Тема 4. Оптимизация и матричные разложения. Нахождение наилучших значений параметров системы. Минимизация затрат и максимизация точности предсказаний. Матричные разложения при построении регрессионных моделей. Уменьшение размерности данных. Частные производные и градиент. Применение градиента. Производные по направлению. Касательная плоскость и линейное приближение. Направление наискорейшего роста. Оптимизация негладких функций. Метод имитации отжига. Генетические алгоритмы и дифференциальная эволюция. Разложение матриц в произведение. Сингулярное разложение. Приближение матрицей меньшего ранга. Связь сингулярного разложения и приближения матрицей меньшего ранга.

Тема 5. Теория вероятностей. Базовые концепции теории вероятностей. Случайность в теории вероятностей и статистике. Свойства вероятности. Условная вероятность. Дискретные и случайные величины. Непрерывные случайные величины. Центральная предельная теорема.

Тема 6. Статистика. Популярные распределения. Статистики оценки параметров. Доверительные интервалы. Оценка распределения по выборке. Важные характеристики распределений.

6. Методические материалы для обучающихся по освоению дисциплины

6.1. Самостоятельная работа обучающихся по дисциплине (модулю)

№ п/п	Раздел (тема) дисциплины	Задание	Часы	Методические рекомендации по выполнению задания (при необходимости)	Форма контроля
1.	Машинное обучение и анализ данных	Основные методы анализа больших данных	10	См.[6, гл.1,2.]	Реферативно
2.	Язык Python	Подключение и использование статистических библиотек	10	См.[3, гл.2, 3.]	Отчет по заданию
3.	Библиотеки Python и линейная алгебра	Решение задач линейной алгебры с помощью среды программирования	10	Х.А.Таха. Введение в исследование операций.	Отчет по заданию
4.	Оптимизация	Решение задач	10	См.[3, гл.4.]	Отчет по

	и матричные разложения	оптимизации с помощью среды программирования			заданию
5.	Теория вероятностей	Решение задач теории игр с помощью Python	10	См.[3, гл.6.]	Отчет по заданию
6.	Статистика	Решение статистических задач с помощью Python	10	См.[1, гл.5.]	Отчет по заданию

6.2. Тематика и задания для практических занятий

Тема 1. Машинное обучение и анализ данных. Основные задачи математического анализа. Нахождение производных. Задачи линейной алгебры для анализа данных. Матричные вычисления. Задачи методов оптимизации. Задачи теории вероятностей и математической статистики.

Тема 2. Язык Python. Работа с Python Notebook. Синтаксис языка. Работа с файлами в Python. Функции и экстремумы. Производная и ее применения.

Тема 3. Библиотека Python и линейная алгебра. Операции с матрицами с помощью библиотек NumPy, SciPy и Matplotlib. Решение оптимизационных задач в SciPy. Векторные пространства. Разрешимость систем линейных уравнений и ранги.

Тема 4. Оптимизация и матричные разложения. Нахождение частных производных. Применение градиента. Повторение: гладкость и градиентный спуск. Методы оптимизации в негладких задачах. Линейная алгебра. Повторение. Матричные разложения.

Тема 5. Теория вероятностей. Работа со случайными величинами (ipython notebook).

Тема 6. Статистика. Анализ больших данных. Построение выборки. Выбор распределения. Оценка распределения по выборке (ipython notebook).

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

а) основная:

1. **Прерва А., Иванова В.** Путь аналитика. Практическое руководство IT-специалиста. – СПб : Питер, 2015. - 304 с.

2. **Тыртышников, Е. Е.** Матричный анализ и линейная алгебра. - М. : Физматлит, 2007. - 480 с.

3. **Гмурман, В.Е.** Теория вероятностей и математическая статистика : [учеб. пособие для студ.]. - М. : Высшая школа, 2002. - 480 с.

б) дополнительная:

4. **Вентцель, Е. С.** Исследование операций: задачи, принципы, методология : учеб. пособие для студ. вузов, напр. "Математика", "Компьютерные науки". - 4-е изд., стер. - Москва : Дрофа, 2006.

5. **Васин, А. А.** Исследование операций : [учеб. пособие для студ. высш. учеб. заведений] : рекомендовано Науч.-метод. советом. - М. : Академия, 2008. - 464 с.

6. **Волков, И. К.** Исследование операций : Учеб. для студ. высш. техн. учеб. заведений / Под ред. В. С. Зарубина, А. П. Крищенко. - 3-е изд., стер. - М. : МГТУ, 2004. - 440 с.

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

Информационно-образовательные ресурсы:

1. Библиотека ГОСТов. Все ГОСТы, [Электронный ресурс], URL:<http://vsegost.com/>
2. Национальный открытый университет <http://intuit.ru/>

Электронные библиотечные системы:

1. ЭБС Университетская библиотека онлайн - <http://biblioclub.ru>
2. ЭБС «Лань» <https://e.lanbook.com>
3. ЭБС «ZNANIUM.COM» <http://znanium.com>

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

Лекционные занятия проводятся в аудиториях с требуемым числом посадочных мест, оборудованные мультимедиа.

Практические занятия проводятся в компьютерных классах. Необходимое программное обеспечение:

Лицензионное программное обеспечение:

Среда программирования Анаконда.

Свободно распространяемое программное обеспечение:

Офисный пакет.