

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

Кафедра «Программное обеспечение автоматизированных систем»



УТВЕРЖДАЮ:  
Первый проректор

Т.Р. Змызгова

« 08 » 2022 г.

Рабочая программа учебной дисциплины

## **Интеллектуальный анализ данных**

образовательной программы высшего образования –  
программы магистратуры

**09.04.04 – Программная инженерия**

Направленность:

**Методы и алгоритмы интеллектуальной обработки данных  
в информационно- вычислительных системах**

Форма обучения: **очная**

Курган 2022

Рабочая программа дисциплины «Интеллектуальный анализ данных» составлена в соответствии с учебным планом программы магистратуры: «Программная инженерия» (Методы и алгоритмы интеллектуальной обработки данных в информационно-вычислительных системах), утвержденным для очной формы обучения 30. 08. 2022 г.

Рабочая программа дисциплины одобрена на заседании кафедры Программного обеспечения автоматизированных систем 30.08.2022 года

Рабочую программу разработал  
профессор кафедры ПОАС



В.А.Симахин

Согласовано:

Заведующий  
кафедрой ПОАС



В. К. Волк

Специалист  
по учебно-методической работе  
Учебно-методического отдела



Г.В. Казанкова

Начальник  
Управления  
образовательной деятельности



И. В. Григоренко

## 1. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость – 8 зач. ед. (288 акад. часов)

Виды учебной работы	Распределение трудоемкости по семестрам и видам учебных занятий (акад. часов)		
	Всего	1-й семестр	2-й семестр
Аудиторные занятия:	<b>104</b>	<b>56</b>	<b>48</b>
Лекции	56	32	24
Практические занятия	48	24	24
Самостоятельная работа	<b>184</b>	<b>88</b>	<b>96</b>
В том числе:			
Выполнение контрольной работы	-	-	-
Подготовка к зачету	18	18	
Подготовка к экзамену	27	-	27
Прочие виды	139	70	69
Вид промежуточной аттестации	Зачет, экзамен	зачет	экзамен
Общая трудоемкость	<b>288</b>	<b>144</b>	<b>144</b>

## **2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ**

Дисциплина «Интеллектуальный анализ данных» является дисциплиной блока 1 модуля « Анализ данных и машинное обучение»

Дисциплина изучается в первом и втором семестре и требует специальной подготовки обучаемых: "Основы программирования", "Вычислительная математика", "Теория вероятностей и математическая статистика", "Математический анализ"; "Системы цифровой обработки данных"; "Методы оптимизации и принятия решений"; «Основы теории управления».

Результаты обучения по дисциплине необходимы при выполнении научно-исследовательской работы, а также при подготовке выпускной квалификационной работы.

## **3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ**

**Цель** изучения дисциплины – это 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 самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте (для ОПК-1)

-Способность применять на практике новые научные принципы и методы исследований (для ОПК-4)

В результате изучения дисциплины обучающийся:

*Должен знать:*

-математические методы для решения естественнонаучных, социально-экономических и профессиональные задач, в том числе нестандартных, в новой или незнакомой среде (для ОПК-1);

- новые научные принципы и методы исследований ( для ОПК 4)

*Должен уметь:*

- самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте (для ОПК-1);

- применять на практике новые научные принципы и методы исследований (для ОПК-4)

#### 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

##### 4.1 Учебно-тематический план

Рубеж	Номер раздела, темы	Наименование раздела, темы	Количество часов контактной работы с преподавателем		
			Лекции	Лабораторные работы	Практические работы
<b>Семестр 1</b>					
	1	Раздел 1 Способы представления и модели порождения экспериментальных данных	4		
	2	Раздел 2 Система предварительной статистической обработки данных	6		4
Рубеж 1	3	Раздел 3 Задачи идентификации систем	8		8
		Рубежный контроль №1 (1 семестр)			1
	4	Раздел 4 Методы редукции факторного пространства	6		4
	5	Раздел 5 Методы планирования экспериментов	8		6
Рубеж 2		Рубежный контроль № 2 (1 семестр)			1
<b>Семестр 2</b>					
Рубеж 1	6	Раздел 6 Задачи распознавания образов	4		6
	7	Раздел 7 Классификаторы	8		6
		Рубежный контроль № 1			1
Рубеж 2	8	Раздел 8 Алгоритмы распознавания образов	8		6
	9	Раздел 9 Программное обеспечение интеллектуального анализа данных	4		4
		Рубежный контроль № 2			1
<b>Всего:</b>			<b>56</b>		<b>44</b>

## 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 класси-

фикаторы 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 поверхности 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. Практические занятия**

Номер раздела, темы	Наименование раздела, темы	Наименование практической работы	Норматив времени, час.
1	Раздел 1 Способы представления и модели порождения экспериментальных данных	Способы представления и модели порождения экспериментальных данных. Основные задачи прикладной статистики. Этапы статистического исследования.	
2	Раздел 2 Система предварительной статистической обработки данных	Система предварительной статистической обработки данных	4
3	Раздел 3 Задачи идентификации систем	Задачи идентификации систем.	4
		Линейная регрессия	4
	Рубежный контроль 1 (1-ый семестр)		1
4	Раздел 4 Методы редукции факторного пространства	Редукция факторного пространства.	4
5	Раздел 5 Методы планирования экспериментов	Методы планирования экспериментов.	6
		Рубежный контроль 2	1
6	Раздел 6 Задачи распознавания образов	Задачи распознавания образов.	6
7	Раздел 7 Классификаторы	Задача классификации	6
		Рубежный контроль 1	1
8	Раздел 8 Алгоритмы распознавания образов	Иерархический кластер анализ. Статистические методы.	6
9	Раздел 9 Программное обеспечение интеллектуального анализа данных	Пакеты прикладных программ	4
		Рубежный контроль № 2 (2 семестр)	1
<b>Всего:</b>			<b>48</b>

## **5 МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

Лекционный курс базируется на пассивном методе обучения, реализующем традиционную объяснительно-иллюстративную образовательную технологию, в рамках которой магистранты 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 практической 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 магистрантом в процессе самостоятельной проработки материала лекции при подготовке к очередному лекционному занятию.

Практические занятия 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 магистрантами, объединяемыми в малые группы (2-3 человека). Задания не имеют однозначного решения и соответствуют целям обучения.

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

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

Для текущего контроля успеваемости 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 практических занятиях в целях лучшего освоения материала и получения высокой оценки по результатам освоения дисциплины.

Выполнение самостоятельной работы подразумевает самостоятельное изучение разделов дисциплины, подготовку к практическим занятиям, рубежным контролям, подготовку к зачету, подготовку к экзамену

Рекомендуемая трудоемкость самостоятельной работы представлена в таблице.



## Рекомендуемый режим самостоятельной работы

Наименование вида самостоятельной работы	Рекомендуемая трудоемкость, акад. час.
Самостоятельное изучение тем дисциплины:	107
Раздел 1 Способы представления и модели порождения экспериментальных данных	18
Раздел 2 Система предварительной статистической обработки данных	18
Раздел 3 Задачи идентификации систем	10
Раздел 4 Методы редукции факторного пространства	14
Раздел 5 Методы планирования экспериментов	10
Раздел 6 Задачи распознавания образов	10
Раздел 7 Классификаторы	10
Раздел 8 Алгоритмы распознавания образов	9
Раздел 9 Программное обеспечение интеллектуального анализа данных	8
Подготовка к практическим занятиям	24
Контрольная работа	-
Подготовка к рубежным контролям (2 часа на каждый рубежный контроль) 2 семестра	4+4=8
Подготовка к зачету	18
Подготовка к экзамену	27
Всего:	184

## 6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

### 6.1. Перечень оценочных средств

1. Балльно-рейтинговая система контроля и оценки академической активности магистрантов в КГУ
2. Отчеты магистрантов по практическим работам.
3. Банк тестовых заданий к рубежным контролям № 1, № 2, №3, №4
4. Вопросы к зачету
5. Вопросы к экзамену

## 6.2. Система балльно-рейтинговой оценки работы обучающихся по дисциплине

№	Наименование	Содержание										
1	Распределение баллов за семестр по видам учебной работы, сроки сдачи учебной работы ( <i>доводятся до сведения магистрантов на первом учебном занятии</i> )	Распределение баллов за 1 семестр										
		<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 15%;">Вид учебной работы:</th> <th style="width: 15%;">Посещение лекций</th> <th style="width: 25%;">Выполнение и защита отчетов по практическим работам</th> <th style="width: 15%;">Рубежный контроль №1</th> <th style="width: 15%;">Рубежный контроль №2</th> <th style="width: 10%;">Зачет</th> </tr> </thead> <tbody> <tr> <td>Балльная оценка:</td> <td>0.56*16=86</td> <td>76*4=286 4пр-4часов 146*1=146 1пр-6часов 42</td> <td style="text-align: center;">10</td> <td style="text-align: center;">10</td> <td style="text-align: center;">30</td> </tr> </tbody> </table>	Вид учебной работы:	Посещение лекций	Выполнение и защита отчетов по практическим работам	Рубежный контроль №1	Рубежный контроль №2	Зачет	Балльная оценка:	0.56*16=86	76*4=286 4пр-4часов 146*1=146 1пр-6часов 42	10
Вид учебной работы:	Посещение лекций	Выполнение и защита отчетов по практическим работам	Рубежный контроль №1	Рубежный контроль №2	Зачет							
Балльная оценка:	0.56*16=86	76*4=286 4пр-4часов 146*1=146 1пр-6часов 42	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 (зачету) магистрант должен набрать не менее 50 баллов, выполнить и защитить все практические работы.</p> <p>Для получения экзаменационной оценки «автоматически» студенту необходимо набрать следующее минимальное количество баллов:</p> <ul style="list-style-type: none"> <li>- 61 для получения «автоматически» зачета.</li> </ul> <p>По согласованию с преподавателем магистранту могут быть добавлены дополнительные (бонусные) баллы за активность на практических занятиях, активное участие в научной и методической работе, оригинальность принятых решений в ходе выполнения практических работ, за участие в значимых учебных и внеучебных мероприятиях кафедры и выставлен зачет автоматически</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 (зачету) магистрант набрал сумму менее 50 баллов, то магистранту НЕОБХОДИМО набрать недостающие баллы и ВЫПОЛНИТЬ ДОПОЛНИТЕЛЬНЫЕ ЗАДАНИЯ до конца последней (зачетной) недели семестра. При этом необходимо проработать материал всех пропущенных практических работ.</p> <p>Формы дополнительных заданий (назначаются преподавателем):</p> <ul style="list-style-type: none"> <li>- выполнение и защита пропущенной практической работы (при невозможности дополнительного ее проведения преподаватель устанавливает форму дополнительного задания по тематике пропущенной практической работы самостоятельно) – до 5 баллов.</li> </ul> <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>										
1	Распределение баллов за	Распределение баллов за 2 семестр										

	семестр по видам учебной работы, сроки сдачи учебной работы ( <i>доводятся до сведения магистрантов на первом учебном занятии</i> )	Вид учебной работы:	Посещение лекций	Выполнение и защита отчетов по практическим работам		Рубежный контроль №1	Рубежный контроль №2	Экзамен
		Балльная оценка:	0.56*12=66	126*3=366 3пр-6часов 86*1=86 1пр-4часов 44		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) магистрант должен набрать не менее 50 баллов, выполнить и защитить все практические работы.</p> <p>Для получения экзаменационной оценки «автоматически» студенту необходимо набрать следующее минимальное количество баллов:</p> <ul style="list-style-type: none"> <li>- 68 для получения «автоматически» экзамена с оценкой удовлетворительно</li> </ul> <p>По согласованию с преподавателем магистранту, набравшему минимум 68 балл, могут быть добавлены дополнительные (бонусные) баллы за активность на практических занятиях, активное участие в научной и методической работе, оригинальность принятых решений в ходе выполнения практических работ, за участие в значимых учебных и вне учебных мероприятиях кафедры и выставлены оценки «хорошо» или «отлично» автоматически</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) магистрант набрал сумму менее 50 баллов, то магистранту НЕОБХОДИМО набрать недостающие баллы и выполнить ДОПОЛНИТЕЛЬНЫЕ ЗАДАНИЯ до конца последней (зачетной) недели семестра. При этом необходимо проработать материал всех пропущенных практических работ.</p> <p>Формы дополнительных заданий (назначаются преподавателем):</p> <ul style="list-style-type: none"> <li>- выполнение и защита пропущенной практической работы (при невозможности дополнительного ее проведения преподаватель устанавливает форму дополнительного задания по тематике пропущенной практической работы самостоятельно) – до 5 баллов.</li> </ul> <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: магистрант, 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 магистранту отводится 1 академический час.

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

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

**Зачет** проводится в традиционной (устной) форме: магистрант выполняет задания билета, включающего два теоретических вопроса, и отвечает преподавателю. Оцениваются полнота и правильность ответов магистранта на теоретические вопросы билета, его эрудиция в смежных вопросах.

Вопросы к зачету доводятся до магистрантов на последней лекции в семестре. Каждый вопрос оценивается в 15 баллов. На подготовку ответа магистранту отводится 1 астрономический час.

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

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

Вопросы к экзамену доводятся до магистрантов на последней лекции в семестре. Каждый вопрос оценивается в 10 баллов. На подготовку ответа магистранту 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 магистранта.

### **6.4 Примеры оценочных средств для рубежных контролей, зачета и экзамена**

1 семестр

#### **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

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

3. Какой метод не используется для решения задачи классификации:

- 1) метод главных компонент
- 2) факторный анализ
- 3) многомерное шкалирование
- 4) регрессионный анализ

4. 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 коррелированы сильно
- 4) разбиение факторного пространства на группы, где в каждой группе факторы коррелированы слабо

5. 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  $z_2$

6. 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

7. Основные задачи временных рядов:

- 1) сезонных и циклических составляющих анализируемого признака
- 2) тренда, сезонных и циклических составляющих анализируемого признака, случайных составляющих

1) выяснить какие из функций  $f, \varphi, \psi$  присутствуют, построить оценки этих функций

2) выбор алгоритмов, в зависимости от функции распределения

3) все вышеперечисленное

4) нет правильного ответа

8. 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

9. Как осуществляется проверка на стационарность?

1) по критерию серий

2) по критерию автоинформативности

3) по критерию внешней информативности

4) по критерию эргодичности

### Примеры тестовых заданий для рубежного контроля №2

1. 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

2. 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

3. 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) не правильного варианта ответа

4. 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

5. 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
- 6. 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
- 7. Для реализации синтеза управления нет необходимости иметь:
  - 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
- 8. Чем метод многомерного шкалирования отличается от факторного анализа:
  - 1) МНШ стремится извлечь больше факторов по сравнению с факторным анализом
  - 2) методы факторного анализа можно применять к любым типам состояний 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
- 9. 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
- 10. Выберите неверное утверждение:
  - 1) системы добычи данных в большей степени ориентированы на практическое приложение полученных результатов, чем на выяснение природы явления
  - 2) в области добычи данных принят такой подход к анализу данных и извлечению знаний, который иногда характеризуют словами "черный ящик"

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

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

## 2 семестр

### 6.4.2. Примеры тестовых заданий для рубежного контроля №1

1. Дайте определение разделяющей и решающей функции. Как определяется принадлежность образа к классу

1) через разделяющие функции; 2) через решающие функции.

2. Приведите примеры непараметрического и параметрического подходов к классификации образов, представленных набором параметров.

3. В каких случаях кластеризация только по пороговому ограничению может дать удовлетворительный результат?

4. Перечислите основные задачи корреляционного анализа при классификации многомерных данных.

5. Дайте определение ковариационной и корреляционной матрицы. Чем они отличаются?

6. С какой целью осуществляется переход к стандартизованной (нормализованной) матрице данных? Какими свойствами обладает эта матрица?

7. Приведите примеры задач, в которых применяется факторный анализ.

8. Какими свойствами обладают общие факторы в модели главных компонент? Что такое характерный фактор?

9. Какой показатель минимизируется (максимизируется) в задаче главных компонент?

### Примеры тестовых заданий для рубежного контроля №2

1. Перечислите основные этапы решения задачи главных компонент в общем виде? Как эта задача решается на практике?

2. Дайте графическую интерпретацию главных компонент для признакового пространства в задаче классификации многомерных данных.

3. В каких задачах распознавания целесообразно применение статистического подхода? Какие данные об объектах распознавания необходимы для его реализации?

4. Объясните содержательно понятие ошибок первого и второго рода. Как проявляются ошибки первого и второго рода при классификации пикселей изображения?

5. 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?



6. Какие стратегии принятия решения, основанные на байесовском подходе, используются:

1) при неизвестных априорных вероятностях; 2) при неизвестных априорных вероятностях и платежных коэффициентах?

7. С какой целью при классификации пикселей изображения Широко дисперсные классы часто разбивают на несколько подклассов?

8. Что такое апостериорная вероятность? Продемонстрируйте графически принцип принятия решения по максимуму правдоподобия в одномерном случае для трех классов с различными средними и дисперсиями.

9. В каком случае результат классификации по принципу максимума правдоподобия совпадает с результатом классификации по минимуму расстояния? Объясните это аналитически.

10. Перечислите основные способы предварительной оценки качества классификации по принципу максимума правдоподобия. В каких случаях целесообразно использовать каждый из этих способов?

#### Примерный перечень вопросов к зачету

1. Способы представления и модели порождения экспериментальных данных
2. Основные задачи прикладной статистики
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. Проверка априорных предположений о данных
12. Критерии согласия, однородности, симметрии независимости, стационарности. Пропущенные значения
13. Анализ резко выделяющихся наблюдений
14. Программное обеспечение подсистемы предварительной статистической обработки данных
15. Задачи идентификации систем. Уровни априорной информации. Типы моделей. Критерии качества
16. Линейный регрессионный анализ.
17. Метод наименьших квадратов
18. Гребневая, полиномиальная регрессии
19. Робастная регрессия
20. Непараметрические модели
21. Метод группового учета аргументов
22. Редукция факторного пространства

23. Метод главных компонент
24. Выбор наиболее информативных факторов
25. Методы планирования экспериментов
26. Критерии оптимальности планов.
27. Полный и дробный факторные эксперименты
28. Планы первого порядка
29. Центральные композиционные планы
30. Задачи распознавания образов
31. Архитектура автоматизированных систем научных исследований
32. Организация пользовательского интерфейса в автоматизированных системах обработки данных

### **Вопросы к экзамену**

1. Понятие образа.
2. Качественное описание задачи распознавания образов.
3. Типы задач распознавания и их характерные черты.
4. Структура системы распознавания образов.
5. Задача распознавания образов как одна из задач анализа данных.
6. Формальная постановка задачи распознавания образов.
7. Признаки и классификаторы.
8. Классификация с обучением и без обучения.
9. Решающие функции.
10. Классификация образов с помощью функций расстояния.
11. Классификация образов с помощью функций правдоподобия.
12. Обучаемые классификаторы образов.
13. Детерминистский подход.
14. Обучаемые классификаторы образов.
15. Статистический подход.
16. Методы распознавания, основанные на сравнении с эталоном.
17. Мера близости, основанная на поиске оптимального пути на графе.
18. Задача сравнения контуров.
19. Статистические методы.
20. Элементы теории статистических решений в распознавании образов.
21. Байесовский подход.
22. Дискриминантные функции и поверхности решения.
23. Алгоритм персептрона.
24. Классификация нейронных сетей.
25. Нейронные сети Хопфилда и Хэмминга.
26. Структурные и синтаксические методы.
27. . Структура системы распознавания образов.
28. Языки описания образов.

## 29. Обработка изображений.

### 6.5 Примеры задач к экзамену

#### Создать нейронную сеть.

##### 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).

*Задание №2.*

Используя алгоритмы:

- Максимины
- К-средних

выполнить кластеризация образов (Варианты 1-10).

Задание 2. Реализовать метод К-средних, минимакса (вариант-1-10)

Задание 3. Реализовать метод голосования (вариант-1-10).

## **6.6 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" - М.: Финансы и статистика, 1983.
3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. "Прикладная статистика. Исследования зависимостей", - М.: Финансы и статистика, 1985.
4. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. "Прикладная статистика. Классификация и снижение размерности", - М.: Финансы и статистика, 1989.

### **7.2 Дополнительная литература**

5. Тюрин 5. Ю.Н., Макаров А.А. "Анализ данных на компьютере", Под ред. В.Э. Фигурного. - М.: ИНФРА-М, Финансы и статистика, 1995

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

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

8. А.В. Духанов, О.Н. Медведева 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, 2016

## **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](http://informatikaplus.narod.ru)

– Сайт о высоких технологиях, новости индустрии из мира компьютерного «железа», тестовые испытания и обзоры оборудования [IXBT.com](http://IXBT.com).

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

-Система поддержки учебного процесса КГУ [dist.kgsu.ru](http://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 СТУДЕНТОВ, 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 сведения обучающихся.

Аннотация  
рабочей программы учебной дисциплины  
**Интеллектуальный анализ данных**

образовательной программы высшего образования –  
программы магистратуры

**09.04.04 – Программная инженерия**

Направленность:

**Методы и алгоритмы интеллектуальной обработки данных  
в информационно- вычислительных системах**

Формы обучения: **очная**

Трудоемкость дисциплины: 8 ЗЕ (288 академических часов)

Семестр: 1

Форма промежуточной аттестации: зачет

Семестр: 2

Форма промежуточной аттестации: экзамен

Содержание дисциплины

**Раздел 1 Способы представления и модели порождения эксперименталь-  
ных 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**