Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Максимов Алексей Борисович
Должность: директор департамента по образовательной политике
Дата подписания: 22.05.2024 17:54:30
Уникальный программный ключ:
8db180d1a3f02ac9e60521a5672742735c18b1d6

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение
высшего образования
«МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»
Факультет информационных технологий

УТВЕРЖДАЮ
Декан факультета
«Информационные технологии»
_____ / Д.Г.Демидов /
«15» февраля 2024г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Математические методы анализа данных»

Направление подготовки/специальность
09.03.01 Информатика и вычислительная техника

Профиль/специализация
«Системная и программная инженерия»

Квалификация
бакалавр

Формы обучения
очная

Москва, 2024г.

Разработчик(и):

к.т.н., доцент / В.Ю. Верещагин /

Согласовано:
Заведующий кафедрой «Инфокогнитивные технологии»,



к.т.н., доцент / Е.А. Пухова /

Содержание

1 Цели, задачи и планируемые результаты обучения по дисциплине 4
2 Место дисциплины в структуре образовательной программы 5
3 Структура и содержание дисциплины 5
3.1 Виды учебной работы и трудоемкость для очной формы обучения 5
3.2 Тематический план изучения дисциплины для очной формы обучения 5
3.3 Содержание дисциплины 6
3.4 Тематика лабораторных занятий 6
4 Учебно-методическое и информационное обеспечение 7
4.1 Нормативные документы и ГОСТы 7
4.2 Основная литература 7
4.3 Дополнительная литература 7
4.4 Электронные образовательные ресурсы 7
4.5 Лицензионное и свободно распространяемое программное обеспечение 7
4.6 Современные профессиональные базы данных и информационные справочные системы 8
5 Материально-техническое обеспечение 8
6 Методические рекомендации 8
6.1 Методические рекомендации для преподавателя по организации обучения 8
6.2 Методические рекомендации для обучающихся по освоению дисциплины 8
7 Фонд оценочных средств 9
7.1 Методы контроля и оценивания результатов обучения 9
7.2 Шкала и критерии оценивания результатов обучения 10
7.3 Оценочные средства 11

1 Цели, задачи и планируемые результаты обучения по дисциплине

Целью дисциплины является углубление знаний прикладной математической статистики, расширение умений и навыков практического решения задач, возникающих в гуманитарных и социально-экономических науках, связанных с обработкой многомерных данных.

Задачи дисциплины: – освоение методологии разработки и реализации статистических методов обработки и анализа многомерных данных, – разработка и анализ эффективности статистических методов анализа многомерных данных, а также отработка навыков применения этой методологии в научных исследованиях и решении практических задач обработки данных и математического моделирования, возникающих в гуманитарных и социально-экономических науках; – формирование умения самостоятельно разбираться в имеющихся концепциях, методах и моделях прикладного статистического анализа данных и применять их для решения прикладных задач, проводить научные исследования в области разработки и применения методов математической статистики

Обучение по дисциплине «Математические методы анализа данных» направлено на формирование у обучающихся следующих компетенций:

Код и наименование компетенций	Индикаторы достижения компетенции
ОПК-8. Способен разрабатывать алгоритмы и программы, пригодные для практического применения	ИОПК-8.1. Знает основные технологии создания и внедрения информационных систем, стандарты управления жизненным циклом информационной системы. ИОПК-8.2. Умеет проектировать блок-схемы алгоритмов, оценивать производительность алгоритмов и затраты памяти на работу алгоритма, разрабатывать программы на основе спроектированного алгоритма и проводить отладку программы, применять методы системного анализа и математического моделирования при разработке и эксплуатации ИС, проводить структурный анализ, функциональный анализ, объектно-ориентированный анализ иерархии классов. осуществлять организационное обеспечение выполнения работ на всех стадиях и в процессах жизненного цикла информационной системы. ИОПК-8.3. Владеет навыками разработки программ, построения блок-схем алгоритмов и оценки производительности алгоритмов, работы с унифицированным языком визуального моделирования, составления плановой и отчетной документации по управлению проектами создания информационных систем на стадиях жизненного цикла.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к части, формируемой участниками образовательных отношений блока Б1 и междисциплинарно связана с дисциплинами: Технологии и методы программирования, Математический анализ, Теория вероятностей и математическая статистика, Методы распределённого хранения и параллельной обработки данных , Методы машинного обучения

3 Структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 4 зачетные единицы (144 часа).

3.1 Виды учебной работы и трудоемкость для очной формы обучения

№ п/п	Вид учебной работы	Количество часов	Семестры	
			5	
1	**Аудиторные занятия**	**54**	54	
	В том числе:			
1.1	Лекции	4	4	
1.2	Семинарские/практические занятия			
1.3	Лабораторные занятия	50	50	
2	**Самостоятельная работа**	**90**	90	
3	**Промежуточная аттестация**			
	Зачет/Диф. Зачет/экзамен		экзамен	
	Итого:	**144**	144	

3.2 Тематический план изучения дисциплины для очной формы обучения

№ п/п	Разделы/темы дисциплины	Трудоемкость, час					
		Всего	Аудиторная работа				Самостоятельная работа
			Лекции	Семинарские/ практические занятия	Лабораторные занятия	Практическая подготовка	
1	Распределение вероятностей случайных величин	22	1		6		15
2	Оценки параметров распределений вероятностей	22	1		6		15
3	Методы анализа законов распределения вероятностей случайных величин	17	1		6		10
4	Проверка гипотез о значениях параметров распределения	17	1		6		10
5	Методы исследования связей между случайными величинами	16			6		10

6	Статистические методы анализа многомерных данных	16			6		10
7	Оценивание плотности и функции вероятности	16			6		10
8	Рубежный контроль	18			8		10
Итого		**144**	**4**		**50**		**90**

3.3 Содержание дисциплины

Предварительные замечания вероятностного и статистического характера.

Дискретные распределения. Непрерывные распределения.

Общие замечания. Оценки типа максимального правдоподобия (M-оценки). Линейные комбинации порядковых статистик (L-оценки). Оценки, получаемые в ранговых критериях (R-оценки). Оценки метода минимального расстояния (R-оценки). Байесовское оценивание.

Робастность в статистике. Общие основы. Зачем нужны робастные процедуры. Качественная робастность. Количественная робастность. Инфинитеземальные аспекты. Общая идея вычисления робастных оценок. Робастность M-оценок. Параметрическая процедура отбраковки аномальных наблюдений

Общие критерии согласия

Специальные критерии согласия

Сравнение параметров распределения.

Непараметрические критерии однородности статистических данных. Критерии тренда и случайности. Толерантные пределы.

Дисперсионный анализ

Корреляционный анализ

Регрессионный анализ

Методы снижения размерности. Методы факторного анализа. Многомерное шкалирование.

Оценивание плотности и функции вероятности. Параметрические оценки плотности. Гистограммы и ядерные оценки плотности. Свойства ядерной оценки одномерной плотности Выбор ширины окна. Референтные эвристические правила. Подстановка. Кроссвалидация на основе наименьших квадратов. Кросс-валидация на основе правдоподобия. Бутстраповские методы. Частотная и ядерная оценка функции вероятности. Ядерное оценивание плотности в случае смешанных дискретных и непрерывных данных Построение доверительных интервалов Проклятие размерности. Оценивание условной плотности. Ядерное оценивание условной плотности. Ядерное оценивание функции условного распределения. Ядерное оценивание условной квантили. Модели бинарного выбора и счетных данных.

3.4 Тематика лабораторных занятий

Предварительные замечания об используемом математическом аппарате.

Многомерное нормальное распределение. Распределение линейной комбинации нормально распределенных величин; независимость величин; частные распределения. Условные распределения и множественный коэффициент корреляции. Характеристическая функция; моменты.

Исследование свойств оценок параметров распределений вероятностей по эмпирическим данным

Байесовское оценивание

Экспериментальное исследование робастности оценок

Критерии проверки отклонения распределения от нормального закона

Специальные критерии согласия
Подбор кривых распределения вероятностей по экспериментальным данным
Дисперсионный анализ
Корреляционный анализ
Алгоритмы восстановления регресии. МНК. Непараметрическая регрессия. Многомерная линейная регрессия. Нелинейные обобщения линейной регрессии. Логистическая регрессия
Методы снижения размерности. Методы факторного анализа.
Многомерное шкалирование.
Оценка плотности совместного распределения

4 Учебно-методическое и информационное обеспечение

4.1 Нормативные документы и ГОСТы

1. Федеральный закон от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации» (с изменениями и дополнениями);
2. Федеральный государственный образовательный стандарт высшего образования - бакалавриат по направлению подготовки 09.03.01 Информатика и вычислительная техника, утвержденный Приказом Министерства образования и науки РФ от 19 сентября 2017 г. N 929 "Об утверждении федерального… Редакция с изменениями N 1456 от 26.11.2020
3. Приказ Министерства образования и науки РФ от 05 апреля 2017 г. № 301 «Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры;
4. Порядок проведения государственной итоговой аттестации по образовательным программам высшего образования – программам бакалавриата, программам специалитета и программам магистратуры, утвержденный приказом Минобрнауки России от 29 июня 2015 г. № 636;
5. Положение о практической подготовке обучающихся, утвержденное приказом Министерства науки и высшего образования Российской Федерации и Министерства просвещения Российской Федерации от 5 августа 2020 г. № 885/390.

4.2 Основная литература

Замятин, А. В. Интеллектуальный анализ данных : учебное пособие / А. В. Замятин. — Томск : ТГУ, 2020. — 196 с. — ISBN 978-5-94621-898-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/202415

4.3 Дополнительная литература

Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/131721

4.4 Электронные образовательные ресурсы

https://online.mospolytech.ru/enrol/index.php?id=278

4.5 Лицензионное и свободно распространяемое программное обеспечение

1. Visual Studio Code
2. Браузеры Chrome, Edge, Firefox

3. OpenVPN с правами для запуска у студентов
4. FileZilla
5. PuTTy
6. Git
7. Python 3.10
8. Anaconda

4.6 Современные профессиональные базы данных и информационные справочные системы

1. https://www.elibrary.ru/
2. http://www.machinelearning.ru/

5 Материально-техническое обеспечение

Для проведения лабораторных работ и самостоятельной работы студентов подходят аудитории, оснащенные компьютерами с программным обеспечением в соответствии со списком в пункте 4.5 и подключенные к интернету.

Число рабочих мест в аудитории должно быть достаточным для обеспечения индивидуальной работы студентов.

Рабочее место преподавателя должно быть оснащено компьютером с подключенным к нему проектором или иным аналогичным по функциональному назначению оборудованием.

6 Методические рекомендации

6.1 Методические рекомендации для преподавателя по организации обучения

1. При подготовке к занятиям следует предварительно проработать материал занятия, предусмотрев его подачу точно в отведенное для этого время занятия. Следует подготовить необходимые материалы – теоретические сведения, задачи и др. При проведении занятия следует контролировать подачу материала и решение заданий с учетом учебного времени, отведенного для занятия.

2. При проверке работ и отчетов следует учитывать не только правильность выполнения заданий, но и оптимальность выбранных методов решения, правильность выполнения всех его шагов.

6.2 Методические рекомендации для обучающихся по освоению дисциплины

Изучение дисциплины осуществляется в строгом соответствии с целевой установкой в тесной взаимосвязи учебным планом. Основой теоретической подготовки студентов являются лекции и самостоятельная работа.

В процессе самостоятельной работы студенты закрепляют и углубляют знания, полученные во время аудиторных занятий, готовятся к промежуточной аттестации, а также самостоятельно изучают отдельные темы учебной программы.

На занятиях студентов, в том числе предполагающих практическую деятельность, осуществляется закрепление полученных, в том числе и в процессе самостоятельной работы, знаний. Особое внимание обращается на развитие умений и навыков установления связи положений теории с профессиональной деятельностью будущего специалиста.

Самостоятельная работа осуществляется индивидуально. Контроль самостоятельной работы организуется в двух формах:

- самоконтроль и самооценка студента;
- контроль со стороны преподавателей (текущий и промежуточный).

Текущий контроль осуществляется на аудиторных занятиях.

Критериями оценки результатов самостоятельной работы студента являются:

- уровень освоения студентом учебного материала;
- умения студента использовать теоретические знания при выполнении практических задач;
- сформированность компетенций;
- оформление материала в соответствии с требованиями.

Приветствуется обсуждение самих заданий с другими студентами: можно как давать, так и получать советы по общей стратегии выполнения и изучения материала, давать и получать помощь в отладке. Однако писать код студент должен самостоятельно. Делиться кодом или писать его совместно запрещено.

7 Фонд оценочных средств

7.1 Методы контроля и оценивания результатов обучения

Приведенные ниже правила выставления оценок и опозданий могут быть изменены, если преподаватель сочтет это необходимым. Важно, чтобы студенты регулярно просматривали план курса, выложенный в СДО, на предмет его обновления или изменения.

Достижение компетенций оценивается с помощью лабораторных работ и рубежных контролей. Индикаторы ИОПК-8.1, ИОПК-8.2, ИОПК-8.3 заложены в темах 1-7.

В соответствии с планом на дисциплины студентам выдаются задания на лабораторные работы. Помимо требований и описания функционала в работе указан крайний срок сдачи. Для сдачи лабораторной работы студенту необходимо прислать ссылку на репозиторий в GitHub Classroom и на хостинг, где размещен результат работы с реализованным функционалом, описанным в задании. Работа считается сданной если в ней реализовано 80% и более требований и функционала, описанного в задании.

Каждый студент имеет право на 6 дней опоздания, которые могут быть потрачены на любые задания в течение семестра. Опоздания предназначены для решения особых ситуаций, таких как болезнь или чрезвычайные семейные обстоятельства.

Когда использованы все дни опоздания за каждый день просрочки начисляется штраф в размере 25% от максимального результата за задание. Задания, присланные позже, чем 4 дня, не будут оцениваться. В связи с зависимостью между работами студентам может потребоваться все равно выполнить предыдущие работы, даже если они не оцениваются.

После сдачи лабораторной работы студент должен ее защитить. Во время защиты лабораторной работы преподаватель проверяет репозиторий, хостинг и выполнение критериев и требований задания, а студент отвечает на вопросы преподавателя по его коду, а также теоретических вопросов, приведенных после текста задания лабораторной работы. Если студент отказывается отвечать на вопросы, или дает полностью неверные ответы, или ответы не по теме, то работа может считаться сданной, но при этом она не оценивается.

Работа должна быть выполнена студентом самостоятельно: в репозитории в системе контроля версий студента содержатся коммиты только за его авторством, по этим коммитам можно проследить как велась работа, студент может объяснить свой код и ход выполнения работы, если эти правила не соблюдаются, то работа не считается сданной и не оценивается.

Рубежные контроли пишутся в аудитории индивидуально по варианту задания, выданному преподавателем в назначенные дни. При отсутствии студента в день написания контрольной работы ему дается еще один шанс ее написать на последнем занятии в семестре, но обязательно очно.

Студенты должны заранее сообщать о том, что у них могут возникнуть трудности со своевременной сдачей задания или проекта. При наличии реальных причин задержки студентам следует как можно скорее связаться с преподавателем и обсудить возможные условия.

7.2 Шкала и критерии оценивания результатов обучения

Лабораторная работа оценивается в процентах степени выполнения следующих критериев и для выставления оценки суммируются проценты за каждый из четырех критериев:

1. Полнота выполнения практического задания (30%): соответствует ли функциональность заданным требованиям и целям, насколько точно и без ошибок код выполняет поставленные задачи, насколько эффективно задание отвечает требованиям целевой аудитории и обеспечивает приятное восприятие.

2. Качество и структура кода (10%): качество, читаемость и организация кода, рациональность выполнения задания, последовательность именования и соблюдение лучших практик.

3. Творчество и инновации (10%): творческий подход студентов к выполнению заданий, насколько студенты вышли за рамки основных требований и реализовали дополнительные возможности или использовали уникальные решения.

4. Ответы на вопросы по коду студента и теории (50%):

Дает краткий ответ, содержащий ошибки или неточности. На наводящие вопросы отвечает неправильно (10% из 50%)

Дает развернутый ответ, содержащий ошибки или неточности. На наводящие вопросы отвечает неверно (20% из 50%)

Дает развернутый ответ, содержащий ошибки или неточности. На наводящие вопросы отвечает правильно (30% из 50%)

Дает правильные и развернутые ответы на вопросы (50% из 50%).

R лабораторные рассчитывается как среднее результатов за все лабораторные работы. За полное и безошибочное выполнение всех лабораторных работ в срок и их защиту можно получить максимум 100 баллов (R лабораторные).

Рубежный контроль оценивается по следующим критериям:

Полнота выполнения практического задания: соответствует ли функциональность заданным требованиям и целям, насколько точно и без ошибок код выполняет поставленные задачи.

Качество и структура кода: качество, читаемость и организация кода, рациональность выполнения задания, последовательность именования и соблюдение лучших практик.

Творчество и инновации: творческий подход студентов к выполнению заданий, насколько студенты вышли за рамки основных требований и реализовали дополнительные возможности или использовали уникальные решения.

Пользовательский опыт: отзывчивость, доступность, насколько эффективно задание отвечает требованиям целевой аудитории и обеспечивает приятное восприятие.

Самостоятельность решения: в репозитории студента есть коммиты только за его авторством, по коммитам в репозитории можно проследить как велась работа, студент может объяснить свой код и ход выполнения работы, если эти правила не соблюдаются, то работа не считается сданной.

Более подробное описание критериев дается в тексте задания рубежного контроля.

За полностью выполненные рубежные контроли также можно получить 100 баллов (R контроль).

Также имеется коэффициент сданных работ K сданные, который равен 1 если все работы сданы и 0 если хотя бы одна работа не сдана.

Итоговый балл рассчитывается по формуле: R сем = (0,5 x R лабораторные + 0,5 x R контроль) * K сданные.

Итоговый балл пересчитывается по шкале ниже и на основании полученной оценки фиксируется результат промежуточной аттестации.

Соответствие баллов в 100 балльной рейтинговой системе оценке по 4-бальной шкале:

0-54 - неудовлетворительно

55-69 - удовлетворительно
70-84 - хорошо
85-100 – отлично

7.3 Оценочные средства

7.3.1 Текущий контроль

ТИПОВОЕ ЗАДАНИЕ Для каждого варианта требуется: 1. Представить опытные данные в сгруппированном виде, разбив на k равноотстоящих частичных интервалов. 2. Найти эмпирическую функцию распределения и построить ее график. 3. Построить полигон и гистограмму относительных частот. 4. Вычислить методом произведений числовые характеристики выборки: выборочную среднюю, выборочную и исправленную дисперсии, выборочное среднее квадратическое отклонение, коэффициенты асимметрии и эксцесса. 5. Найти точечные оценки параметров нормального закона распределения и плотность вероятностей f(x). 6. Проверить, согласуется ли принимаемая гипотеза о нормальном распределении генеральной совокупности с эмпирическим распределением выборки, используя критерии Пирсона и Колмогорова (при уровнях значимости 0,05; 0,01).

7.3.2 Промежуточная аттестация

КОНТРОЛЬНЫЕ ВОПРОСЫ (к экзамену) 1. Виды случайных событий. 2. Классическое и статистическое определения вероятности появления события. Основные формулы комбинаторики. 3. Принцип статистической устойчивости относительных частот. Связь и различие между классическим и статистическим определениями вероятности. 4. Геометрическое определение вероятности. Задача Бюффона. 5. Алгебра событий. Понятия суммы и произведения событий, их геометрическая интерпретация. Основные законы алгебры событий. 6. Теорема сложения вероятностей для несовместных событий. Следствия из нее. 7. Теорема сложения вероятностей для совместных событий. 8. Теорема умножения вероятностей для зависимых и независимых событий. Понятие условной вероятности. 9. Теорема о вероятности появления хотя бы одного события. 10. Формула полной вероятности. 11. Формула Бернулли. 12. Локальная и интегральная теоремы Лапласа. 13. Определение и типы случайных величин. Понятие закона распределения случайной величины. Ряд распределения. 14. Интегральная функция распределения вероятностей. Определение, вероятностный смысл и свойства. 15. Плотность вероятностей. Определение, вероятностный смысл и свойства. 16. Связь между интегральной функцией распределения вероятностей и плотностью вероятностей. 17. Определение, вероятностный смысл и свойства математического ожидания для дискретных и непрерывных случайных величин. 18. Определение, вероятностный смысл и свойства дисперсии. 19. Биномиальный закон распределения. 20. Среднее и наивероятнейшее числа появлений события при биномиальном распределении. 21. Закон распределения Пуассона. 22. Равномерный закон распределения вероятностей. 23. показательный закон распределения вероятностей. 24. Нормальный закон распределения вероятностей. Вероятность попадания нормально распределенной случайной величины на произвольный конечный интервал. 25. Вероятность попадания нормально распределенной случайной величины на интервал, симметричный относительно среднего значения. Правило трех сигм. Теорема Ляпунова. 26. Предельные теоремы теории вероятностей. 27. Предмет и основные задачи математической статистики. 28. Понятие о выборочном методе. Генеральная и выборочная совокупности. 29. Основные понятия математической статистики (вариационный ряд, частота, относительная частота,

статистическое распределение выборки). 30. Эмпирическая функция распределения выборки и ее свойства. 31. Полигон частот и полигон относительных частот. 32. Гистограмма частот и относительных частот. 33. Точечные оценки параметров распределения. Требования к оценкам. 34. Выборочная средняя. Свойство устойчивости выборочных средних. 35. Выборочная и исправленная дисперсии. 36. Интервальные оценки. Доверительный интервал и доверительная вероятность. Понятие точности оценки. 37. Построение доверительного интервала для оценки генеральной средней при известном среднем квадратическом отклонении. 38. Построение доверительного интервала для оценки генеральной средней при не известном среднем квадратическом отклонении. Случай малой выборки. Распределение Стьюдента. 39. Упрощенные методы расчета статистических характеристик выборки. Условные варианты. Метод произведений. 40. Критерии согласия. Уровень значимости. 41. Критерий Пирсона. 42. Корреляционная и регрессионная зависимости. 43. Уравнение выборочной регрессии. 44. Выборочный коэффициент регрессии. 45. Выборочный коэффициент корреляции. 46. Связь между выборочными коэффициентами регрессии корреляции.