Федеральное государственное бюджетное образовательное учреждение высшего образования
Московский государственный университет имени М.В. Ломоносова
ФИЛИАЛ МГУ В Г. ДУБНЕ

УТВЕРЖДАЮ

И.о. директора

филиала МГУ в г. Дубне

_____/ Э.Э. Боос /

«01» сентября 2025г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины:

Статистические методы обработки данных

Уровень высшего образования:

Магистратура

Направление подготовки / специальность:

01.04.02 «Прикладная математика и информатика»

Направленность (профиль) ОПОП:

Методы и технологии обработки данных в гетерогенных вычислительных средах

Форма обучения:

очная

Дубна 2025

Рабочая программа дисциплины (модуля) разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 01.04.02 "Прикладная математика и информатика" программы магистратуры в редакции приказа МГУ от «___»_____20___ г. №_____.

1. Место дисциплины (модуля) в структуре ОПОП ВО:
Дисциплина относится к вариативной части ОПОП ВО.

2. Входные требования для освоения дисциплины (модуля), предварительные условия (если есть):

Студенты должны обладать знаниями по математической статистике и теории вероятностей, программированию, численным методам оптимизации.

3. Результаты обучения по дисциплине (модулю):

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	**Индикатор (показатель) достижения компетенции**	**Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций**
СПК-2. Способность понимать и применять в исследовательской и прикладной деятельности современные суперкомпьютерные технологии, математический аппарат, вычислительные методы для проведения крупномасштабного математического моделирования и обработки данных на современных высокопроизводительных вычислительных системах	СПК-2.1. Применяет в исследовательской и прикладной деятельности современные суперкомпьютерные технологии, математический аппарат, вычислительные методы для проведения крупномасштабного математического моделирования и обработки данных на современных высокопроизводительных вычислительных системах	**Знать** Погрешность измерений, и их классификация. Виды и типы данных. Эксперименты физики высоких энергий **Уметь** Проводить регрессионный анализ данных **Иметь практический опыт** Разработки статистических моделей для проверки гипотез и определения статистической значимости
СПК-3. Способность разрабатывать гетерогенные вычислительные среды для решения научных задач крупных проектов, включая проекты класса мегасайенс.	СПК-3.1. Разрабатывает гетерогенные вычислительные среды для решения научных задач крупных проектов, включая проекты класса мегасайенс.	**Знать** Основы статистического анализа экспериментальных данных **Уметь** Создавать модели для анализа временных рядов. **Иметь практический опыт** Разработки статистических моделей для решения научных задач крупных проектов, включая проекты класса мегасайенс

4. Объем дисциплины составляет 3 з.е., в том числе 64 академических часов, отведенных на контактную работу обучающихся с преподавателем, 44 академических часов на самостоятельную работу обучающихся.

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий:

5.1. Структура дисциплины (модуля) по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий (в строгом соответствии с учебным планом)

Наименование разделов и тем дисциплины (модуля), **Форма промежуточной аттестации по дисциплине (модулю)**	**Номинальные трудозатраты обучающегося**			**Всего академических часов**	**Форма текущего контроля успеваемости* (наименование)**
	Контактная работа (работа во взаимодействии с преподавателем) Виды контактной работы, академические часы		**Самостоятельная работа обучающегося, академические часы**		
	Занятия лекционного типа	Занятия семинарского типа			
Тема 1. Погрешность измерений, и их классификация. Виды и типы данных. Эксперименты физики высоких энергий.	4	4	4	12	опрос
Тема 2. Основы статистического анализа экспериментальных данных.	4	4	4	12	опрос
Тема 3. Регрессионный анализ ч.I.	4	4	4	12	опрос
Тема 4. Регрессионный анализ ч.II.	4	4	4	12	опрос
Тема 5. Проверка гипотез и статистическая значимость.	4	4	6	14	опрос
Тема 6. Методы кластеризации.	4	4	6	14	опрос
Тема 7. Анализ временных рядов.	4	4	8	16	опрос
Тема 8. Введение в машинное обучение.	4	4	8	16	опрос
Другие виды самостоятельной работы (отсутствуют)	—	—			—
Промежуточная аттестация (зачет)					
Итого	*32*	*32*	*40*	108	—

5.2. Содержание разделов (тем) дисциплины

№	Наименование разделов (тем)	Содержание разделов (тем) дисциплин

п/п	дисциплины	
1.	Тема 1. Погрешность измерений, и их классификация. Виды и типы данных. Эксперименты физики высоких энергий.	Природа погрешностей и их влияние на интерпретацию измерений. Обзор видов и типов данных. Примеры измерений и данных в экспериментах физики высоких энергий с акцентом на эксперименты мегасайенс проекта НИКА.
2.	Тема 2. Основы статистического анализа экспериментальных дынных.	Графические методы представления и анализа данных измерений. Точечные и интервальные оценки неизвестных параметров распределений. Сравнение двух выборочных совокупностей.
3.	Тема 3. Регрессионный анализ ч.I.	Линейная регрессия. Метод максимального правдоподобия и метод наименьших квадратов. Понятие робастных методов, взвешенный метод наименьших квадратов.
4.	Тема 4. Регрессионный анализ ч.II.	Нелинейная регрессия. Понятие целевой функции. Градиентные и неградиентные методы численной оптимизации.
5.	Тема 5. Проверка гипотез и статистическая значимость.	Методы проверки статистических гипотез. Статистическая значимость. Статистический критерий. Байесовский анализ.
6.	Тема 6. Методы кластеризации.	Задача кластеризации и основные методы её решения. Классификация методов кластеризации. Меры расстояний.
7.	Тема 7. Анализ временных рядов.	Методы анализа временных рядов. Быстрое Фурье-преобразование. Вейвлет-преобразование и алгоритм быстрого вейвлет-преобразование. Сглаживание данных и частотный анализ временных рядов.
8.	Тема 8. Введение в машинное обучение.	Разведывательный анализа данных. Метод опорных векторов. Бэггинг и случайный лес. Деревья решений и градиентный бустинг на решающих деревьях. Задача классификации и регрессии.

6. Фонд оценочных средств (ФОС, оценочные и методические материалы) для оценивания результатов обучения по дисциплине (модулю).

6.1. Типовые контрольные задания или иные материалы для проведения текущего контроля успеваемости

Примеры тем для устного опроса:

Тема 1: Погрешность измерений, и их классификация. Виды и типы данных. Эксперименты физики высоких энергий.

1. Что такое погрешность измерений? Какие основные типы погрешностей существуют?
2. Какие виды и типы данных вы знаете? Приведите примеры каждого вида.
3. Приведите примеры данных, получаемых в экспериментах физики высоких энергий. Какие ошибки в данных характерны для приведённого примера?

Тема 2: Основы статистического анализа экспериментальных дынных

4. Какие графические методы представления данных используются для анализа экспериментальных результатов?
5. Что такое точечные и интервальные оценки неизвестных параметров распределений? В чем их различия?
6. Как можно сравнить две выборочные совокупности?

Тема 3: Регрессионный анализ I.

7. Метод максимального правдоподобия и метод наименьших квадратов.
8. В чем заключается понятие робастных методов линейной регрессии, и когда их следует применять?

Тема 4: Регрессионный анализ II.

9. В чем основные принципы нелинейной регрессии? Какие методы используются для её аппроксимации?
10. Как можно описать целевую функцию для нелинейной регрессии? Какие методы численной оптимизации применяются для её минимизации?
11. Какие преимущества и недостатки градиентных и неградиентных методов численной оптимизации?

Тема 5: Проверка гипотез и статистическая значимость

12. Какие методы используются для проверки статистических гипотез? Как определить статистическую значимость?
13. Что такое статистический критерий? Какие основные критерии существуют?
14. В чем основные отличия между классическим и байесовским подходами к проверке гипотез?

Тема 6: Методы кластеризации

15. Какова цель задачи кластеризации? Приведите пример методов кластеризации.
16. Приведите примеры меры расстояний для определения сходства между объектами.
17. Классифиция методов кластеризации.

Тема 7: Анализ временных рядов

18. Какие особенности характерны для анализа временных рядов, и почему они требуют специальных методов обработки?
19. Быстрое Фурье-преобразование. Как его можно применить к анализу временных рядов?
21. Основные принципы вейвлет-преобразования? Быстрое-вейвлет преобразование.

Тема 8: Введение в машинное обучение

22. Что такое разведывательный анализ данных (EDA)? Какие методы он в себя включает?
23. Что такое метод опорных векторов (SVM), и в каких задачах его можно применять?
24. Деревья решений, бэггинг и случайный лес.
25. Градиентный бустинг на решающих деревьях.

6.2. Типовые контрольные задания или иные материалы для проведения промежуточной аттестации по дисциплине, критерии и шкалы оценивания

Примеры вопросов к зачету

1. Какие требования предъявляют к оценкам? Поясните понятия: точечная оценка; генеральное и выборочное среднее; генеральная и выборочная дисперсия.
2. В чем заключается метод максимального правдоподобия?
3. В чем заключается метод моментов? Что такое интервальное оценивание, доверительный интервал, доверительная вероятность?
4. Что такое статистическая гипотеза? Приведите примеры нулевой и конкурирующей гипотез. Что означают ошибки 1 и 2 рода при проверке статистических гипотез. Приведите порядок проверки гипотезы.
5. Как проверить гипотезы о сравнении средних при различных условиях? Как проверить гипотезы о сравнение дисперсий? Критерий Фишера. Что такое непараметрические и параметрические гипотезы?
6. Чем различаются корреляционный анализ и регрессионный анализ?
7. Что такое линия регрессии?
8. Какой порядок выполнения регрессионного анализа?
9. Особенности регрессионного анализа при полиномиальной зависимости?
10. Когда применима линеаризация функциональной зависимости? Какие её недостатки?
11. Метод Гаусса-Ньютона. Когда применим и основные положения?
12. Шаг λ и демпфер α? Как они выбираются? По каким критериям останавливается процесс минимизации? Что такое Робастные оценки?
13. Априорная информация и как она учитывается в программах обработки данных?

14. Сравните метод максимального правдоподобия и метод наименьших квадратов оценки параметров. В чём достоинства и недостатки каждого из методов?
15. Что такое псевдослучайные числа? В каких случаях вместо последовательности случайных чисел может быть использована последовательность псевдослучайных чисел?
16. В чём заключается особенность экспериментальных данных в физике частиц?

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине				
Оценка / виды оценочных средств	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)
Знания *(виды оценочных средств: опрос, тесты)*	Отсутствие знаний	Фрагментарные знания	Общие, но не структурированные знания	Сформированные систематические знания
Умения *(виды оценочных средств: практические задания)*	Отсутствие умений	В целом успешное, но не систематическое умение	В целом успешное, но содержащее отдельные пробелы умение (допускает неточности непринципиального характера)	Успешное и систематическое умение
Навыки (владения, опыт деятельности) *(виды оценочных средств: выполнение и защита курсовой работы, отчет по практике, отчет по НИР и т.п.)*	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но используемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач

7. Ресурсное обеспечение:

7.1. Перечень основной и дополнительной литературы

Основная литература

1. З. Брандт. Анализ данных. Мир. 2003, Москва. (S.Brandt. Data analysis. Springer.)
2. Ватутин В. А., Ивченко Г.И., Медведев Ю. И., Чистяков В. П., Теория вероятностей и математическая статистика. Агар, М. 2003.
3. Г.В. Вавилова. Математическая обработка результатов измерения. Томский политехнический университет. 2013

Дополнительная литература

1. Злоказов В. Б. Математические методы для анализа экспериментальных спектров и спектро-подобных распределений. ФЭЧАЯ, 1985, т. 16-5, с.1126-1163.

2. Гришин В. К., Живописцев Ф. А., Иванов В. А., Математический анализ и интерпретация физического эксперимента. М., МГУ, 1988.
3. Pecharsky V.K., Zavalij P.Y. Fundamentals of powder diffraction and structural characterization of materials. - Springer, 2005.
4. Громилов С.А. Введение в рентгенографию поликристаллов. Учеб.-метод. пособие. – Новосибирск: НГУ, 2009.

7.2. Перечень лицензионного программного обеспечения, в том числе отечественного производства

При реализации дисциплины может быть использовано следующее программное обеспечение:

- Операционная система Astra Linux (https://astralinux.ru/) или аналог, с офисным пакетом, пакетами разработчика.
- Интерактивный блокнот с web-интерфейсом Google Colab (http://colab.research.google.com) и/или JupyterHub в экосистеме для задач машинного/глубокого обучения и анализа данных Гетерогенной вычислительной платформы «HybriLIT» (http://hlit.jinr.ru),
- Библиотеки языка программирования Python, такие, например, как NumPy, SciPy, Matplotlib, Pandas, Sklearn.

7.3. Перечень профессиональных баз данных и информационных справочных систем

1. http://www.edu.ru – портал Министерства образования и науки РФ
2. http://www.ict.edu.ru – система федеральных образовательных порталов «ИКТ в образовании»
3. http://www.openet.ru - Российский портал открытого образования
4. http://www.mon.gov.ru - Министерство образования и науки Российской Федерации
5. http://www.fasi.gov.ru - Федеральное агентство по науке и инновациям

7.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Math-Net.Ru [Электронный ресурс] : общероссийский математический портал / Математический институт им. В. А. Стеклова РАН ; Российская академия наук, Отделение математических наук. - М. : [б. и.], 2010. - Загл. с титул. экрана. - Б. ц.
URL: http://www.mathnet.ru
2. Университетская библиотека Online [Электронный ресурс] : электронная библиотечная система / ООО "Директ-Медиа" . - М. : [б. и.], 2001. - Загл. с титул. экрана. - Б. ц. URL: www.biblioclub.ru
3. Универсальные базы данных East View [Электронный ресурс] : информационный ресурс / East View Information Services. - М. : [б. и.], 2012. - Загл. с титул. экрана. - Б. ц.
URL: www.ebiblioteka.ru
4. Научная электронная библиотека eLIBRARY.RU [Электронный ресурс] : информационный портал / ООО "РУНЭБ" ; Санкт-Петербургский государственный университет. - М. : [б. и.], 2005. - Загл. с титул. экрана. - Б. ц.
URL: www.eLibrary.ru

7.5. Описание материально-технического обеспечения.

Филиал МГУ в г. Дубне, ответственный за реализацию данной Программы, располагает соответствующей материально-технической базой, включая современную вычислительную технику, объединенную в локальную вычислительную сеть, имеющую выход в Интернет. Используются специализированные компьютерные классы, оснащенные современным оборудованием. Материальная база подразделения соответствует действующим санитарно-техническим нормам и обеспечивает проведение всех видов занятий (лабораторной, практической, дисциплинарной и междисциплинарной подготовки) и научно-исследовательской работы обучающихся, предусмотренных учебным планом.

8. Методические рекомендации по организации изучения дисциплины

8.1. Формы и методы преподавания дисциплины

(перечисляются в соответствии с таблицей 5.1.)

Используемые формы и методы обучения:

лекции

семинарские занятия

самостоятельная работа студентов.

В процессе преподавания дисциплины преподаватель использует как классические формы и методы обучения (лекции и семинарские занятия), так и активные методы обучения.

При проведении лекционных занятий преподаватель использует аудиовизуальные, компьютерные и мультимедийные средства обучения, а также демонстрационные и наглядно-иллюстрационные (в том числе раздаточные) материалы.

Семинарские занятия проводятся в форме проблемной ситуации, когда некоторый аспект рассмотренной темы излагается преподавателем более подробно. Часть информации конспектируется. Большая часть времени выделена на работу с использованием компьютерной техники и программного обеспечения.

В рамках курса используются активные и интерактивные методы обучения в процессе проведения занятий. Основными особенностями интерактивных занятий являются интерактивные практические упражнения и задания, которые выполняются обучающимися не только и не столько на закрепление изученного материала, но и на самостоятельное изучение нового.

8.2. Методические рекомендации преподавателю

Перед началом изучения дисциплины преподаватель должен ознакомить студентов с видами учебной и самостоятельной работы, перечнем литературы и интернет-ресурсов, формами текущей и промежуточной аттестации, с критериями оценки качества знаний для итоговой оценки по дисциплине.

При проведении лекций, преподаватель:

1) формулирует тему и цель занятия;

2) излагает основные теоретические положения;

3) с помощью мультимедийного оборудования и/или под запись дает определения основных понятий, расчетных формул;

4) проводит примеры из отечественного и зарубежного опыта, дает текущие статистические данные для наглядного и образного представления изучаемого материала;

5) в конце занятия дает вопросы для самостоятельного изучения.

Для семинарских занятий

Подготовка к проведению занятий проводится регулярно. Организация преподавателем семинарских занятий должна удовлетворять следующим требования: количество занятий должно соответствовать учебному плану программы, содержание планов должно соответствовать программе, план занятий должен содержать перечень рассматриваемых вопросов.

Во время семинарских занятий используются словесные методы обучения, как беседа и дискуссия, что позволяет вовлекать в учебный процесс всех слушателей и стимулирует творческий потенциал обучающихся.

При подготовке семинарскому занятию преподавателю необходимо знать план его проведения, продумать формулировки и содержание учебных вопросов, выносимых на обсуждение.

В начале занятия преподаватель должен раскрыть теоретическую и практическую значимость темы занятия, определить порядок его проведения, время на обсуждение каждого учебного вопроса. В ходе занятия следует дать возможность выступить всем желающим и предложить выступить тем слушателям, которые проявляют пассивность.

Целесообразно, в ходе обсуждения учебных вопросов, задавать выступающим и аудитории дополнительные и уточняющие вопросы с целью выяснения их позиций по существу обсуждаемых проблем, а также поощрять выступление с места в виде кратких дополнений. На занятиях проводится отработка практических умений под контролем преподавателя

8.3. Методические рекомендации студентам по организации самостоятельной работы.

Приступая к изучению новой учебной дисциплины, студенты должны ознакомиться с учебной программой, учебной, научной и методической литературой, имеющейся в библиотеке университета, встретиться с преподавателем, ведущим дисциплину, получить в библиотеке рекомендованные учебники и учебно-методические пособия, осуществить запись на соответствующий курс в среде электронного обучения университета.

Глубина усвоения дисциплины зависит от активной и систематической работы студента на лекциях и практических занятиях, а также в ходе самостоятельной работы, по изучению рекомендованной литературы.

На лекциях важно сосредоточить внимание на ее содержании. Это поможет лучше воспринимать учебный материал и уяснить взаимосвязь проблем по всей дисциплине. Основное содержание лекции целесообразнее записывать в тетради в виде ключевых фраз, понятий, тезисов, обобщений, схем, опорных выводов. Необходимо обращать внимание на термины, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации. Желательно оставлять в конспектах поля, на которых делать пометки из рекомендованной литературы, дополняющей материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений. С целью уяснения теоретических положений, разрешения спорных ситуаций необходимо задавать преподавателю уточняющие вопросы. Для закрепления содержания лекции в памяти, необходимо во время самостоятельной работы внимательно прочесть свой конспект и дополнить его записями из учебников и рекомендованной литературы. Конспектирование читаемых лекций и их последующая доработка способствует более глубокому усвоению знаний, и поэтому являются важной формой учебной деятельности студентов.

Методические указания для самостоятельной работы обучающихся

Прочное усвоение и долговременное закрепление учебного материала невозможно без продуманной самостоятельной работы. Такая работа требует от студента значительных усилий, творчества и высокой организованности. В ходе самостоятельной работы студенты выполняют следующие задачи: дорабатывают лекции, изучают рекомендованную литературу, готовятся к практическим занятиям, к коллоквиуму, контрольным работам по отдельным темам дисциплины. При этом эффективность учебной деятельности студента во многом зависит от того, как он распорядился выделенным для самостоятельной работы бюджетом времени.

Результатом самостоятельной работы является прочное усвоение материалов по предмету согласно программы дисциплины. В итоге этой работы формируются профессиональные умения и компетенции, развивается творческий подход к решению возникших в ходе учебной деятельности проблемных задач, появляется самостоятельности мышления.

Решение задач

При самостоятельном решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения проблемы (задачи), то нужно сравнить их и выбрать самый рациональный. Полезно до начала вычислений составить краткий план решения проблемы (задачи).

Решение проблемных задач или примеров следует излагать подробно, вычисления располагать в строгом порядке, отделяя вспомогательные вычисления от основных. Решения при необходимости нужно сопровождать комментариями, схемами, чертежами и рисунками.

Следует помнить, что решение каждой учебной задачи должно доводиться до окончательного логического ответа, которого требует условие, и по возможности с выводом.

Полученный ответ следует проверить способами, вытекающими из существа данной задачи. Полезно также (если возможно) решать несколькими способами и сравнить полученные результаты.

Решение задач данного типа нужно продолжать до приобретения твердых навыков в их решении.

Задача — это цель, заданная в определенных условиях, решение задачи — процесс достижения поставленной цели, поиск необходимых для этого средств.

Алгоритм решения задач:

1. Внимательно прочитайте условие задания и уясните основной вопрос, представьте процессы и явления, описанные в условии.

2. Повторно прочтите условие для того, чтобы чётко представить основной вопрос, проблему,

цель решения, заданные величины, опираясь на которые можно вести поиски решения.

3. Произведите краткую запись условия задания.

4. Если необходимо составьте таблицу, схему, рисунок или чертёж.

5. Определите метод решения задания, составьте план решения.

6. Запишите основные понятия, формулы, описывающие процессы, предложенные заданной системой.

7. Найдите решение в общем виде, выразив искомые величины через заданные.

9. Проверьте правильность решения задания.

10. Произведите оценку реальности полученного решения.

11. Запишите ответ.

9. Разработчик (разработчики) программы.

канд. физ.-мат. наук Александр Сержикович Айриян