

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Поверинов Игорь Евгеньевич

Должность: Проректор по учебной работе

Дата подписания: 08.07.2024 14:36:22

Уникальный программный ключ:

6d465b936eef331cede482bdeedd12ab9821b052f016465d53b72a2eab0de1b2

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Декан инженерно – экономического
факультета

« ____ » _____ 2021 г.

**МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**

Б1.В.02 Анализ многомерных данных

Профиль подготовки

«Искусственный интеллект и анализ больших данных в банковской сфере»

Квалификация выпускника

Магистр

Формы обучения

очная

г. Ульяновск, 2021

1. Методические рекомендации по изучению дисциплин

До начала изучения дисциплины «Анализ многомерных данных» студентам необходимо ознакомиться с содержанием рабочей программы изучаемой дисциплины (далее - РПД), а именно: с целями и задачами дисциплины, ее связями с другими дисциплинами образовательной программы, с основной и дополнительной литературой, с видами самостоятельной работы, с требованиями к промежуточной аттестации по дисциплине.

Необходимо проверить наличие индивидуального доступа к материалам дисциплины в системе дистанционного обучения (далее СДО). Изучить находящиеся в доступе материалы, объявления и требования.

Дисциплина предполагает лекционные и лабораторные занятия студентов. Для успешного усвоения теоретического материала и приобретения практических навыков необходимо регулярно посещать лекции (вне зависимости от формата их проведения), активно работать на лабораторных занятиях, перечитывать лекционный материал, значительное внимание уделять самостоятельному изучению дисциплины. Поэтому, важным условием успешного освоения дисциплины обучающимися является создание оптимальной системы тайм-менеджмента, позволяющей распределить учебную нагрузку равномерно в соответствии с графиком образовательного процесса. Большую помощь в этом может оказать составление плана работы на семестр, месяц, неделю, день. Все задания к лабораторным занятиям, а также задания, вынесенные на самостоятельную работу, рекомендуется выполнять непосредственно после изучения соответствующей темы лекционного курса. Это способствует лучшему усвоению материала, позволяет своевременно выявить и устранить «пробелы» в знаниях, систематизировать ранее пройденный материал, на его основе приступить к овладению новыми знаниями и навыками.

2. Методические рекомендации по подготовке к лекциям

Наиболее важная часть теоретического материала рассматривается в рамках лекций. Однако, небольшое количество часов, выделенное учебным планом на лекционные занятия, обосновывает необходимость самостоятельного изучения теоретических аспектов дисциплины.

В ходе лекций преподаватель излагает и разъясняет основные, наиболее сложные понятия в рамках тем, предусмотренных РПД, а также связанные с ней теоретические и практические проблемы, дает рекомендации на практические занятия, вместе с тем, четко формулирует и указывает на самостоятельную работу. Лекционные занятия преимущественно носят интерактивный характер в форме «вопрос-ответ» или мини дискуссии. Вопросы для мини дискуссий могут быть сформулированы преподавателем заранее, тогда студенты готовят свои аргументы еще до лекционного занятия. Дискуссии могут возникать и в процессе лекции. А так же студенты

сами могут готовить к конкретной теме лекционного занятия свои вопросы, имеющие для них практический и (или) научный интерес.

Для подготовки к лекциям (или в рамках самостоятельной работы после изучения темы на лекционном занятии) преподаватель может рекомендовать изучения учебной, научной литературы, статей, или официальных сайтов. При работе с данными источниками обучающимся рекомендуется вести конспект, для обеспечения систематизации материала.

В СДО студентам доступны презентационные материалы по темам лекционных занятий в рамках дисциплины. Для оптимизации процесса обучения, студенты могут пользоваться этими презентациями на лекционных занятиях, оставляя собственные пометки (комментарии, уточнения), которые помогут им в последующем подготовиться к практическим занятиям и промежуточной аттестации.

Студентам целесообразно разработать собственную систему сокращений, аббревиатур и символов (в дополнение к общепринятым сокращениям). Однако при дальнейшей работе с конспектом символы лучше заменить обычными словами для быстрого зрительного восприятия текста. Работая над конспектом лекций, всегда необходимо использовать не только учебник, но и ту литературу, которую дополнительно рекомендовал лектор. Именно такая серьезная, кропотливая работа с лекционным материалом позволит глубоко овладеть теоретическими знаниями.

3. Методические рекомендации по подготовке к лабораторным занятиям

Лабораторные занятия служат для закрепления изученного материала, развития умений, навыков и защиты выдвигаемых положений, приобретения практических навыков решения задач, разбора кейсов, а также для контроля преподавателем степени подготовленности студентов по изучаемой дисциплине. Лабораторное занятие – это занятие, проводимое под руководством преподавателя в учебной аудитории (или онлайн), направленное на углубление научно – теоретических знаний и овладение определенными методами самостоятельной, индивидуальной и групповой работы. В процессе таких занятий вырабатываются практические умения. Перед лабораторным занятием следует изучить конспект лекции и рекомендованную преподавателем литературу, обращая внимание на практическое применение теории и на методику решения типовых задач. На лабораторном занятии главное – уяснить связь решаемых задач с теоретическими положениями. Логическая связь лекций и лабораторных занятий заключается в том, что информация, полученная на лекции, в процессе самостоятельной работы на практическом занятии осмысливается и перерабатывается, при помощи преподавателя анализируется до мельчайших подробностей, после чего прочно усваивается.

В рамках лабораторных занятий предусмотрены собеседования.

Процедура выполнения лабораторных работ и собеседования по ним:

Количество проводимых лабораторных работ в течение всего периода освоения дисциплины	8 работ
Формат представления результатов	Бумажный отчет
Общее количество вопросов для собеседования	4-5 вопросов
Количество основных задаваемых при собеседовании вопросов	3-4 вопроса
Методические рекомендации (при необходимости)	Размещаются в электронной образовательной среде ВУЗа

Примерный перечень вопросов при собеседованиях по лабораторным работам

- 1) Назовите основные критерии классификации данных.
- 2) В чем состоит различие пространственных, временных и пространственно-временных данных?
- 3) Назовите основные типы переменных в зависимости от шкалы измерения.
- 4) Назовите основные числовые характеристики центра группирования количественных данных.
- 5) Какие показатели вариации количественных данных Вы можете назвать? В чем состоят различия между ними?
- 6) Что понимается под точечной и интервальной оценками?
- 7) Какие законы распределения выборочных характеристик используются для получения интервальных оценок?
- 8) Какая зависимость называется корреляционной?
- 9) Какие коэффициенты корреляции Вы знаете? В чем заключается различие между ними?
- 10) Какова цель расчета матрицы коэффициентов парной корреляции?
- 11) Как интерпретируется множественный коэффициент детерминации?
- 12) Какую информацию можно получить, анализируя корреляционное поле?
- 13) Что представляет собой каноническая корреляция и в чем состоит ее отличие от линейных коэффициентов корреляции?
- 14) На какой идее основан метод наименьших квадратов и для какой цели он используется?
- 15) Как интерпретируется значение коэффициента регрессии?
- 16) Перечислите свойства оценок случайных возмущений.
- 17) Каковы свойства МНК-оценок в классической модели линейной регрессии?
- 18) Что характеризуют элементы ковариационной матрицы?
- 19) Какая идея лежит в основе расчета коэффициента множественной детерминации в регрессионном анализе?
- 20) Для решения каких задач регрессионного анализа требуется условие нормальности регрессионных остатков?
- 21) Какова цель проведения компонентного анализа?
- 22) Опишите модель метода главных компонент.

- 23) Что представляют собой собственные векторы и собственные значения корреляционной матрицы, и как они могут быть использованы для получения матрицы весовых коэффициентов?
- 24) Дайте определение квадратичных форм и главных компонент. Укажите главные компоненты для двумерного, трехмерного и конечномерного пространства.
- 25) Как получают и для чего используют матрицы индивидуальных значений главных компонент?
- 26) Каковы свойства ортогональной матрицы собственных векторов в модели метода главных компонент?
- 27) В чем сущность регрессии на главные компоненты?
- 28) В чем состоит основная проблема применимости метода регрессии на главные компоненты?
- 29) Как определить относительный вклад m первых главных компонент в суммарную дисперсию?
- 30) Сколько главных компонент (факторов) следует выделять при снижении размерности признакового пространства?
- 31) Как проинтерпретировать выделенные главные компоненты (факторы)?
- 32) С какой целью проводится кластерный анализ?
- 33) Назовите способы представления информации для проведения кластерного анализа.
- 34) Как можно рассчитать расстояние между объектами (признаками)?
- 35) В каких случаях может быть использовано обычное евклидово расстояние?
- 36) Как можно рассчитать расстояние между объектами, если они представлены дихотомическими признаками?
- 37) Что представляет собой функционал качества разбиения, какова цель его использования?
- 38) Какие иерархические кластер-процедуры Вы знаете?
- 39) В чем состоит различие агломеративных и дивизимных кластер-процедур?
- 40) Что представляет собой дендрограмма, с какой целью она используется?
- 41) В каких случаях требуется нахождение робастных оценок параметров?
- 42) В чем состоит основной недостаток непосредственного применения правила «трех сигм» для выявления аномальных наблюдений?
- 43) Как выявить аномальные наблюдения с помощью критерия Диксона?
- 44) В каком случае критерий Титъена — Мура эквивалентен критерию Граббса при выявлении аномальных наблюдений?
- 45) Для чего производят цензурирование выборки?
- 46) Как получить робастные оценки параметров линейной регрессионной модели?

Шкала оценивания с учетом срока сдачи

Критерии оценивания	Балл
Студент правильно выполнил задание работы, продемонстрировал знания теоретического и практического материала	Отлично
Студент правильно выполнил задание работы, продемонстрировал неполные знания теоретического и практического материала	Хорошо
Студент выполнил задание работы, но допустил значительные неточности при выполнении, продемонстрировал неполные знания теоретического и практического материала	Удовлетворительно
Студент неправильно выполнил задание работы, не продемонстрировал знания теоретического и практического материала	Неудовлетворительно

4. Методические рекомендации по самостоятельной работе

Целью самостоятельной работы является освоение фундаментальных знаний, прикладных навыков, развитие ответственности и организованности, умений самостоятельно работать с учебным материалом и приобретение навыков поиска и реферирования доступной научной информации. Самостоятельная работа во внеаудиторное время подразумевает:

- повторение лекционного материала;
- изучения учебной и научной литературы; – изучения нормативных правовых актов (в том числе в электронных базах данных);
- проведение самоконтроля путем ответов на вопросы текущего контроля знаний, решения задач, кейсов;
- подготовку к докладам на практических занятиях;
- подготовку к прохождению промежуточной аттестации.

Самостоятельная работа оценивается в рамках:

- лекционных занятий (при наличии предварительных заданий по теме);
- лабораторных занятий путем собеседования;
- выполнения контрольных работ;
- промежуточной аттестации (ответы на вопросы).

В рамках самостоятельной работы обучающиеся могут использовать следующие ресурсы:

1. Образовательный портал УлГТУ <https://virtual.ulstu.ru/>
2. Научная библиотека УлГТУ <http://lib.ulstu.ru/>
3. ЭБС Юрайт <https://urait.ru/>

4. Росстат <https://rosstat.gov.ru/>
5. База данных «Статистические показатели банковского сектора РФ» ЦБ РФ https://cbr.ru/statistics/bank_sector/review/
6. База рецензируемых журнальных статей и глав книг (включая материалы открытого доступа) <https://www.sciencedirect.com/>
7. База данных Научной электронной библиотеки eLIBRARY.RU <https://www.elibrary.ru/defaultx.asp>

5. Рекомендации по работе с научной и учебной литературой, другими источниками информации

Работа с учебной и научной литературой является главной формой самостоятельной работы и необходима при подготовке к мини дискуссиям на лекционных занятиях, опросу, участию в обсуждении на практических занятиях, тестированию, докладам, промежуточной аттестации. Она включает проработку лекционного материала – изучение рекомендованных источников и литературы по тематике лекций.

В процессе работы с учебной и научной литературой студент может:

- делать записи по ходу чтения в виде простого или развернутого плана (создавать перечень основных вопросов, рассмотренных в источнике);
- составлять тезисы (цитирование наиболее важных мест статьи или монографии, короткое изложение основных мыслей автора);
- готовить аннотации (краткое обобщение основных вопросов работы);
- создавать конспекты (развернутые тезисы).

Выбрав нужный источник, следует найти интересующий раздел по оглавлению или алфавитному указателю, а также одноименный раздел конспекта лекций или учебного пособия. В случае возникших затруднений в понимании учебного материала следует обратиться к другим источникам, где изложение может оказаться более доступным. Необходимо отметить, что работа с литературой не только полезна как средство более глубокого изучения любой дисциплины, но и является неотъемлемой частью профессиональной деятельности будущего выпускника.

Обучающийся может использовать любую другую литературу, кроме рекомендованной в РПД, размещенную в доступных электронных библиотечных системах (далее ЭБС) и или библиотеке вуза. Высокий динамизм пополнения ЭБС приводит к тому, что в течение семестра изучения дисциплины могут появляться очень интересные с теоретической и практической точки зрения новые издания, которые не были перечислены в РПД.

6. Методические рекомендации к промежуточной аттестации по дисциплине

Главная задача промежуточной аттестации по дисциплине – проверка качества усвоения содержания дисциплины, уровня сформированности компетенций, заявленных в РПД.

Курсовая работа

1. Процедура проведения
Этапы проведения КП (КР) с указанием сроков выполнения:
 - 1) Выдача задания (3 неделя семестра).
 - 2) Ознакомление с темой работы и заданием (4 неделя семестра).
 - 3) Изучение теории (5-6 неделя семестра).
 - 4) Проведение расчетов, получение выводов (12-13 неделя семестра).
 - 5) Оформление пояснительной записки (14-15 неделя семестра).

2. Шкала оценивания с учетом срока сдачи¹

Критерии оценивания уровня сформированности компетенций	Балл
Выставляется при выполнении работы в полном объеме; оформлении с соблюдением установленных правил; студент свободно владеет теоретическим материалом, безошибочно применяет его при практическом исследовании; применении адекватных методов и алгоритмов для выполнения задания; на все вопросы дает правильные и обоснованные ответы, убедительно защищает свою точку зрения.	Отлично
Выставляется при выполнении работы в полном объеме; оформлении с соблюдением установленных правил; студент твердо владеет теоретическим материалом, может применять его при практическом исследовании; использованы правильные методы и алгоритмы для выполнения задания работы с несущественными неточностями; на большинство вопросов даны правильные ответы, защищает свою точку зрения достаточно обоснованно.	Хорошо
Выставляется при выполнении работы в полном объеме, работа оформлена с соблюдением установленных правил; при выполнении работы без достаточно глубокой проработки вопросов применены правильные методы и алгоритмы для выполнения задания с существенными неточностями; на вопросы отвечает неуверенно или допускает ошибки, неуверенно защищает свою точку зрения.	Удовлетворительно
Выставляется, когда студент не может защитить свои решения, допускает грубые ошибки и искажение фактов при ответах на поставленные вопросы или не отвечает на них.	Неудовлетворительно

3. Варианты (темы) курсовой работы

Тематика курсовых работ по дисциплине определяется преподавателем кафедры. При этом выбор основывается как на государственном стандарте, так и на направлениях научно-исследовательской и учебно-методической работы,

¹ За несвоевременную сдачу обучающемуся могут быть начислены штрафные баллы.

актуальных направлениях работы других организаций, деятельность которых связана с разработкой информационного и программного обеспечения.

При разработке тематики возможно предложение решения следующих задач:

- 1) Изучить теоретические основы анализа многомерных данных и составить реферат.
- 2) Произвести сбор статистических данных по актуальной проблеме, определить адекватные методы их анализа (включая методы, изученные в курсе) и применить их.
- 3) По всем пунктам сделать выводы и дать практические рекомендации.

Экзамен

1. Процедура проведения

Общее количество вопросов к экзамену (зачету с оценкой)	16 вопросов
Количество вопросов в билете	2 вопросов
Наличие задач в билете	1 задача
Формат проведения	Устно и письменно

Каждый обучающийся получает билет, содержащий 2 теоретических вопроса и задачу. На подготовку предоставляется 40 минут. Далее беседа с преподавателем, принимающим экзамен. возможны дополнительные вопросы преподавателя как по вопросам экзаменационного билета, так и по всему курсу в целом.

2. Шкала оценивания с учетом текущего контроля работы обучающегося в семестре

Критерии оценки уровня сформированности компетенций по дисциплине	Балл
Обучающийся показал глубокие знания теоретического материала по поставленному вопросу, грамотно, логично и стройно его изложил, правильно решена задача	Отлично
Обучающийся твердо знает материал, грамотно его излагает, но допустил несущественные неточности в ответе на вопрос / при решении задачи	Хорошо
Обучающийся показал знание только основных положений по поставленному вопросу, требует в отдельных случаях наводящих вопросов для принятия правильного решения, допустил отдельные неточности в ответе на вопрос / при решении задачи	Удовлетворительно
Обучающийся допустил грубые ошибки в ответе на поставленный вопрос / при решении задачи	Неудовлетворительно

3. Примерный перечень вопросов к экзамену

- 1) Основные критерии классификации наборов данных и виды классификации.
- 2) Основные числовые характеристики количественных данных.
- 3) Корреляционный анализ взаимосвязи количественных и качественных признаков.
- 4) Методы оценивания параметров регрессии. Оценивание параметров двумерного и множественного уравнения регрессии методом наименьших квадратов.
- 5) Свойства оценок параметров регрессии, полученных методом наименьших квадратов.
- 6) Регрессионные модели с переменной структурой.
- 7) Основные понятия и задачи снижения размерности признакового пространства.
- 8) Метод главных компонент.
- 9) Факторный анализ.
- 10) Многомерное шкалирование.
- 11) Кластерный анализ.
- 12) Дискриминантный анализ.
- 13) Метод канонических корреляций.
- 14) Методы выявления аномальных значений в выборочных данных.
- 15) Методы устойчивого параметрического оценивания.
- 16) Основные подходы к оцениванию параметров при отсутствии информации о законе распределения значений признаков.

4. Примерные варианты задач к экзамену

1. По выборке из 10 наблюдений роста (в см) отцов (X) и их взрослых сыновей (Y), представленной в таблице, найти выборочный коэффициент корреляции и проверить его значимость при $\alpha = 0,05$ (распределение случайных величин X и Y предполагается нормальным):

x_i	180	172	173	169	175	170	179	170	167	174
y_i	186	180	176	171	182	166	182	172	169	177

2. Две группы экспертов проанализировали 12 проектов с точки зрения их эффективности:

Группы экспертов	Номера проектов и присвоенные им ранги											
	№1	№2	№3	№4	№5	№6	№7	№8	№9	№10	№11	№12
Первая	1	3	4	2	5	6	12	7	8	9	10	11
Вторая	2	3	1	4	6	5	9	7	8	10	12	11

Рассчитайте коэффициент Спирмена и сделайте вывод о согласованности мнений экспертов.

3. По данным 15 предприятий отрасли рассчитан коэффициент корреляции $r_{x_1x_2} = -0,7$ между себестоимостью продукции $x^{(1)}$ и производительностью труда $x^{(2)}$. Требуется:

- а) построить матрицу собственных значений и матрицу факторных нагрузок;
- б) определить дисперсию и относительный вклад первой главной компоненты в обобщенную дисперсию исходных признаков.

4. По выборке из 10 наблюдений роста (в см) отцов (X) и их взрослых сыновей (Y), представленной в таблице, рассчитать оценки параметров линейной регрессии, описывающей зависимость роста сыновей от роста отцов:

x_i	180	172	173	169	175	170	179	170	167	174
y_i	186	180	176	171	182	166	182	172	169	177

Дать интерпретацию полученным результатам.

5. Деятельность 10 промышленных предприятий региона охарактеризована двумя показателями $x^{(1)}$ — общие затраты на руль товарной продукции и $x^{(2)}$ — фондоотдача (см. табл.):

№ предприятия / показатели	1	2	3	4	5	6	7	8	9	10
$x^{(1)}$	0,92	0,72	0,83	0,81	0,82	0,93	0,84	0,89	0,89	0,95
$x^{(2)}$	0,51	0,59	1,03	1,21	0,63	0,68	0,57	1,52	1,04	0,99

Требуется:

- проранжировать предприятия по первой главной компоненте;
- отобразить объекты (предприятия) в пространстве двух главных компонент, отложив по оси ординат значения первого главного фактора, а по оси абсцисс — второго;
- дать интерпретацию полученным результатам.

6. Дана информация по шести объектам, каждый из которых характеризуется двумя признаками:

Номер объекта (i)	1	2	3	4	5	6
x_{i1}	5	6	5	10	11	10
x_{i2}	10	12	13	9	9	7

Проведите классификацию:

- используя обычное евклидово расстояние и метод «ближнего соседа»;
- используя обычное евклидово расстояние и метод «дальнего соседа»;
- используя обычное евклидово расстояние и метод «средней связи».

7. По иерархическому агломеративному алгоритму проведите классификацию четырех хозяйств, работа которых характеризуется показателями объема реализованной продукции $x^{(1)}$ — растениеводства и $x^{(2)}$ — животноводства (млн руб./га):

Номер хозяйства (i)	1	2	3	4
x_{i1}	1	7	1	9
x_{i2}	5	9	3	7

Постройте дендрограмму. Для расчета расстояния между объектами используйте взвешенное евклидово расстояние с весами $w_1 = 0,1$; $w_2 = 0,9$, а расстояние между кластерами измеряйте по принципу «дальнего соседа».

8. Данные о деятельности четырех фирм по показателям $x^{(1)}$ и $x^{(2)}$ представлены в таблице:

Номер фирмы (i)	1	2	3	4
x_{i1}	2	2	8	10
x_{i2}	3	5	9	7

Рассчитайте:

- а) расстояние между вторым и четвертым объектами ($d_E(2,4)$), исходя из обычного евклидова расстояния;
- б) расстояние между первым и третьим объектами ($d_{BE}(1,3)$), исходя из взвешенного евклидова расстояния с весами $w_1 = 0,8$; $w_2 = 0,2$.

9. Расстояния между тремя объектами равны $d_{1,2} = 7,8$; $d_{1,3} = 6,2$; $d_{2,3} = 2,4$. Постройте матрицу расстояний.

10. Дана информация о шести регионах по двум показателям:

Номер региона (i)	1	2	3	4	5	6
x_{i1}	35	31	32	36	30	34
x_{i2}	126	112	123	128	115	123

При решении задачи классификации регионов использован метод k -средних. Координаты первого и второго регионов были взяты в качестве центров группирования. В результате проведения первой итерации алгоритма регионы были разделены на группы следующим образом: I группа — №1, №3, №4, №6; II группа — №2, №5. Проведите вторую и (при необходимости) последующие итерации алгоритма для получения окончательного ответа данной задачи.