

Федеральное государственное бюджетное образовательное учреждение
высшего образования
Московский государственный университет имени М.В. Ломоносова
Факультет вычислительной математики и кибернетики

УТВЕРЖДАЮ

декан факультета вычислительной
математики и кибернетики



/И.А. Соколов /

«14» октября 2021г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Наименование дисциплины (модуля):

Прикладной многомерный статистический анализ

Уровень высшего образования:

магистратура

Направление подготовки / специальность:

01.04.02 "Прикладная математика и информатика" (3++)

Направленность (профиль) ОПОП:

Искусственный интеллект в кибербезопасности

Форма обучения:

очная

Рабочая программа рассмотрена и утверждена
на заседании Ученого совета факультета ВМК
(протокол № 7, от 29 сентября 2021 года)

Москва 2021

Рабочая программа дисциплины (модуля) разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования (ФГОС ВО) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки 01.04.02 "Прикладная математика и информатика" утвержденного Приказом Министерства образования и науки РФ от 10 января 2018 г. N 13.

1. Место дисциплины (модуля) в структуре ОПОП ВО:

Дисциплина (модуль) относится к части дисциплин основной профессиональной образовательной программы, формируемых участниками образовательных отношений.

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

Учащиеся должны владеть знаниями по теории вероятностей, математической статистике, машинному обучению в объеме, соответствующем основным образовательным программам бакалавриата по укрупненным группам направлений и специальностей 01.00.00 «Математика и механика», 02.00.00 «Компьютерные и информационные науки» и другим направлениям подготовки бакалавриата.

3. Результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников.

Планируемые результаты обучения по дисциплине (модулю)		
Содержание и код компетенции.	Индикатор (показатель) достижения компетенции	Планируемые результаты обучения по дисциплине, сопряженные с индикаторами достижения компетенций
ПК-5. Способен руководить проектами по созданию систем искусственного интеллекта применением новых методов и алгоритмов машинного обучения со стороны заказчика	ПК-5.1. Руководит разработкой архитектуры комплексных систем искусственного интеллекта со стороны заказчика ПК-5.2. Руководит созданием систем искусственного интеллекта на основе моделей искусственных нейронных сетей и инструментальных средств	ПК-5.1. 3-1. Знает возможности современных инструментальных средств и систем программирования для решения задач машинного обучения ПК-5.1. У-1. Умеет проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения ПК-5.2. 3-1. Знает функциональность современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения ПК-5.2. 3-2. Знает принципы построения систем искусственного интеллекта, методы и подходы к планированию и реализации проектов по созданию систем искусственного интеллекта, методы интеллектуального планирования экспериментов ПК-5.2. У-1. Умеет применять современные инструментальные средства и системы программирования для разработки новых методов и моделей машинного обучения ПК-5.2. У-2. Умеет руководить выполнением коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта

4. Объем дисциплины (модуля) составляет 3 з.е., в том числе 36 академических часа, отведенных на контактную работу обучающихся с преподавателем, 72 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:

В рамках данного курса будут рассмотрены основные задачи многомерного статистического анализа. А именно, будет дано описание математических моделей и методов таких разделов математической статистики как корреляционный анализ, регрессионный анализ, дисперсионный анализ, дискриминантный анализ, кластерный анализ. Предложенные методы и алгоритмы иллюстрируются с помощью более - менее реальных примеров.

Наименование и краткое содержание разделов и тем дисциплины, форма промежуточной аттестации по дисциплине	Всего (часы)	В том числе								
		Контактная работа (работа во взаимодействии с преподавателем), часы						Самостоятельная работа учащегося, часы		
		из них						из них		
		Занятия лекционного типа	Занятия семинарского типа	Групповые консультации	Индивидуальные консультации	Учебные занятия, направленные на проведение текущего контроля успеваемости: коллоквиумы, практические контрольные занятия и др.	Всего	Чтение учебников и оригинальных журнальных статей	Подготовка рефератов и т.п..	Всего
Тема 1. Основные задачи и методы прикладного многомерного статистического анализа Обзор основных задач многомерного статистического анализа. Задача о наилучшей линейной оценке. Коэффициенты корреля-	14	4		-	-		4	10	-	10

ции и их свойства. Типы многомерных данных. Многомерное нормальное распределение и его свойства.										
Тема 2. Регрессионный анализ Классическая многомерная модель регрессии. Метод наименьших квадратов (МНК). Свойства оценок параметров. Проверка линейных гипотез. Коэффициент детерминации и его свойства. Выбор оптимальной модели. Модели с гетероскедастичностью и автокорреляцией ошибок и их оценивание.	34	12		-	-		12	22	-	22
Тема 3. Дисперсионный анализ. Однофакторный дисперсионный анализ. Двухфакторный дисперсионный анализ.	21	6		-	-		6	15	-	15
Тема 4. Дискриминантный анализ. Постановка задачи, опи-	21	6		-	-		6	15	-	15

сание общего метода решения задачи дискриминации. Решение задачи дискриминации для случая многомерного нормального распределения с известными параметрами. Решение задачи дискриминации для случая многомерного нормального распределения с неизвестными параметрами.										
Тема 5. Кластерный анализ. остановка задачи, описание общего метода решения задачи построения кластеров. Описание методов близости и различия кластеров. Описание алгоритма построения кластеров.	14	4	-		-		4	10	-	10
Промежуточная аттестация - зачет	4						4			
Итого	108	36						72		

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, критерии и шкалы оценивания

Возможные темы для домашних заданий (рефератов)

1. Основные задачи многомерного статистического анализа:
 - корреляционный анализ
 - регрессионный анализ
 - снижение размерности
 - дисперсионный анализ
 - дискриминантный анализ
 - кластерный анализ
2. Гильбертово пространство случайных величин .
3. Что такое наилучшая линейная оценка.
4. Лемма о перпендикуляре.
5. Простой коэффициент корреляции и что он измеряет.
6. Множественный коэффициент корреляции и что он измеряет.
7. Частный коэффициент корреляции и что он измеряет.
8. Множественная линейная регрессия: модель и основные ограничения.
9. Описание МНК для оценки параметров.
10. Явный вид оценок параметров по МНК.
11. Общая схема проверки гипотезы о параметре.
12. Для чего используется Т-критерий.
13. Основное различие Т-критерия и F-критерия в задаче проверки значимости влияния фактора.
14. Адекватность модели. Постановка задачи.
15. Коэффициент детерминации и что он измеряет.
16. Основная задача в однофакторном дисперсионном анализе.
17. Основная задача в двухфакторном дисперсионном анализе.
18. Основная задача дискриминантного анализа.
19. Кластерный анализ: постановка задачи.

20. Кластерный анализ: последовательное построение факторов.

Методические материалы для проведения процедур оценивания результатов обучения

Выполнение домашних заданий для текущего контроля успеваемости и задания промежуточного контроля оценивается по пяти-балльной системе. Итоговая оценка вычисляется как среднее арифметическое полученных результатов.

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

Список тем вопросов для зачета.

1. Основные задачи многомерного статистического анализа.
2. Гильбертово пространство случайных величин. Задача о наилучшей линейной оценке.
3. Корреляционный и регрессионный анализ.
4. Коэффициенты корреляции.
5. Простая линейная регрессия. Метод наименьших квадратов. Свойства оценок.
6. Множественная линейная регрессия. МНК. Свойства оценок.
7. Т-критерий значимости влияния фактора.
8. Проверка линейных гипотез. F-критерий.
9. Проверка адекватности модели. Коэффициент детерминации.
10. Равенство уравнений регрессии. Тест Чоу.
11. Фиктивные переменные.
12. Модель линейной регрессии с гетероскедастичностью.
13. Модель линейной регрессии с автокорреляцией в ошибках. Критерий Дарбина-Уотсона.
14. Однофакторный дисперсионный анализ.
15. Двухфакторный дисперсионный анализ.
16. Дискриминантный анализ: постановка задачи и ее решение в случае известных параметров.
17. Решение задачи дискриминантного анализа в случае неизвестных параметров.
18. Кластерный анализ: постановка задачи и основные понятия.

19. Кластерный анализ: схема последовательного построения факторов.

ШКАЛА И КРИТЕРИИ ОЦЕНИВАНИЯ результатов обучения (РО) по дисциплине (модулю)				
Оценка РО и соответствующие ви- ды оценочных средств	2 (не зачтено)	3 (зачтено)	4 (зачтено)	5 (зачтено)
Знания <i>Зачет</i>	Отсутствие знаний	Фрагментарные знания	Общие, но не структуриро- ванные знания	Сформированные систе- матические знания
Умения <i>Практические зада- ния</i>	Отсутствие умений	В целом успешное, но не си- стематическое умение	В целом успешное, но со- держащее отдельные пробелы умение (допускает не- точности непринципиально- го характера)	Успешное и системати- ческое умение
Навыки (владения, опыт дея- тельности) <i>Зачет, практические занятия</i>	Отсутствие навыков (владений, опыта)	Наличие отдельных навыков (наличие фрагментарного опыта)	В целом, сформированные навыки (владения), но ис- пользуемые не в активной форме	Сформированные навыки (владения), применяемые при решении задач

7. Ресурсное обеспечение:

7.1. Перечень основной и дополнительной литературы

Основная литература

1. Андерсон Т. Статистический анализ временных рядов / Андерсон Т. – М.: Книга по Требованию, 2012. – 752 с. ISBN 978-5-458-26140-1

Дополнительная литература

- 1) Факторный, дискриминантный и кластерный анализы. – М.: Финансы и статистика, 1989.

- 2) Шеффе Г. Дисперсионный анализ. – М.: Наука, 1980.
- 3) Болч К., Хуань К.Дж. Многомерные статистические методы для экономики. – М.: Статистика, 1979.
- 4) Hardle W., Simar L. Applied Multivariate Statistical Analysis. - <http://www.xplorestat.de/ebooks/ebooks.html>
- 5) Кулаичев А.П. Методы и средства комплексного анализа данных. – М.: Форум-Инфра-М, 2006.
- 6) Андерсон Т. Введение в многомерный статистический анализ. – М.: Наука, 1963.
- 7) Johnson R.A. and Winchern D.W. Applied Multivariate Statistical Analysis. – Pearson Prentice Hall, 2007.
- 8) Хохлов Ю.С. Эконометрика. Вводный курс: Учебное пособие. М.: Издательский отдел факультета ВМиК МГУ, 2006. – 100 с.

7.2. Перечень лицензионного программного обеспечения, в том числе отечественного производства При реализации дисциплины может быть использовано следующее программное обеспечение:

Программное обеспечение для подготовки слайдов лекций MS PowerPoint, MS Word
 Программное обеспечение для создания и просмотра pdf-документов Adobe Reader
 Издательская система LaTeX
 Язык программирования Python и среда разработки Jupiter Notebook (вместе с библиотеками numpy, scikit-learn, pandas)
 Язык программирования R и среда разработки R Studio

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.

Факультет ВМК, 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 факультета 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. Соответствие результатов обучения по данному элементу ОПОП результатам освоения ОПОП указано в Общей характеристике ОПОП.

9. Разработчик (разработчики) программы.

д.ф.- м.н., профессор Хохлов Юрий Степанович (yskhokhlov@yandex.ru)

10. Язык преподавания - русский.