

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Курганский государственный университет»
(КГУ)

Кафедра «Программное обеспечение автоматизированных систем»

УТВЕРЖДАЮ:



Ректор
Н. В. Дубив
2020 г.

Рабочая программа учебной дисциплины

Методы и алгоритмы анализа данных

образовательной программы высшего образования –
программы бакалавриата

09.03.04 – Программная инженерия
Направленность:

Программное обеспечение автоматизированных систем

Форма обучения: **очная и заочная**

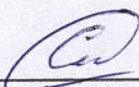
Курган 2020

Рабочая программа дисциплины «**Методы и алгоритмы анализа данных**» составлена в соответствии с учебным планом программы бакалавриата: «Программная инженерия» (**Программное обеспечение автоматизированных систем**), утвержденным:

- для очной формы обучения 28.08. 2020 г.
- для заочной формы обучения 28.08. 2020 г.


Рабочая программа дисциплины одобрена на заседании кафедры «Программное обеспечение автоматизированных систем» 31.08.2020 года, протокол № 1.

Рабочую программу составил:
профессор кафедры ПОАС



_____/В. А. Симахин/

Согласовано:

Заведующий кафедрой
«Программное обеспечение
автоматизированных систем»


_____/В. К. Волк/

Специалист
по учебно-методической работе
Учебно-методического отдела


_____/Г. В Казанкова/

1. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость – 3зач. ед. (108 акад. часов)

Вид учебной работы	Распределение трудоемкости по семестрам и видам учебных занятий, акад. часов			
	Очная форма обучения		Заочная форма обучения	
	Всего	Семестр	Всего	Семестр
		5		7
Аудиторные занятия в том числе:	64	64	12	12
Лекции	32	32	6	6
Лабораторные работы	32	32	6	6
Самостоятельная работа в том числе:	44	44	96	96
Контрольная работа	8	8	18	18
Подготовка к экзамену	27	27	27	27
Другие виды самостоятельной работы	9	9	51	51
Общая трудоемкость дисциплины	108	108	108	108
Виды промежуточной аттестации	Экзамен		Экзамен	

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Интеллектуальная обработка данных» является дисциплиной обязательной части блока Б1 модуля «Математические и естественно-научные дисциплины».

Дисциплина изучается в 5 семестре для очной формы обучения, в 4 семестре для заочной формы обучения и требует специальной подготовки обучающихся: "Основы программирования", "Вычислительная математика", "Теория вероятностей и математическая статистика", "Математический анализ"; "Системы цифровой обработки данных"; «Основы теории информации».

Результаты обучения по данной дисциплине необходимы при выполнении научно-исследовательской работы, а также при написании курсовых работ и подготовке выпускной квалификационной работы.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Цель изучения дисциплины

изучение математических методов и получение практических навыков при статистическом анализе исходной информации, применительно к задачам, возникающим при исследовании сложных систем.

Задачами дисциплины является изучение:

современного математического аппарата прикладной статистики, привитие практических навыков статистического исследования реальных задач и статистической обработки результатов эксперимента с использованием современного программного обеспечения ЭВМ и статистического моделирования.

Компетенции, формируемые в результате освоения дисциплины:

– Способность осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач (УК-1);

– Способность применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности; (ОПК-1);

– Способность использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности; (ОПК-2);

– Способность применять в практической деятельности основные концепции, принципы, теории и факты, связанные с информатикой (ОПК-7).

В результате освоения дисциплины обучающийся должен продемонстрировать следующие **результаты обучения**:

Должен знать:

– Основы критического анализа, синтеза информации, системного подхода для решения поставленных задач (УК-1);

– Методы математического анализа и моделирования, теоретического и экспериментального исследования (ОПК-1);

– Современные информационные технологии и программные средства, в том числе отечественного производства, (ОПК-2);

– Основные концепции, принципы, теории и факты, связанные с информатикой (ОПК-7).

Должен уметь:

– Осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач (УК-1);

– Применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности (ОПК-1);

– Использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности (ОПК-2);

– Применять в практической деятельности основные концепции, принципы, теории и факты, связанные с информатикой (ОПК-7).

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1 Учебно-тематический план

Очная форма обучения

Рубеж	Номер раздела, темы	Наименование раздела, темы	Количество часов контактной работы с преподавателем	
			Лекции	Лабораторные работы
Семестр 5				
Рубеж 1	1	Способы представления и модели порождения экспериментальных данных	2	
	2	Система предварительной статистической обработки данных	4	4
	3	Задачи идентификации систем	4	8
		Рубежный контроль №1		1
Рубеж 2	4	Методы редукции факторного пространства	4	4
	5	Методы планирования экспериментов	4	2
	6	Задачи распознавания образов	4	4
	7	Классификаторы.	4	4
	8	Алгоритмы распознавания образов.	4	4
	9	Программное обеспечение интеллектуального анализа данных	2	
			Рубежный контроль № 2	
Всего:			32	32

Заочная форма обучения

Рубеж	Номер раздела, темы	Наименование раздела, темы	Количество часов контактной работы с преподавателем	
			Лекции	Лабораторные работы
Семестр 7				
	1	Способы представления и модели порождения экспериментальных данных	1	
	2	Система предварительной статистической обработки данных	1	2
	3	Задачи идентификации систем	1	2
	4	Методы редукции факторного пространства	1	
	5	Методы планирования экспериментов	0.5	2
	6	Задачи распознавания образов	0.5	
	7	Классификаторы	0.5	
	8	Алгоритмы распознавания образов	0.5	
Всего:			6	6

4.2 Содержание лекционных занятий

Раздел 1 Способы представления и модели порождения экспериментальных данных

Способы представления и модели порождения экспериментальных данных. Основные задачи прикладной статистики. Этапы статистического исследования.

Статистические данные. Документирование, ввод, хранение и редактирование данных. Шкалы измерений: номинальная; порядковая; количественная.

Унифицирование разнотипных данных. Организация статистических данных на ЭВМ

Раздел 2 Система предварительной статистической обработки данных

Система предварительной статистической обработки данных.

Изучение эмпирических распределений. Оценивание числовых характеристик распределений. Робастные оценки сдвига и масштаба. Проверка априорных предположений о данных. Критерии согласия, однородности, симметрии независимости, стационарности. Пропущенные значения. Анализ резко выделяющихся наблюдений. Программное обеспечение подсистемы предварительной статистической обработки данных

Раздел 3 Задачи идентификации систем

Задачи идентификации систем. Уровни априорной информации. Типы моделей. Критерии качества. Линейный регрессионный анализ. Метод наименьших квадратов. Гребневая, полиномиальная регрессии. Робастная регрессия. Нелинейная регрессия. Непараметрические модели. Сплайны. Метод группового учета аргументов

Раздел 4 Методы редукции факторного пространства

Редукция факторного пространства. Метод главных компонент. Выбор наиболее информативных факторов. Визуализация многомерных данных

Раздел 5 Методы планирования экспериментов

Методы планирования экспериментов. Критерии оптимальности планов. Полный и дробный факторные эксперименты. Планы первого порядка. Центральные композиционные планы. Симплекс - планы

Раздел 6 Задачи распознавания образов

Общая характеристика проблемы распознавания объектов и явлений.

Понятие образа. Качественное описание задачи распознавания образов. Типы задач распознавания и их характерные черты. Структура системы распознавания образов. Задача распознавания образов как одна из задач анализа данных.

Раздел 7 Классификаторы

Формальная постановка задачи распознавания образов. Признаки и классификаторы. Классификация с обучением и без обучения. Решающие функции. Классификация образов с помощью функций расстояния. Классификация образов с помощью функций правдоподобия. Обучаемые классификаторы образов. Детерминистский подход. Обучаемые классификаторы образов. Статистический подход. Показатели эффективности распознавания. Информативные параметры

Раздел 8 Алгоритмы распознавания образов

Методы распознавания, основанные на сравнении с эталоном. Мера близости, основанная на поиске оптимального пути на графе. Задача сравнения контуров.

Статистические методы. Элементы теории статистических решений в распознавании образов. Байесовский подход. Дискриминантные функции и поверхность решения. Структурные и синтаксические методы. Методы преобразования. Языки описания образов. Обработка изображений.

Раздел 9 Программное обеспечение интеллектуального анализа данных

Архитектура автоматизированных систем научных исследований.

Организация пользовательского интерфейса в автоматизированных системах обработки данных. Библиотеки и пакеты прикладных программ по статистическому анализу. Сравнительные характеристики ППП

4.3. Лабораторные занятия

Номер раздела, темы	Наименование раздела, темы	Наименование лабораторной работы	Норматив времени, час.	
			очная форма обучения	заочная форма обучения
2	Система предварительной статистической обработки данных	Система предварительной статистической обработки данных	4	2
3	Задачи идентификации систем	Задачи идентификации систем	4	2
		Линейная регрессия	4	
	Рубежный контроль № 1		1	
4	Методы редукции факторного пространства	Редукция факторного пространства.	4	
5	Методы планирования экспериментов	Методы планирования экспериментов.	2	2
6	Задачи распознавания образов	Задачи распознавания образов.	4	
7	Классификаторы	Задача классификации	4	
8	Алгоритмы распознавания образов	Иерархический кластер анализ.	4	
		Статистические методы.		
	Рубежный контроль № 2		1	
Всего:			32	6

4.4 Контрольная работа (для очной и заочной форм обучения)

Требования к контрольной работе и варианты приведены в методических указаниях.

Варианты заданий №1

В таблице приведены данные о среднедушевом прожиточном минимуме в день на одного работающего x (в рублях) и данные о средней заработной плате за один рабочий день y (в рублях) в 15-ти регионах.

1. Постройте уравнение парной регрессии y от x .
2. Рассчитайте коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.
3. Оцените статистическую значимость параметров регрессии и уравнения регрессии с помощью F -критерия Фишера и t -критерия Стьюдента.
4. Найдите доверительные интервалы для коэффициентов регрессии и уравнения регрессии на уровне значимости $\alpha = 0,05$.
5. Найдите и удалите из выборки две точки, наиболее удалённые от линии регрессии. Постройте линию регрессии для этой выборки. Сравните результаты.

Вариант 1		Вариант 2		Вариант 3	
Прожи- точный ми- нимум	Заработ- ная плата	Прожи- точный ми- нимум	Заработ- ная плата	Прожи- точный ми- нимум	Заработ- ная плата
234	445	215	486	206	369
246	484	226	531	216	410
261	518	239	569	229	441
237	457	217	500	208	383
267	524	245	574	235	443
318	623	292	682	279	526
201	396	184	433	177	335
264	517	242	566	232	436
219	434	201	476	192	369
261	517	239	567	229	439
228	449	209	492	200	380
345	685	316	752	303	583
207	419	190	460	182	360
252	526	231	579	221	459
276	553	253	607	242	472

Варианты заданий №2

По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%) (смотри таблицу своего варианта).

Требуется:

1. Построить линейную модель множественной регрессии и выполнить анализ результатов.

Вариант 1

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,6	9	11	9	6,3	21
2	6	3,6	12	12	11	6,4	22
3	6	3,9	14	13	11	7	24
4	7	4,1	17	14	12	7,5	25
5	7	3,9	18	15	12	7,9	28
6	7	4,5	19	16	13	8,2	30
7	8	5,3	19	17	13	8	30
8	8	5,3	19	18	13	8,6	31
9	9	5,6	20	19	14	9,5	33
10	10	6,8	21	20	14	9	36

5 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Лекционный курс базируется на пассивном методе обучения, реализующем традиционную объяснительно-иллюстративную образовательную технологию, в рамках которой студенты выступают в роли слушателей, воспринимающих учебный материал и участвующих в дискуссиях и экспресс-опросах.

При прослушивании лекций рекомендуется в конспекте отмечать все важные моменты, на которых заостряет внимание преподаватель, в частности те, которые направлены на качественное выполнение соответствующей лабораторной работы.

Преподавателем запланировано использование при чтении лекций технологии учебной дискуссии. Поэтому рекомендуется фиксировать для себя интересные моменты с целью их активного обсуждения на дискуссии в конце лекции.

Конспект каждой лекции завершается перечнем контрольных вопросов, ответы на которые должны быть получены студентом в процессе самостоятельной проработки материала лекции при подготовке к очередному лекционному занятию.

Лабораторные работы проводятся на основе интерактивных методов в виде творческих заданий экспериментального характера, направленных не столько на закрепление уже изученного материала, сколько на изучение нового, и выполняемые студентами, объединяемыми в малые группы (2-3 человека). Задания не имеют однозначного решения и соответствуют целям обучения.

Залогом качественного выполнения лабораторных работ является самостоятельная подготовка к ним накануне путем повторения материалов лекций. Рекомендуется подготовить вопросы по неясным моментам и обсудить их с преподавателем в начале лабораторного занятия.

Преподавателем запланировано применение на лабораторных занятиях технологий развивающейся кооперации, коллективного взаимодействия, разбора конкретных ситуаций.

Для текущего контроля успеваемости по очной форме обучения преподавателем используется балльно-рейтинговая система контроля и оценки академической активности. Поэтому настоятельно рекомендуется тщательно прорабатывать материал дисциплины при самостоятельной работе, участвовать во всех формах обсуждения и взаимодействия, как на лекциях, так и лабораторных занятиях в целях лучшего освоения материала и получения высокой оценки по результатам освоения дисциплины.

Выполнение самостоятельной работы подразумевает самостоятельное изучение разделов дисциплины, подготовку к лабораторным занятиям, рубежным контролям для очной формы обучения, подготовку к экзамену, выполнение контрольной работы для очной и заочной формы обучения

Рекомендуемая трудоемкость самостоятельной работы представлена в таблице.

Рекомендуемый режим самостоятельной работы

Наименование вида самостоятельной работы	Рекомендуемая трудоемкость, акад. час.	
	Очная форма обучения	Заочная форма обучения
Самостоятельное изучение тем дисциплины:	4	48
<p>Раздел 1 Способы представления и модели порождения экспериментальных данных Способы представления и модели порождения экспериментальных данных. Основные задачи прикладной статистики. Этапы статистического исследования. Статистические данные. Документирование, ввод, хранение и редактирование данных. Шкалы измерений: номинальная; порядковая; количественная. Унифицирование разнотипных данных. Организация статистических данных на ЭВМ</p>	1	6
<p>Раздел 2 Система предварительной статистической обработки данных Система предварительной статистической обработки данных Изучение эмпирических распределений. Оценивание числовых характеристик распределений. Робастные оценки сдвига и масштаба. Проверка априорных предположений о данных. Критерии согласия, однородности, симметрии независимости, стационарности. Пропущенные значения. Анализ резко выделяющихся наблюдений. Программное обеспечение подсистемы предварительной статистической обработки данных</p>		6
<p>Раздел 3 Задачи идентификации систем Задачи идентификации систем. Уровни априорной информации. Типы моделей. Критерии качества. Линейный регрессионный анализ. Метод наименьших квадратов. Гребневая, полиномиальная регрессии. Робастная регрессия. Нелинейная регрессия. Непараметрические модели. Сплаины. Метод группового учета аргументов</p>		6
<p>Раздел 4 Методы редукции факторного пространства Редукция факторного пространства. Метод главных компонент. Выбор наиболее информативных факторов. Визуализация многомерных данных</p>	1	6
<p>Раздел 5 Методы планирования экспериментов Методы планирования экспериментов. Критерии оптимальности планов. Полный и дробный факторные эксперименты. Планы первого порядка. Центральные композиционные планы. Симплекс- планы</p>		6
<p>Раздел 6 Задачи распознавания образов Общая характеристика проблемы распознавания объектов и явлений. Понятие образа. Качественное описание задачи распознавания образов. Типы задач распознавания и их характерные черты. Структура системы распознавания образов. Задача распознавания</p>	1	6

образов как одна из задач анализа данных.		
Раздел 7 Классификаторы Формальная постановка задачи распознавания образов. При-знаки и классификаторы. Классификация с обучением и без обучения. Решающие функции. Классификация образов с помощью функций расстояния. Классификация образов с помощью функций правдоподобия. Обучаемые классификаторы образов. Детерминистский подход. Обучаемые классификаторы образов. Статистический подход. Показатели эффективности распознавания. Информативные параметры		6
Раздел 8 Алгоритмы распознавания образов Методы распознавания, основанные на сравнении с эталоном. Мера близости, основанная на поиске оптимального пути на графе. Задача сравнения контуров. Статистические методы. Элементы теории статистических решений в распознавании образов. Байесовский подход. Дискриминантные функции и поверхности решения. Структурные и синтаксические методы. Методы предобработки. Языки описания образов. Обработка изображений.	1	2
Раздел 9 Программное обеспечение интеллектуального анализа данных Архитектура автоматизированных систем научных исследований. Организация пользовательского интерфейса в автоматизированных системах обработки данных. Библиотеки и пакеты прикладных программ по статистическому анализу. Сравнительные характеристики ИПП.		4
Подготовка к лабораторным занятиям (0,5 часа на каждое занятие (для очной формы обучения и 1 час на каждое занятие для заочной формы обучения)	4	3
Контрольная работа	8	18
Подготовка к рубежным контролям (0,5 часа на каждый рубежный контроль)	1	-
Подготовка к экзамену	27	27
Всего:	44	96

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

6.1. Перечень оценочных средств

1. Балльно-рейтинговая система контроля и оценки академической активности студентов в КГУ (для очной формы обучения)
2. Отчеты студентов по лабораторным работам.
3. Банк тестовых заданий к рубежным контролям № 1, № 2 (для очной формы обучения)
4. Контрольная работа (для очной и заочной формы обучения)
5. Вопросы к экзамену

6.2. Система балльно-рейтинговой оценки работы обучающихся по Дисциплине

№	Наименование	Содержание				
Очная форма обучения						
1	Распределение баллов за семестр по видам учебной работы, сроки сдачи учебной работы (доводятся до сведения студентов на первом учебном занятии)	Распределение баллов за 5 семестр				
		Вид учебной работы:	Выполнение и защита отчетов по лабораторным работам	КР	Рубежный контроль №1	Рубежный контроль №2
	Балль-ная оценка:	26*16=326	18	10	10	30
2	Критерий пересчета баллов в традиционную оценку по итогам работы в семестре и зачета	60 и менее баллов – неудовлетворительно; не зачтено; 61...73 – удовлетворительно; зачтено; 74... 90 – хорошо; 91... 100 – отлично				
3	Критерии допуска к промежуточной аттестации, возможности получения автоматического зачета (экзаменационной оценки) по дисциплине, возможность получения бонусных баллов	<p>Для допуска к промежуточной аттестации (экзамену) студент должен набрать не менее 50 баллов, выполнить и защитить все лабораторные работы и контрольную работу</p> <p>Для получения экзаменационной оценки «автоматически» студенту необходимо набрать следующее минимальное количество баллов:</p> <ul style="list-style-type: none"> - 68 для получения «автоматически» оценки «удовлетворительно». <p>По согласованию с преподавателем студенту, набравшему 68 баллов, могут быть добавлены дополнительные (бонусные) баллы за активность на лабораторных занятиях, активное участие в научной и методической работе, оригинальность принятых решений в ходе выполнения лабораторных работ, за участие в значимых учебных и внеучебных мероприятиях кафедры и выставлена автоматически оценка «хорошо» или «отлично».</p>				
4	Формы и виды учебной работы для неуспевающих (восстановившихся на курсе обучения) магистрантов для получения недостающих баллов в конце семестра	<p>В случае, если к промежуточной аттестации (экзамену) студент набрал сумму менее 50 баллов, то студенту необходимо набрать недостающие баллы и выполнить дополнительные задания до конца последней (зачетной) недели семестра. При этом необходимо проработать материал всех пропущенных лабораторных работ.</p> <p>Формы дополнительных заданий (назначаются преподавателем):</p> <ul style="list-style-type: none"> - выполнение и защита пропущенной лабораторной работы (при невозможности дополнительного ее проведения преподаватель устанавливает форму дополнительного задания по тематике пропущенной лабораторной работы самостоятельно) – до 5 баллов. При прохождении рубежного контроля- баллы в зависимости от рубежа и контрольной работы-до 7 баллов. <p>Ликвидация академических задолженностей, возникших из-за разности в учебных планах при переводе или восстановлении, проводится путем выполнения дополнительных заданий, форма и объем которых определяется преподавателем</p>				

6.3 Процедура оценивания результатов освоения дисциплины

Рубежный контроль осуществляется в форме фронтального тестирования по разделам дисциплины в первом и втором семестре. Тест по каждому разделу содержит 10 вопросов, для каждого рубежного контроля. Оценивается количество правильных ответов на задания теста: студент, ответивший правильно менее, чем на 3 задания теста, считается не прошедшим тестирование и обязан повторно пройти этот тест во время консультации по дисциплине, а также во время проведения консультаций по дисциплине в форме собеседования.

На каждое тестирование при рубежном контроле студенту отводится 1 академический час.

Преподаватель оценивает в баллах результаты тестирования каждого студента по количеству правильных ответов и заносит в ведомость учета текущей успеваемости.

Примерные тестовые задания приведены ниже. Каждый вопрос оценивается в один балл.

Экзамен проводится в традиционной (устной) форме: студент выполняет задания билета, включающего два теоретических вопроса и одну задачу, и отвечает преподавателю. Оцениваются полнота и правильность ответов студента на теоретические вопросы билета, его эрудиция в смежных вопросах.

Вопросы к экзамену доводятся до студентов на последней лекции в семестре. Каждый вопрос оценивается в 10 баллов. На подготовку ответа студенту отводится 1 астрономический час.

Результаты текущего контроля успеваемости и экзамена заносятся преподавателем в экзаменационную ведомость, которая сдается в организационный отдел института в день экзамена, а также выставляются в зачетную книжку студента.

6.4 Примеры оценочных средств для рубежных контролей, экзамена

6.4.1. Примеры тестовых заданий для рубежного контроля №1

1. Алгоритмы распознавания образов достаточно трудоемки, все это вызывает проблему
 - 1) уменьшения размерности факторного пространства
 - 2) увеличения размерности факторного пространства
 - 3) уменьшения критерия оптимальности
 - 4) увеличения критерия оптимальности
2. Что из перечисленного не является предпосылкой для задачи редукции
 - 1) дублирование информации в исходных факторах
 - 2) высокая вариабельность среды
 - 3) неинформативность факторов
 - 4) возможность агрегирования
3. Какой из критериев связан с задачами регрессионного анализа и классификации
 - 1) критерий автоинформативности

- 2) критерий внешней информативности
 - 3) критерий внутренней информативности
 - 4) нет правильного варианта ответа
4. Какой метод не используется для решения задачи классификации:
- 1) метод главных компонент
 - 2) факторный анализ
 - 3) многомерное шкалирование
 - 4) регрессионный анализ
5. Метод экстремальной группировки факторов подразумевает
- 1) разбиение факторного пространства на группы
 - 2) объединение факторного пространства по группам
 - 3) разбиение факторного пространства на группы, где в каждой группе факторы значительно коррелированы
 - 4) разбиение факторного пространства на группы, где в каждой группе факторы коррелированы слабо
6. Первая главная компонента характеризует
- 1) проекцию точек на z_1
 - 2) проекцию точек на x_1
 - 3) проекцию точек на y_1
 - 4) проекцию точек на z_2
7. Выберите наиболее полный ответ. Нестационарный случайный процесс находится под воздействием
- 1) сезонных и циклических составляющих анализируемого признака
 - 2) тренда, сезонных и циклических составляющих анализируемого признака
 - 3) случайных составляющих, сезонных и циклических составляющих анализируемого признака
 - 4) тренда, сезонных и циклических составляющих анализируемого признака, случайных составляющих
8. Основные задачи временных рядов
- 1) выяснить какие из функций f, φ, ψ присутствуют, построить оценки этих функций
 - 2) выбор алгоритмов, в зависимости от функции распределения
 - 3) все вышеперечисленное
 - 4) нет правильного ответа
9. Процесс $X(t)$ называется стационарным в широком смысле
- 1) если его n -мерная функция распределения зависит от сдвига $F_n(x_1 \dots x_n; t_1 \dots t_n)$ по времени
 - 2) если его n -мерная функция распределения не зависит от сдвига $F_n(x_1 \dots x_n; t_1 \dots t_n)$ по времени
 - 3) если его среднее значение не зависит от времени
 - 4) если его среднее значение зависит от времени

10. Осуществление проверки на стационарность
- 1) по критерию серий
 - 2) по критерию автоинформативности
 - 3) по критерию внешней информативности
 - 4) по критерию эргодичности
11. Выберите верное утверждение, связанное с критерием серий
- 1) чем больше зависимость в данных, тем длиннее серии
 - 2) чем больше зависимость в данных, тем короче серии
 - 3) чем больше серий, тем больше зависимости в данных
 - 4) чем меньше серий, тем больше зависимости в данных
12. Методы сглаживания временного ряда
- 1) направлены на выделение случайной составляющей
 - 2) направлены на выделение неслучайной составляющей
 - 3) направлены на выделение тренда
 - 4) направлены на выделение сезонных и циклических составляющих
13. В параметрических методах сглаживания временного ряда применяется
- 1) только МНК - метод наименьших квадратов
 - 2) только МНМ – метод наименьших модулей
 - 3) МНК или МНМ (МНК можно заменить МНМ, если есть выбросы)
 - 4) нет правильного варианта ответа
14. В статистической теории классификации класс - это
- 1) генеральная совокупность, описываемая унимодальной функцией распределения
 - 2) генеральная совокупность, описываемая функцией распределения
 - 3) генеральная совокупность, описываемая модальной функцией распределения
 - 4) нет правильного варианта ответа
15. В факторном анализе сходства между объектами (например, переменными) выражаются с помощью
- 1) произвольного типа матрицы сходства объектов
 - 2) матрицы коэффициентов корреляций
 - 3) матрицы коэффициентов сходства объектов
 - 4) матрицы числовых значений координат
16. Целью факторного анализа является
- 1) поиск латентных переменных
 - 2) интерпретация наблюдаемых переменных, дающих возможность пользователю объяснить сходства между объектами
 - 3) поиск и интерпретация непосредственно не наблюдаемых переменных, дающих возможность пользователю объяснить сходства между объектами
 - 4) верны все варианты
17. Для синтеза управления

1) в факторном анализе ориентация осей может быть выбрана произвольной, в многомерном шкалировании нет

2) в факторном анализе ориентация осей не может быть выбрана произвольной, в многомерном шкалировании может

3) и в факторном анализе, и в многомерном шкалировании ориентация осей может быть выбрана произвольной

4) ни в факторном анализе, ни в многомерном шкалировании ориентация осей не может быть выбрана произвольной

18. Отличие метода многомерного шкалирования от факторного анализа заключается в том, что

1) МНШ стремится извлечь больше факторов по сравнению с факторным анализом

2) методы факторного анализа можно применять к любым типам расстояний или сходств, а методы МНШ нет

3) методы МНШ требуют, чтобы исследуемые данные подчинялись многомерному нормальному распределению, а зависимости были линейными, факторный анализ не накладывает таких ограничений

4) факторный анализ требует, чтобы исследуемые данные подчинялись многомерному нормальному распределению, а зависимости были линейными, методы МНШ не накладывает таких ограничений

19. Выберите неверное утверждение

1) более сильные зависимости между переменными являются более значимыми

2) объем выборки не влияет на значимость зависимости

3) слабые связи могут быть значимо доказаны только на больших выборках

4) чем больше величина зависимости (связи) между переменными в выборке обычного объема, тем более она надежна

20. Выберите неверное утверждение

1) системы получения данных в большей степени ориентированы на практическое приложение полученных результатов, чем на выяснение природы явления

2) в области получения данных принят такой подход к анализу данных и извлечению знаний, который иногда характеризуют словами "черный ящик"

3) в области получения данных не используются такие методы как нейронные сети

4) получение данных трактуется как «смесь статистики, методов искусственного интеллекта и анализа баз данных»

6.4.2. Примеры тестовых заданий для рубежного контроля №2

1. Дайте определение разделяющей и решающей функции.

2. Как определяется принадлежность образа к классу

1) через разделяющие функции;

2) через решающие функции.

3. Приведите примеры непараметрического и параметрического подходов к классификации образов, представленных набором параметров.
 4. В каких случаях кластеризация только по пороговому ограничению может дать удовлетворительный результат.
 5. Перечислите основные задачи корреляционного анализа при классификации многомерных данных.
 6. Дайте определение ковариационной и корреляционной матрицы. Чем они отличаются.
 7. С какой целью осуществляется переход к стандартизованной (нормализованной) матрице данных. Какими свойствами обладает эта матрица.
 8. Приведите примеры задач, в которых применяется факторный анализ.
 9. Какими свойствами обладают общие факторы в модели главных компонент.
 10. Какой показатель минимизируется (максимизируется) в задаче главных компонент.
 11. Перечислите основные этапы решения задачи главных компонент в общем виде. Как эта задача решается на практике.
 12. Дайте графическую интерпретацию главных компонент для признакового пространства в задаче классификации многомерных данных.
 13. В каких задачах распознавания целесообразно применение статистического подхода. Какие данные об объектах распознавания необходимы для его реализации.
 14. Объясните содержательно понятие ошибок первого и второго рода. Как проявляются ошибки первого и второго рода при классификации пикселей изображения.
 15. Объясните, как влияют на значение коэффициента правдоподобия
 - 1) априорные вероятности появления классов
 - 2) значения платежных коэффициентов.
- Что называют нуль-единичной байесовской стратегией и в каких случаях ее целесообразно использовать.
16. Какие стратегии принятия решения, основанные на байесовском подходе используются
 - 1) при неизвестных априорных вероятностях
 - 2) при неизвестных априорных вероятностях и платежных коэффициентах
 17. С какой целью при классификации пикселей изображения широко дисперсные классы часто разбивают на несколько подклассов
 18. Что такое апостериорная вероятность. Продемонстрируйте графически принцип принятия решения по максимуму правдоподобия в одномерном случае для трех классов с различными средними и дисперсиями.
 19. В каком случае результат классификации по принципу максимума правдоподобия совпадает с результатом классификации по минимуму расстояния. Объясните это аналитически.
 20. Перечислите основные способы предварительной оценки качества классификации по принципу максимума правдоподобия. В каких случаях

целесообразно использовать каждый из этих способов.

Примерный перечень вопросов к экзамену

1. Способы представления и модели порождения экспериментальных данных
2. Основные задачи анализа данных
3. Этапы статистического исследования
4. Статистические данные
5. Документирование, ввод, хранение и редактирование данных
6. Шкалы измерений: номинальная; порядковая; количественная
7. Организация статистических данных на ЭВМ
8. Система предварительной статистической обработки данных
9. Изучение эмпирических распределений
10. Оценивание числовых характеристик распределений. Робастные оценки сдвига и масштаба
11. Программное обеспечение подсистемы предварительной статистической обработки данных
12. Задачи идентификации систем. Уровни априорной информации. Типы моделей. Критерии качества
13. Линейный регрессионный анализ.
14. Метод наименьших квадратов
15. Гребневая, полиномиальная регрессии
16. Робастная регрессия
17. Непараметрические модели
18. Редукция факторного пространства
19. Метод главных компонент
20. Задачи распознавания образов
21. Архитектура автоматизированных систем научных исследований
22. Организация пользовательского интерфейса в автоматизированных системах обработки данных.
23. Понятие образа.
24. Качественное описание задачи распознавания образов
25. Типы задач распознавания и их характерные черты
26. Структура системы распознавания образов.
27. Задача распознавания образов как одна из задач анализа данных
28. Признаки и классификаторы
29. Классификация образов с обучением и без обучения
30. Классификация образов с помощью функций расстояния
31. Методы распознавания, основанные на сравнении с эталоном
32. Элементы теории статистических решений в распознавании образов
33. Дискриминантные функции и поверхности решения
34. Алгоритм персептрона
35. Классификация нейронных сетей
36. Нейронные сети Хопфилда и Хэмминга

Примеры задач

Задание №1.

Реализовать метод К-средних, минимакса, метод голосования и выполнить кластеризацию образов

Вариант 1

Образы: четырехугольники.

Признаки: четыре длины сторон четырехугольника.

Класс «маленький четырехугольник»: (2,2,2,2), (2,2,3,2), (2,3,2,2), (3,2,2,2), (3,3,2,2), (2,3,3,2), (3,2,3,2), (3,3,3,2), (2,2,2,3), (2,2,3,3), (2,3,2,3), (3,2,2,3), (3,3,2,3), (2,3,3,3), (3,2,3,3), (3,3,3,3).

Класс «большой четырехугольник»: (9,9,9,9), (9,9,11,9), (9,11,9,9), (11,9,9,9), (11,11,9,9), (11,9,11,9), (9,11,11,9), (11,11,11,9), (9,9,9,11), (9,9,11,11), (9,11,9,11), (11,9,9,11), (11,11,9,11), (11,9,11,11), (9,11,11,11), (11,11,11,11).

Класс «очень маленький четырехугольник»: (0.1,0.1,0.1,0.1).

Вариант 2

Образы: книги.

Признаки: количество страниц, количество картинок, количество таблиц, средняя длина слова.

Класс «детская книга»: (200,50,0,4), (150,37,0,5).

Класс «учебник»: (400,15,90,8), (500,3,105,7).

Класс «журнал»: (80,200,5,6), (100,198,4,7).

Вариант 3

Образы: погода на сутки (летом).

Признаки: количество часов с дождем, количество солнечных часов, средняя температура воздуха, количество детей во дворе днем, скорость ветра.

Класс «хорошая погода»: (0,17,25,15,2), (1,16,24,13,3), (1,17,23,15,2).

Класс «плохая погода»: (20,0,18,1,7), (17,2,18,0,8), (18,1,19,2,7).

Вариант 4

Образы: погода на сутки (зимой).

Признаки: количество часов со снегом, количество солнечных часов, средняя температура воздуха, количество детей во дворе днем, скорость ветра.

Класс «хорошая погода»: (0,13,-10,15,2), (1,12,-12,13,3), (1,13,-13,15,2).

Класс «плохая погода»: (20,0,-21,1,8), (17,2,-24,0,9), (18,1,-23,2,8).

Вариант 5

Образы: животные.

Признаки: количество полосок, рост (см), длина хвоста (см).

Класс «пчела»: (7,1,0), (8,2,0), (6,1,0).
Класс «зебра»: (30,150,50), (35,160,60), (33,155,50).
Класс «тигр»: (60,100,100), (70, 110,110), (55,100,110).
Класс «муравей»: (0,0.5,0), (0,0.6,0).

Вариант 6

Образы: одежда.

Признаки: длина рукава, количество пуговиц, длина изделия.

Класс «футболка»: (15,0,60), (20,0,70).

Класс «рубашка»: (50,8,60), (60,9,60), (55,8,70).

Класс «летняя рубашка»: (15,6,60), (20,8,70), (20,7,65).

Класс «халат»: (15,9,120), (20,10,130), (25,9,115).

Вариант 7

Образы: школьники.

Признаки: оценка по алгебре, оценка по геометрии, оценка по физике, оценка по истории, оценка по литературе, оценка по иностранному языку, оценка по физкультуре.

Класс «отличник»: (5,5,5,5,5,5,5), (4,5,5,5,5,5,5), (5,4,5,5,5,5,5), (5,5,4,5,5,5,5), (5,5,5,4,5,5,5), (5,5,5,5,4,5,5), (5,5,5,5,5,4,5), (5,5,5,5,5,5,4).

Класс «троечник»: (3,3,3,3,3,3,3), (4,3,3,3,3,3,3), (3,4,3,3,3,3,3), (3,3,4,3,3,3,3), (3,3,3,4,3,3,3), (3,3,3,3,4,3,3), (3,3,3,3,3,4,3), (3,3,3,3,3,3,4), (3,3,3,3,3,3,5).

Класс «математик»: (5,5,5,3,3,3,4), (5,5,5,3,4,3,3), (5,5,5,4,3,3,4), (5,5,5,3,3,4,3), (5,5,5,3,4,4,3), (5,5,5,4,4,3,4), (5,5,5,4,3,4,3).

Класс «гуманитарий»: (3,3,3,5,5,5,4), (3,4,3,5,5,5,3), (4,3,3,5,5,5,4), (3,3,4,5,5,5,3), (3,4,4,5,5,5,3), (4,4,3,5,5,5,4), (4,3,4,5,5,5,3).

Вариант 8

Образы: ягоды.

Признаки: количество семечек, радиус сечения (см), количество на одном растении.

Класс «вишня»: (1,2,300), (1,1.5,400), (1,1,350).

Класс «арбуз»: (50,20,1), (60,25,1), (40,21,1).

Класс «облепиха»: (6,0.3,500), (8,0.2,550), (8,0.3,600).

Класс «виноград»: (6,1,1500), (8,1.5,1400), (8,0.7,1700).

Вариант 9

Образы: пятиугольники.

Признаки: пять длин сторон пятиугольника.

Класс «маленький пятиугольник»: (2,2,2,2,2), (2,2,3,2,2), (2,3,2,2,3), (3,2,2,2,3), (3,3,2,2,2), (2,3,3,2,3), (3,2,3,2,2), (3,3,3,2,3), (2,2,2,3,2), (3,3,3,3,3).

Класс «большой пятиугольник»: (9,9,9,9,9), (9,9,11,9,11), (9,11,9,9,9), (11,9,9,9,11), (11,11,9,9,9), (11,9,11,9,11), (9,11,11,9,9), (11,11,11,9,9), (11,11,11,11,11).

Класс «очень маленький пятиугольник»: (0.1,0.1,0.1,0.1,0.1).

Вариант 10

Образы: треугольники.

Признаки: три длины сторон треугольника.

Класс «маленький треугольник»: (2,2,2), (2,2,3), (2,3,2), (3,2,2), (3,3,2), (2,3,3), (3,2,3), (3,3,3).

Класс «большой треугольник»: (9,9,9), (9,9,11), (9,11,9), (11,9,9), (11,11,9), (11,9,11), (9,11,11), (11,11,11).

Класс «очень маленький треугольник»: (0.1,0.1,0.1).

6.5 Фонд оценочных средств

Полный банк заданий для текущего и рубежных контролей и промежуточной аттестации по дисциплине, показатели, критерии шкалы оценивания компетенций, методические материалы, определяющие процедуры оценивания образовательных результатов, приведены в УМК дисциплины.

7. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ УЧЕБНАЯ ЛИТЕРАТУРА

7.1 Основная литература

1. Макшанов, А. В. "Технологии интеллектуального анализа данных" [Электронный ресурс] : учеб. пособие / А. В. Макшанов, А. Е. Журавлев. — 2-е изд., стер. — Санкт-Петербург : Лань, 2019. — 212 с. – Доступ из ЭБС "Лань".

2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. "Прикладная статистика. Классификация и снижение размерности", - М.: Финансы и статистика, 1989.

7.2 Дополнительная литература

3. Тюрин С. Ю.Н., Макаров А.А. "Анализ данных на компьютере", Под ред. В.Э. Фигурного. - М.: ИНФРА-М, Финансы и статистика, 1995

4. Дилигенская А.Н. "Идентификация объектов управления", Самара, СГТУ, 2009. Режим доступа: свободный. <http://www.fineprint.com>

5. Сухарев М.Г. "Методы прогнозирования", Учебное пособие М. РГУ нефти и газа им. И.М. Губкина, 2009 г. Режим доступа: свободный. <http://www.fineprint.com>

6. Духанов А.В., Медведева О.Н. "Имитационное моделирование сложных систем", Курс лекций, Владимир, ВГУ, 2010 Режим доступа: свободный. <http://www.fineprint.com>

8. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

1. Задания для выполнения лабораторных работ и методические указания. Курган, КГУ, 2019

9. РЕСУРСЫ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫЕ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

– Сайт дистанционного обучения в НОУ (Национальный Открытый Университет) «ИНТУИТ» содержит бесплатные курсы, программы повышения квалификации и профессиональной переподготовки, интересные доклады и другую полезную информацию <http://www.intuit.ru>.

– Федеральный портал «Российское образование» <http://www.edu.ru/>

– Информационный сайт, содержащий справочные материалы по информатике, которые включают в себя курс лекций, схемы, презентации, рефераты и др. informatikaplus.narod.ru

– Сайт о высоких технологиях, новости индустрии из мира компьютерного «железа», тестовые испытания и обзоры оборудования IXBT.com.

– Портал «Информационно-коммуникационные технологии в образовании» <http://www.ict.edu.ru>.

-Система поддержки учебного процесса КГУ dist.kgsu.ru.

<http://www.edu.ru> Федеральный портал «Российское образование

Образовательный портал «УМНИК»: <http://new.volsu.ru/umnik/>

Федеральный образовательный портал. Библиотека. Единое окно доступа к образовательным ресурсам: <http://window.edu.ru/library>

9. РЕСУРСЫ СЕТИ «ИНТЕРНЕТ», НЕОБХОДИМЫЕ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

– Сайт дистанционного обучения в НОУ (Национальный Открытый Университет) «ИНТУИТ» содержит бесплатные курсы, программы повышения квалификации и профессиональной переподготовки, интересные доклады и другую полезную информацию <http://www.intuit.ru>.

– Федеральный портал «Российское образование» <http://www.edu.ru/>

– Информационный сайт, содержащий справочные материалы по информатике, которые включают в себя курс лекций, схемы, презентации, рефераты и др. informatikaplus.narod.ru

– Сайт о высоких технологиях, новости индустрии из мира компьютерного «железа», тестовые испытания и обзоры оборудования IXBT.com.

– Портал «Информационно-коммуникационные технологии в образовании» <http://www.ict.edu.ru>.

-Система поддержки учебного процесса КГУ dist.kgsu.ru.

<http://www.edu.ru> Федеральный портал «Российское образование

Образовательный портал «УМНИК»: <http://new.volsu.ru/umnik/>

Федеральный образовательный портал. Библиотека. Единое окно доступа к образовательным ресурсам: <http://window.edu.ru/library>

10. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ И ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ

1. ЭБС «Лань»
2. ЭБС «Консультант студента»
3. ЭБС «Znanium.com»

4. «Гарант» - справочно-правовая система

11. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Материально-техническое обеспечение по реализации дисциплины осуществляется с требованиями ФГОС ВО по данной образовательной программе.

12. ДЛЯ ОБУЧАЮЩИХСЯ С ИСПОЛЬЗОВАНИЕМ ДИСТАНЦИОННЫХ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

При использовании электронного обучения и дистанционных образовательных технологий (далее ЭО и ДОТ) занятия полностью или частично проводятся в режиме онлайн. Объем дисциплины и распределение нагрузки по видам работ соответствует п 4.1. Распределение баллов соответствует п 6.2 либо может быть изменено в соответствии с решением кафедры, в случае перехода на ЭО и ДОТ в процессе обучения. Решение кафедры об используемых технологиях и системе оценивания достижений обучающихся принимается с учетом мнения ведущего преподавателя и доводится до обучающихся.

Аннотация
рабочей программы учебной дисциплины
«Методы и алгоритмы анализа данных»

образовательной программы высшего образования –
программы бакалавриата
09.03.04 – Программная инженерия
Направленность:
Программное обеспечение автоматизированных систем

Формы обучения: **очная, заочная**

Трудоемкость дисциплины: 3 ЗЕ (108 академических часов)

Семестр: 5(очная форма обучения)

Семестр: 7(заочная форма обучения)

Форма промежуточной аттестации: экзамен

Содержание дисциплины

Раздел 1 Способы представления и модели порождения экспериментальных данных

Раздел 2 Система предварительной статистической обработки данных

Раздел 3 Задачи идентификации систем

Раздел 4 Методы редукции факторного пространства

Раздел 5 Методы планирования экспериментов

Раздел 6 Задачи распознавания образов

Раздел 7 Классификаторы

Раздел 8 Алгоритмы распознавания образов

Раздел 9 Программное обеспечение интеллектуального анализа данных