



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение высшего образования  
**«Дальневосточный федеральный университет»**  
(ДВФУ)  
ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ (ШКОЛА)

«СОГЛАСОВАНО»

Руководитель ОП

С.Л.Бедрина

«УТВЕРЖДАЮ»

Директор департамента Информационных и компьютерных систем

подпись  
«15» июля 2021 г.



**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**  
**Теоретические основы и технология обработки больших данных**  
**Направление подготовки 09.03.03 Прикладная информатика**  
(Прикладная информатика в экономике)  
**Форма подготовки очная**

курс 3, 4 семестр 6, 7  
лекции 36 час.  
практические занятия 00 час.  
лабораторные работы 38 час.  
в том числе с использованием МАО лек. 0 /пр. 0 /лаб. 34 час.  
всего часов аудиторной нагрузки 126 час.  
в том числе с использованием МАО 34 час.  
самостоятельная работа 182 час.  
в том числе на подготовку к экзамену 27 час.  
контрольные работы (количество) не предусмотрены  
курсовая работа / курсовой проект не предусмотрены  
зачет 6 семестр  
экзамен 7 семестр

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта по направлению подготовки 09.03.03 Прикладная информатика утвержденного приказом Министерства образования и науки РФ от 12 октября 2017 г. № 922 (с изменениями и дополнениями)


Рабочая учебная программа обсуждена на заседании департамента «Информационных и компьютерных систем», протокол № 7 от «09» июля 2021 г.

Директор департамента Информационных и компьютерных систем: д.ф.-м. н., доцент Пустовалов Е.В.  
Составитель ст.преподаватель Е.И.Шувалова\_

**Оборотная сторона титульного листа РПД**

**I. Рабочая программа пересмотрена на заседании департамента:**

Протокол от «17» сентября 2021 г. № 1

Директор департамента \_\_\_\_\_  \_\_\_\_\_ Пустовалов Е.В.  
(подпись) (И.О. Фамилия)

**II. Рабочая программа пересмотрена на заседании департамента:**

Протокол от «\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Директор департамента \_\_\_\_\_ \_\_\_\_\_  
(подпись) (И.О. Фамилия)

**III. Рабочая программа пересмотрена на заседании департамента:**

Протокол от «\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Директор департамента \_\_\_\_\_ \_\_\_\_\_  
(подпись) (И.О. Фамилия)

**IV. Рабочая программа пересмотрена на заседании департамента:**

Протокол от «\_\_\_\_\_» \_\_\_\_\_ 20\_\_ г. № \_\_\_\_\_

Директор департамента \_\_\_\_\_ \_\_\_\_\_  
(подпись) (И.О. Фамилия)

## 1. Цели и задачи освоения дисциплины:

**Цель:** изучение некоторых современных информационных технологий, предназначенных для интеллектуального анализа данных, направленных на формирование целостного представления об анализе и интерпретации экспериментальных и статистических данных, как о процессе поиска, так и применения скрытых в них закономерностей для достижения поставленных целей.

Дисциплина «Теоретические основы и технология обработки больших данных» логически и содержательно связана с такими курсами как «Программирование», «Основы теории управления», «Базы данных», «Информационные системы и технологии».

### Задачи

В результате освоения дисциплины студент должен:

- представления о феномене больших данных, о научных и технических проблемах и возможностях, связанных с их появлением, о трендах в области технологий хранения и анализа больших данных;

- знания причин возникновения тренда больших данных, процессов анализа больших данных, основных подходов к обработке больших массивов данных;

- умения формулировать алгоритмы, выбрать подходящий инструмент анализа больших данных, выбрать подходящую технологию хранения больших данных.

Для успешного изучения дисциплины «Теоретические основы и технология обработки больших данных» у обучающихся должны быть сформированы следующие предварительные компетенции:

- способность анализировать социально-экономические задачи и процессы с применением методов системного анализа и математического моделирования,

- способность использовать основные законы естественнонаучных дисциплин и современные информационно-коммуникационные технологии в профессиональной деятельности,

- способность программировать приложения и создавать программные прототипы решения прикладных задач,

- способность применять системный подход и математические методы в формализации решения прикладных задач.

В результате изучения данной дисциплины у студентов формируются следующие компетенции.

**Профессиональные компетенции выпускников и индикаторы их достижения:**

Задача профессиональной деятельности	Объект или область знания	Код и наименование профессиональной компетенции	Код и наименование индикатора достижения профессиональной компетенции
Тип задач профессиональной деятельности: производственно-технологический			
<p>Осуществление эксплуатации и сопровождения сервисов и информационных систем в целом и ее отдельных компонентов. Проведение работ по установке и тестированию программного обеспечения, загрузке баз данных, ведение технической документации. Начальное обучение и консультирование пользователей по вопросам эксплуатации информационных систем, ее компонентов и информационных сервисов.</p>	<p>Прикладные и информационные процессы Информационные системы Информационные технологии</p>	<p>ПК-5. Способность использовать информационные сервисы и международные информационные ресурсы для автоматизации прикладных и информационных процессов.</p>	<p>ПК 5.1. Знает назначение и классы интеллектуальных информационных систем; модели представления знаний; ПК 5.2. Умеет разрабатывать модель знаний прикладной области; выбирать инструментальные средства и технологии проектирования экспертных систем; создавать и сопровождать большие базы данных ПК 5.3. Владеет навыками работы с инструментальными средствами представления модели знаний предметной области</p>

Для формирования вышеуказанных компетенций в рамках дисциплины «Теоретические основы и технология обработки больших данных» применяются следующие методы активного/ интерактивного обучения: лекция-визуализация, мини-лекция, лабораторный метод (работа), творческое задание и работа в малых группах.

## **I. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА**

*6 Семестр (18часов)*

**Раздел I. Введение в интеллектуальный анализ больших данных ( 4 час.)**

**Тема 1. Большие данные. Хранилища данных ( 2 час.)**

Большие данные и проблемы их обработки. Методы представления больших данных. Концепция хранилищ данных. Организация Хранилищ

данных. Многомерная модель данных. Определение OLAP-системы. Архитектура OLAP-систем: MOLAP, ROLAP, HOLAP.

## **Тема 2. Введение в интеллектуальный анализ (Data Mining) ( 2 час.)**

Анализ данных методами Data Mining. Задачи Data Mining: классификация, регрессионный анализ, кластеризация, поиск ассоциативных правил. Области применения Data Mining. Модели Data Mining: предсказательные и описательные. Технологии и методы обработки больших данных: классические (классификация, регрессия, поиск ассоциативных правил, кластеризация) и интеллектуальные (нейронные сети, нечеткие или размытые (fuzzy) алгоритмы, генетические алгоритмы) подходы. Обнаружение знаний.

## **Раздел II. Мягкие вычисления в анализе больших данных ( 12 час.)**

### **Тема 1. Нейросетевые технологии анализа больших данных ( 4 час.)**

#### **Нейронные сети. Базовые понятия.**

Биологические представления о нейроне. Понятие нейрокомпьютера. Классификация нейронных сетей. Задача распознавания и линейная машина. Искусственный нейрон. Проблема линейной делимости. Правило обучения Хебба. Концепция входной и выходной звезды. Парадигмы обучения. Предварительная обработка информации и оценка качества работы нейросети.

#### **Однослойные нейронные сети.**

Описание искусственного нейрона. Персептрон. Линейная нейронная сеть. Рекуррентный метод наименьших квадратов. Линейная сеть с линией задержки.

#### **Нейронные сети прямого распространения.**

Топология и свойства. Алгоритм обратного распространения ошибки. Реализация логических функций. Аппроксимация функций. Распознавание символов. Моделирование статических зависимостей. Масштабирование и восстановление данных.

#### **Модели ассоциативной памяти.**

Нейронная сеть Элмана. Сети Хопфилда. Двухнаправленная ассоциативная память. Нейронная сеть Хэмминга. Адаптивные резонансные нейронные сети.

#### **Нейронные сети Кохонена.**

Структура сети Кохонена. Обучение сети Кохонена. Слой Кохонена. Самоорганизующиеся карты Кохонена. Нейронные сети классификации.

### **Радиальные нейронные сети.**

Структура радиальной нейронной сети. Расчет параметров радиальной нейронной сети. Обучение радиальной нейронной сети. Радиальные нейронные сети и нечеткие системы.

### **Методы прогнозирования на базе нейронных сетей**

### **Методы классификации и кластеризации на базе нейронных сетей**

## **Тема 2. Технологии размытой (нечеткой) логики в анализе больших данных ( 4 час.)**

Введение в теорию множеств. Операции над множествами.

Введение в теорию нечетких множеств. Операции над нечеткими множествами. Нечеткие отношения. Нечеткие отображения.

Нечеткость в описании прикладных задач. Нечеткость в принятии решений. Нечеткие множества в анализе данных.

## **Тема 3. Генетические алгоритмы ( 2 час.)**

Понятие о генетическом алгоритме. Этапы работы генетического алгоритма. Кодирование информации и формирование популяции. Оценивание популяции. Селекция. Скрещивание и формирование нового поколения. Мутация. Настройка параметров генетического алгоритма. Канонический генетический алгоритм. Пример работы генетического алгоритма. Рекомендации к программной реализации генетического алгоритма. Применение генетического алгоритма для решения задач оптимизации и аппроксимации.

## **Тема 4. Анализ текстовой информации ( 2 час.)**

Задачи анализа текстов: этапы анализа текстов, предобработка текстов. Извлечение ключевых понятий из текста: описание процесса извлечения понятий из текста; стадия локального анализа; стадия интеграции и вывода понятий.

Классификация текстовых документов: задачи и методы классификации текстов.

Методы кластеризации текстовых документов: представление текстовых документов, иерархические и бинарные методы кластеризации.

Задачи аннотирования текстов: выполнение аннотирования текстов; методы извлечения фрагментов для аннотации.

Средства анализа текстовой информации: Oracle Text (Oracle), Intelligent Miner for Text (IBM), Text Miner (SAS Institute), TextAnalyst (Мегакомпьютер Интеллидженс).

### **Раздел III. Распределенный анализ данных ( 2 час.)**

#### **Тема 1. Мультиагентный подход к распределенному анализу данных ( 2 час.)**

Системы мобильных агентов: стандарты мультиагентных систем, системы мобильных агентов.

Использование мобильных агентов для анализа данных: проблемы распределенного анализа данных, агенты-аналитики, варианты анализа распределенных данных).

Система анализа распределенных данных: общий подход, агент для сбора информации о базе данных, агент для сбора статистической информации о данных, агент для решения задачи интеллектуального анализа данных, агент для решения интегрированной задачи интеллектуального анализа данных.

*7 Семестр (18 часов)*

### **Раздел IV. Классические подходы Data Mining ( 18 час.)**

#### **Тема 1 Введение в интеллектуальный анализ (Data Mining) ( 2 час.)**

Анализ данных методами Data Mining. Задачи Data Mining: классификация, регрессионный анализ, кластеризация, поиск ассоциативных правил. Области применения Data Mining. Модели Data Mining: предсказательные и описательные. Технологии и методы обработки больших данных: классические (классификация, регрессия, поиск ассоциативных правил, кластеризация) и интеллектуальные (нейронные сети, нечеткие или размытые (fuzzy) алгоритмы, генетические алгоритмы) подходы. Обнаружение знаний.

**Deductor . Назначение. Общие сведения.**

#### **Тема 2 Задача классификации ( 4 час.)**

Постановка задачи. Представление результатов: правила классификации; деревья решений; математические функции.

Методы построения правил классификации: построение элементарных правил; методика Байеса.

Методы построения деревьев решений.

### **Тема 3 Регрессионный анализ ( 2 час.)**

Общий вид математических зависимостей. Простая и множественная линейная регрессия. Метод наименьших квадратов. Ограничения применимости регрессионных моделей.

Нелинейные модели.

Прогнозирование временных рядов: постановка задачи, методы прогнозирования временных рядов.

### **Тема 4 Поиск ассоциативных правил ( 4 час.)**

Постановка задачи, поиск последовательностей, обзор задач поиска ассоциативных правил. Представление результатов. Алгоритм Apriori: свойство антимонотонности, описание алгоритма.

### **Тема 5 Кластеризация ( 4 час.)**

Постановка задачи. Меры близости. Представление результатов. Алгоритмы кластеризации: базовые (иерархические (агломеративные и дивизимные), неиерархические (k-means, fuzzy c-means, Гюстафсона-Кесселя)) и адаптивные.

Модельные (алгоритм EM), концептуальные (алгоритм Cobweb), сетевые (метод WaveCluster) алгоритмы.

### **Тема 6 Стандарты Data Mining ( 2 час.)**

Обзор стандартов. Стандарт CWM: назначение, структура и состав. Стандарт CRISP: структура, фазы и задачи стандарта. Стандарт PMML. Другие стандарты: SQL/MM, OLE DB, JDMAPI.

## **II. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ**

### **КУРСА**

#### **Лабораторные работы ( 90 час.)**

*V Семестр (36 часов)* Применяются МАО лабораторный метод, творческое задание и работа в малых группах(**18 час.**)

**Лабораторная работа №1.** Создание модели многослойной нейронной сети, оценка параметров многослойной нейронной сети прямого распространения градиентным алгоритмом обучения (для построения статической модели) (4 час.)

Цель работы: получение навыков по формированию обучающей выборки и оценке параметров многослойной нейронной сети прямого



распространения градиентным алгоритмом обучения (для построения статической модели линейных и нелинейных объектов). Разработка программного обеспечения в среде Scilab (или другой).

**Лабораторная работа №2.** Создание модели рекуррентной нейронной сети (многослойной с обратными связями, сети Элмана), оценка параметров рекуррентной нейронной сети градиентным алгоритмом обучения (для построения динамической модели) (6 час.)

Цель работы: получение навыков по формированию обучающей выборки и оценке параметров рекуррентной нейронной сети градиентным алгоритмом обучения (для построения динамической модели линейных и нелинейных объектов). Разработка программного обеспечения в среде Scilab (или другой).

**Лабораторная работа №3.** Программные реализации моделей нейронных сетей (10 час.)

Цель работы: получение навыков по моделированию радиально базисных сетей, сетей Кохонена и их обучению (применительно к использованию решения задач кластеризации). Моделирование работы сетей Хэмминга и Хопфилда. Исследование стохастических методов обучения нейронных сетей. Разработка программного обеспечения в среде Scilab (или другой).

**Лабораторная работа №4.** Нечеткие системы (6 час.)

Цель работы: создание нечеткого описания системы, изучение операций над нечеткими множествами.

**Лабораторная работа №5.** Генетический алгоритм (6 час.)

Цель работы: моделирование работы генетических алгоритмов для решения задач оптимизации.

**Лабораторная работа №6.** Моделирование систем анализа распределенных данных в среде AnyLogic (4 час.)

Цель работы: моделирование систем анализа распределенных данных в среде AnyLogic.

*VI Семестр (54 час.)* Применяются МАО лабораторный метод, творческое задание и работа в малых группах(36 час.)

**Лабораторная работа №1 Знакомство с аналитической платформой Deductor. Начало работы с системой (4час.)**

Цель работы: получение навыков установки и работы с проектами.

**Лабораторная работа №2 Визуализация данных. (4час.)**

Цель работы: получение навыков работы с *Мастером визуализации*

**Лабораторная работа №3 Обработка данных в таблицах. (6час.)**

Цель работы: получение навыков обработки исходных данных

**Лабораторная работа №4 Работа с таблицами - сортировка записей, редактирование полей и данных. (4час.)**

Цель работы: получение навыков сортировки, замены, фильтрации данных.

**Лабораторная работа №5 Прогнозирование с помощью построения пользовательских моделей (4час.)**

Цель работы: получение навыков построения пользовательских моделей и прогнозирования с их использованием.

**Лабораторная работа №6 Классификация с помощью деревьев решений (4 час.)**

Цель работы: получение навыков решения задач классификации с помощью деревьев решений и прогнозирования с их использованием.

**Лабораторная работа №7 Кластеризация с помощью алгоритма k-means (4час.)**

Цель работы: получение навыков решения задач кластеризации с помощью алгоритма k-means и прогнозирования с использованием алгоритма.

**Лабораторная работа №8 Кластеризация с помощью самоорганизующейся карты Кохонена (4час.)**

Цель работы: получение навыков решения задач кластеризации с помощью самоорганизующейся карты Кохонена и построения прогноза.

**Лабораторная работа №9 Регрессионный анализ ( 6 час.)**

Цель работы: получение навыков прогнозирования социальных и экономических явлений.

**Лабораторная работа №10 Поиск ассоциативных правил (4час.)**

Цель работы: получение навыков нахождения закономерности между связанными событиями.

**Лабораторная работа №11 Разработка программы по ИЗ (10 час.)**

Цель работы: получение навыков создания программного прототипа решения прикладной задачи.

### **III. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ**

Самостоятельная работа заключается:

-в самостоятельной подготовке студента к лекции

–чтение конспекта предыдущих лекций и учебников, что помогает лучше понять материал новой лекции, в начале лекции проводится устный или письменный экспресс

-в подготовке к практическим (лабораторным) занятиям по основным и дополнительным источникам литературы, по разобранным задачам;

-в выполнении домашних заданий; в подготовке ответов на контрольные вопросы по темам;

-в самостоятельном изучении отдельных тем или вопросов по учебникам или учебным пособиям;

-в подготовке и выполнении контрольных мероприятий по дисциплине;

-в подготовке к выполнению и защите результатов лабораторных работ.

### **План организации самостоятельной работы**

<b>№ раздела дисциплины</b>	<b>Форма внеаудиторной самостоятельной работы</b>	<b>Примерные нормы времени на выполнение</b>	<b>Форма контроля</b>
1	Выполнение сформулированных письменных и устных домашних заданий; подготовка к лабораторным занятиям.	3	Отчет по л.р. и д.з.
2	Выполнение сформулированных письменных и устных домашних заданий; подбор статистических данных по теме; подготовка к лабораторным занятиям.	4	Отчет по л.р. и д.з.
3	Выполнение сформулированных письменных и устных домашних заданий; подбор статистических данных по теме; подготовка к лабораторным занятиям.	4	Отчет по л.р. и д.з., т ест.
4	Выполнение сформулированных письменных и устных домашних заданий; подготовка к лабораторным занятиям, подбор статистических данных по теме.	4	Отчет по л.р. и д.з. , т ест.
5	Выполнение сформулированных письменных и устных домашних заданий.	4	Отчет по л.р. и д.з.
6	Выполнение сформулированных письменных и устных домашних заданий; подбор статистических данных по теме; подготовка к лабораторным занятиям.	4	Отчет по л.р. и д.з.
7	Выполнение сформулированных письменных и устных домашних заданий; подготовка обзоров научной литературы, подготовка к лабораторным занятиям.	4	Отчет по л.р. и д.з. , т ест.
	Подготовка к экзамену	27	

## **IV. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА**

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций		Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Раздел I. Введение в интеллектуальный анализ больших данных	ПК-5	знает основные понятия, классификацию и характеристики больших данных	тест	зачет
			умеет реализовывать модели данных		
			владеет терминологией больших данных и навыками работы с инструментальными средствами представления модели знаний предметной области		
2	Раздел II. Мягкие вычисления в анализе больших данных	ПК-5	Знает методы синтеза и анализа нейросетевых, нечетких систем и генетические алгоритмы	Отчет по лабораторным работам	зачет
			умеет синтезировать и применять на практике нейросетевые, нечеткие системы анализа данных		
			владеет математическим аппаратом теории систем нейронных сетей, нечеткой логики		
3	Раздел III Распределенный анализ данных.	ПК-5	Знает методы мультиагентного моделирования	Отчет по лабораторным работам	зачет
			Умеет использовать программные продукты для анализа распределенных данных		
			Владеет математическим аппаратом описания распределенных систем анализа данных		
4	Раздел IV. Классические подходы Data Mining	ПК-5	Знает назначение и классы интеллектуальных информационных систем; модели представления знаний;	Отчет по лабораторным работам	экзамен
			умеет разрабатывать модель знаний прикладной области; выбирать инструментальные средства и технологии проектирования экспертных систем; создавать и сопровождать большие базы данных;		
			владеет - навыками работы с инструментальными средствами представления модели знаний предметной области.		

## **V. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### **Основная литература**

*(электронные и печатные издания)*

1 Базы знаний интеллектуальных систем : учебное пособие для вузов / Т. А. Гаврилова, В. Ф. Хорошевский. — Санкт-Петербург : Питер, 2001—382 с.(32 экз)

2 Ларичев, О.И. Теория и методы принятия решений: учебник / О.И. Ларичев. – М.: Логос, 2008. – 391 с.(10 экз)

### **Дополнительная литература**

*(печатные и электронные издания)*

1 Юре, Л. Анализ больших наборов данных [Электронный ресурс] / Л. Юре, Р. Ананд, Д.У. Джеффри ; пер. с англ. Слинкин А.А.. — Электрон. дан. — Москва : ДМК Пресс, 2016. — 498 с. — Режим доступа: <https://e.lanbook.com/book/93571> .

2 Павлов С.Н. Системы искусственного интеллекта. Часть 1,2 [Электронный ресурс] : учебное пособие / С.Н. Павлов. — Электрон. текстовые данные. — Томск: Томский государственный университет систем управления и радиоэлектроники, Эль Контент, 2011. — 176 с. — 978-5-4332-0013-5. — Режим доступа: <http://www.iprbookshop.ru/13974.html>

### **Перечень ресурсов информационно-телекоммуникационной сети «Интернет»**

- <http://alife.narod.ru/>
- <http://cyberleninka.ru/article/n/problema-izvlecheniya-znaniy-v-informatsionnyh-sistemah>
- Сайты, обнаруживаемые в Yandex.ru по запросам:
- "Интеллектуальная обработка данных (data mining)";
- "Поддержка принятия решений";
- "Экспертные системы".
- "Когнитивное моделирование";
- "Нейронные сети".

### **Перечень информационных технологий и программного обеспечения**

Программные продукты:

- - программный пакет SCILAB версии не ранее 5.5.0;
- - пакет Deductor Studio.
- - Visual Studio.
- - AnyLogic.

## **VI. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ**

**Рекомендации по планированию и организации времени, отведенного на изучение дисциплины.** Дисциплина направлена на формирование научного подхода к исследованию интеллектуальных систем анализа больших данных, что включает умение правильно организовать рабочее время обучающегося.

**Описание последовательности действий обучающихся, или алгоритм изучения дисциплины.** Изучение дисциплины предполагает поступательный подход по принципу усложнения, от знакомства с теорией и закрепление их на лабораторных работах

**Рекомендации по работе с литературой.** В связи с тем, что литература по данной дисциплине постоянно обновляется, рекомендуется ориентироваться на самые современные источники (журналы списка ВАК, Scopus, WebOfScience и др.). Тем не менее, нельзя игнорировать ставшие классическими исследования.

**Рекомендации по подготовке к экзамену.** Экзамен включает привычные экзаменационные вопросы.

## **VII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

Для проведения лабораторных занятий требуется компьютерный класс, оборудованный ЭВМ, с установленной на них операционной системой Windows версии не ранее XP. Требуемое программное обеспечение - Scilab (Matlab), пакет Deductor Studio, Visual Studio, AnyLogic.

Для проведения лекционных занятий требуется компьютерный класс, оборудованный мультимедийным оборудованием.

## **VIII. ФОНДЫ ОЦЕНОЧНЫХ СРЕДСТВ**

## Паспорт ФОС

Код и формулировка компетенции	Этапы формирования компетенции	
ПК-5. Способность использовать информационные сервисы и международные информационные ресурсы для автоматизации прикладных и информационных процессов	Знает	Знает назначение и классы интеллектуальных информационных систем; модели представления знаний
	Умеет	разрабатывать модель знаний прикладной области; выбирать инструментальные средства и технологии проектирования экспертных систем; создавать и сопровождать большие базы данных
	Владеет	навыками работы с инструментальными средствами представления модели знаний предметной области

№ п/п	Контролируемые разделы / темы дисциплины	Коды и этапы формирования компетенций	Оценочные средства		
			текущий контроль	промежуточная аттестация	
1	Раздел I. Введение в интеллектуальный анализ больших данных	ПК-5	знает основные понятия, классификацию и характеристики больших данных	тест	зачет
			умеет реализовывать модели данных		
			владеет терминологией больших данных и навыками работы с инструментальными средствами представления модели знаний предметной области		
2	Раздел II. Мягкие вычисления в анализе больших данных	ПК-5	Знает методы синтеза и анализа нейросетевых, нечетких систем и генетические алгоритмы	Отчет по лабораторным работам	зачет
			умеет синтезировать и применять на практике нейросетевые, нечеткие системы анализа данных		
			владеет математическим аппаратом теории систем нейронных сетей, нечеткой логики		
	Раздел III. Распределенный анализ данных.	ПК-5	Знает методы мультиагентного моделирования	Отчет по лабораторным работам	зачет
			Умеет использовать программные продукты для анализа распределенных данных		

			Владеет математическим аппаратом описания распределенных систем анализа данных		
	Раздел IV. Классические подходы Data Mining	ПК-5	Знает назначение и классы интеллектуальных информационных систем; модели представления знаний;	Отчет по лабораторным работам	экзамен
			умеет разрабатывать модель знаний прикладной области; выбирать инструментальные средства и технологии проектирования экспертных систем; создавать и сопровождать большие базы данных;		
			владеет - навыками работы с инструментальными средствами представления модели знаний предметной области.		

### Перечень оценочных средств

Код ОС	Наименование оценочного средства	Краткая характеристика оценочного средства
ПР-1	Тест	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося.
ПР-2	Контрольная работа	Средство проверки умений применять полученные знания для решения задач определенного типа по теме или разделу
ПР-6	Лабораторная работа	Средство для закрепления и практического освоения материала по определенному разделу.

### Критерии оценочного средства Отчет по лабораторной работе

Балл (интервал баллов)	Уровень освоения	Критерии оценивания уровня освоения компетенций*
4	Максимальный уровень (интервал)	Лабораторная работа выполнена в полном соответствии с требованиями, студент представил отчет без погрешностей и замечаний, на все вопросы при защите лабораторной работы дал правильные ответы.
[3.1; 3.9]	Средний уровень (интервал)	Лабораторная работа выполнена в полном соответствии с требованиями, студент представил отчет с небольшими погрешностями в оформлении и/или реализации требований к составу



		описаний, на защите затруднялся при ответах на некоторые вопросы, нуждался в уточняющих вопросах и подсказках со стороны преподавателя
3	Минимальный уровень (интервал)	Лабораторная работа выполнена в соответствии с требованиями, студент представил отчет с существенными погрешностями в оформлении, неспособен правильно интерпретировать полученные результаты, на защите затруднялся и/или не ответил на большинство вопросов, нуждался в уточняющих вопросах и подсказках со стороны преподавателя
0	Минимальный уровень (интервал) не достигнут.	студент самостоятельно выполнил лабораторную работу, неспособен пояснить содержание отчета, не ответил ни на один контрольный вопрос на защите

Программой дисциплины предусмотрены лабораторные работы. По каждой лабораторной работе студенту выдаётся задание, которое он должен выполнить за текущее занятие.

Рейтинговый контроль по лабораторным работам производится при их сдаче во время лабораторных занятий.

### **Методические рекомендации, определяющие процедуры оценивания результатов освоения дисциплины**

#### **Вопросы к зачету**

1. Что такое искусственная нейронная сеть? Что такое нейрокомпьютер и нейрокомпьютинг?
2. Какие принципы используются при классификации нейронных сетей?
3. Каковы три основные парадигмы обучения нейронных сетей? Чем отличается детерминированное обучение от стохастического?
4. Какие операции могут выполняться при предварительной обработке обучающих данных для нейросети?
5. Как оценить качество обучения нейросети?
6. Как используется МНК при обучении линейной нейронной сети? Как оценить качество обучения линейной ИНС при использовании МНК? Возможно ли использование МНК при обучении персептрона?
7. Как можно использовать нейронные сети ADALIN? Какие существуют ограничения при использовании ADALIN? Что представляют собой нейронные сети MADALINE?

8. Каковы свойства искусственных нейронных сетей прямого распространения? Каким свойством должна обладать ИНС прямого распространения, чтобы быть универсальным аппроксиматором?

9. Какова последовательность действий при использовании алгоритма обратного распространения ошибки? В чем состоит явление переобучения НС?

10. Что собой представляет градиентный метод поиска минимума функции?

11. В чем заключается основной недостаток алгоритма обратного распространения ошибки для обучения нейронных сетей?

12. Какова структура нейронной сети Элмана? Какие активационные функции используются в разных слоях сети Элмана?

13. Какую активационную функцию используют нейроны сети Хопфилда? Какова структура нейронной сети Хэмминга? Как вычисляется расстояние Хэмминга? Какие активационные функции используются в однослойной сети Хэмминга?

14. Какие задачи решают нейронные сети Кохонена? Сколько слоев имеет нейронная сеть Кохонена? Какие существуют варианты сетей Кохонена? Что такое топология карты Кохонена? Какие бывают варианты топологий? Каков механизм реализации обучения без учителя в сетях Кохонена? Какие проблемы могут возникать при обучении сетей Кохонена?

15. Какова основная идея использования генетического алгоритма для глобальной оптимизации? Что представляет собой хромосома при генетическом синтезе нейронной сети? Какие существуют генетические операторы? Что такое пригодность хромосомы и как ее можно оценить? В чем измеряется время работы генетического алгоритма?

16. Сколько слоев содержит RBF-сеть? Что представляет собой радиально-базисная функция? Является ли RBF-сеть универсальным аппроксиматором? Каковы возможные варианты радиально-базисных функций?

17. В чем состоит главное отличие RBF-сетей от обычных многослойных сетей прямого распространения?

18. Какую активационную функцию имеют нейроны выходного слоя RBF-сети?

19. Какой метод используется для приближенного выбора коэффициентов RBF-сети?

20. Какие параметры RBF-сети полагаются заданными при использовании метода наименьших квадратов?

21. С какой целью используется кластеризация при обучении RBF-сети?
22. Какие задачи можно решить с помощью генетического алгоритма при обучении RBF-сети?
23. Каковы принципы использования генетического алгоритма при обучении RBF-сети? Что представляют собой сети GRNN? Что представляют собой сети PNN?
24. Возникновение нечеткой логики. Нечеткие множества. Сущность и определения.
25. Дайте определение понятий «лингвистическая переменная» и «нечеткое множество», поясните их на примере. Какие операции можно выполнять над нечеткими множествами? Рассмотрите математические операции над нечеткими множествами: дополнение, объединение и пересечение множеств.
26. Операции над нечеткими множествами. Логические операции.
27. Операции над нечеткими множествами. Алгебраические операции. Построение функций принадлежности. Нечеткие и лингвистические переменные.
28. Дайте определение нечеткого отношения и расскажите о свойствах нечетких отношений. Каким образом, и с какой целью используются нечеткие отношения в ИИС?
29. Что такое нечеткая импликация? Какими способами ее можно реализовать для правил с одним выходом и двумя выходами. Приведите примеры.
30. Расскажите о нечетком логическом выводе. Дайте определение композиции нечетких отношений. Приведите пример нечеткой композиции.
31. Нечеткие алгоритмы и выводы. Расскажите об организации ЭС с нечетким логическим выводом. Какие способы используются в таких системах для представления и обработки знаний?
32. Программные пакеты в области нечеткой логики.
33. Среда моделирования Anylogic. Основные инструменты создания и работы с имитационной моделью.
34. Агентные модели. Особенности создания. Дать определение агента и описать его свойства.
35. Что называется многоагентной системой? Какие динамические свойства системы обнаруживаются в результате взаимодействия агентов?
36. Какие элементы динамической модели предприятия можно описать с помощью агентного моделирования? Области приложения агентного моделирования.

## Вопросы к экзамену

- 1 Определите понятие нечеткой логики.
- 2 Охарактеризуйте функцию принадлежности.
- 3 Опишите понятие дефазификации нечеткого множества.
- 4 Каким образом коэффициент уверенности выражается через меры доверия и недоверия?
- 5 Приведите соотношение между мерами доверия, полученными при независимом учете первого и второго свидетельства и объединенной мерой доверия, полученной при учете двух свидетельств.
- 6 Охарактеризуйте нечеткие правила вывода в экспертных системах.
- 7 Опишите структуру генетического алгоритма.
- 8 Охарактеризуйте целочисленное и вещественное кодирование в ГА.
- 9 Опишите канонический генетический алгоритм.
- 10 Охарактеризуйте операторы кроссовера и мутации.
- 11 Опишите биологические нейронные сети.
- 12 Охарактеризуйте понятие формального нейрона.
- 13 Опишите существующие модели искусственных нейронных сетей.
- 14 Приведите примеры активационных функций.
- 15 Опишите процесс обучения нейронной сети.
- 16 Интеллектуальный анализ данных (data mining).
- 17 Типы выявляемых закономерностей data mining: ассоциация, последовательность, классификация, кластеризация, прогнозирование.
- 18 Математический аппарат data mining: деревья решений, регрессионный анализ, нейронные сети, временные ряды.
- 19 Области применения технологий интеллектуального анализа данных: розничная торговля, банковская деятельность, страховой бизнес, производство, автоматизированные системы для интеллектуального анализа данных.

### Критерии выставления оценки студенту на зачете/ экзамене по дисциплине «Теоретические основы и технология обработки больших данных»:

Баллы (рейтинговой оценки)	Оценка экзамена	Требования к сформированным компетенциям
86-100	зачтено/отлично	Оценка «отлично» выставляется студенту при: - глубоком и прочном усвоении материала курса «Теоретические основы и технология обработки больших данных», - исчерпывающе, последовательно, четко и

		<p>логически стройно излагает материал,</p> <ul style="list-style-type: none"> <li>- умеет увязывать теорию с практикой,</li> <li>- использует в ответе материал монографической литературы,</li> <li>- правильно обосновывает принятое решение.</li> </ul>
76-85	зачтено/ хорошо	<p>Оценка «хорошо» выставляется студенту при:</p> <ul style="list-style-type: none"> <li>- твердом знании материала курса «Теоретические основы и технология обработки больших данных»,</li> <li>- грамотно и по существу излагает материал, не допуская существенных неточностей в ответе на вопрос,</li> <li>- правильно применяет теоретические положения при решении практических вопросов и задач,</li> <li>- владеет необходимыми навыками и приемами их выполнения.</li> </ul>
61-75	зачтено/ удовлетворительно	<p>Оценка «удовлетворительно» выставляется студенту при:</p> <ul style="list-style-type: none"> <li>- знания только основного материала, но не усвоения его деталей,</li> <li>- допуске неточностей, недостаточно правильно сформулированном изложении материала курса,</li> <li>- затруднении при выполнении лабораторных работ.</li> </ul>
<60	не зачтено/ неудовлетворительно	<p>Оценка «неудовлетворительно» выставляется студенту:</p> <ul style="list-style-type: none"> <li>- который не знает значительной части материала,</li> <li>- допускает существенные ошибки,</li> <li>- неуверенно, с большими затруднениями выполняет лабораторные работы.</li> </ul>