

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Макаренко Елена Николаевна

Должность: Ректор

Дата подписания: 28.04.2023 17:24:48

Уникальный программный ключ:

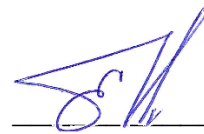
c098bc0c1041cb2a4cf926cf171d6715d99a6ae00adc8e27b55cbe1e2dbd7c78

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования «Ростовский государственный экономический университет (РИНХ)»

УТВЕРЖДАЮ

Директор Института магистратуры



Иванова Е.А.

«22» февраля 2022 г.

Рабочая программа дисциплины

Математические методы анализа больших данных

Направление 09.04.03 Прикладная информатика

магистерская программа

09.04.03.03 Машинное обучение и технологии больших данных

Для набора 2022 года

Квалификация

магистр

Кафедра Фундаментальной и прикладной математики

Составители рабочей программы:

к.ф.-м.н., доцент Богачев Тарас Викторович

СОДЕРЖАНИЕ

I. Цели и задачи освоения дисциплины	4
II. Место дисциплины в структуре образовательной программы	4
III. Требования к результатам освоения дисциплины	6
IV. Содержание и структура дисциплины	7
4.1. Содержание дисциплины, структурированное по темам	7
4.2. План внеаудиторной самостоятельной работы	8
4.3. Содержание учебного материала	10
V. Образовательные технологии	12
VI. Учебно-методическое обеспечение дисциплины	12
6.1. Основная литература	12
6.2. Дополнительная литература	12
6.3. Периодические издания	13
6.4. Перечень ресурсов сети Интернет	13
VII. Материально-техническое обеспечение дисциплины	13
VIII. Методические указания для обучающихся по освоению дисциплины	14
IX. Учебная карта дисциплины	15
X. Фонд оценочных средств	16
10.1. Паспорт фонда оценочных средств	16
10.2. Вопросы для собеседования	16
10.3. Практикум	17
10.4. Экзаменационные вопросы и билеты	20

I. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цели освоения дисциплины:

- формировании у студентов способности применять математические методы и модели для обработки и анализа больших данных при решении профессиональных задач

Задачи освоения дисциплины:

- развитие у обучающихся умения выбирать и применять математические методы для решения задач анализа больших данных.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина относится к модулю обязательных профессиональных дисциплин обязательной части образовательной программы.

Для изучения данной дисциплины необходимы знания, умения и навыки, формируемые предшествующими элементами образовательной программы:

Наименование дисциплины (модуля), практики	Требуемые знания, умения, навыки
Методы машинного обучения	<p><i>Знания:</i></p> <ul style="list-style-type: none">– Знает классы методов и алгоритмов машинного обучения.– Знает возможности современных инструментальных средств и систем программирования для решения задач машинного обучения.– Знает функциональность современных инструментальных средств и систем программирования в области создания моделей и методов машинного обучения. <p><i>Умения:</i></p> <ul style="list-style-type: none">– Умеет ставить задачи и адаптировать методы и алгоритмы машинного обучения.– Умеет проводить сравнительный анализ и осуществлять выбор инструментальных средств для решения задач машинного обучения.– Умеет применять современные инструментальные средства и системы программирования для разработки новых методов и моделей машинного обучения.
Современные проблемы и методы прикладной информатики	<p><i>Знания:</i></p> <ul style="list-style-type: none">– Современного состояния и тенденций научно-технического развития информационного общества;– Направлений современных исследований и аналитические инструменты в прикладной информатике;– Основных тенденции и прогноза научно-технического развития в области информационно-коммуникационных технологий.– Современных методов и средств информатики для решения прикладных задач;– Научные фронтиры в области компьютерных наук: последние достижения, современные вызовы и открытые вопросы.– Знает содержание, объекты и субъекты информационного общества и цифровой экономики, критерии эффективности функционирования информационного общества, теоретические проблемы информатики, искусственного интеллекта, современные методы, средства, стандарты информатики для решения прикладных задач различных классов; правовые, экономические, социальные и психологические аспекты информатизации деятельности организационно-экономических систем.– Знает состав современных методов и средств информатики,

Наименование дисциплины (модуля), практики	Требуемые знания, умения, навыки
	<p>передовые методы искусственного интеллекта для решения задач профессиональной деятельности.</p> <p><i>Умения:</i></p> <ul style="list-style-type: none"> – Анализировать особенности и состояние современного информационного общества и пути его развития. – Анализировать возможности и выбирать современные методы и средства информатики для решения прикладных задач. – Умеет применять при решении задач профессиональной деятельности критерии эффективности функционирования информационного общества и цифровой экономики; структуру интеллектуального капитала, методы оценки эффективности. – Умеет проводить анализ современных методов и средств информатики и искусственного интеллекта для решения прикладных задач различных классов. <p><i>Навыки:</i></p> <ul style="list-style-type: none"> – Применения аналитических инструментов и методов для исследования современного состояния и тенденций научно-технического развития информационного общества. – Обобщения результатов проведенного анализа и исследования.

Знания, умения и навыки, формируемые данной дисциплиной, потребуются при освоении следующих элементов образовательной программы:

- Технологии анализа больших данных;
- Интеллектуальные геоинформационные системы;
- Системы аналитики больших данных;
- Математические методы и модели поддержки принятия решений;
- Нейронные сети и глубокое обучение;
- производственная практика, проектно-технологическая практика;
- производственная практика, преддипломная практика.

III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины направлено на формирование следующих компетенций в соответствии с образовательной программой:

Перечень планируемых результатов обучения по дисциплине, соотнесённых с индикаторами достижения компетенций

Компетенция	Индикаторы достижения компетенции	Результаты обучения
ПК-1. Способен адаптировать и применять методы и алгоритмы машинного обучения для решения прикладных задач в различных предметных областях	ПК-1.1. Ставит задачи по адаптации или совершенствованию методов и алгоритмов для решения комплекса задач предметной области	<i>Знания:</i> <ul style="list-style-type: none">– Знает классы методов и алгоритмов машинного обучения.– Знает математические модели, методы и алгоритмы для обработки и анализа больших данных. <i>Умения:</i> <ul style="list-style-type: none">– Умеет ставить задачи и адаптировать методы и алгоритмы машинного обучения.– Умеет выбирать и применять математические модели, методы и алгоритмы для решения прикладных задач анализа больших данных.
ПК-3. Способен руководить проектами со стороны заказчика по созданию, поддержке и использованию системы искусственного интеллекта на основе нейросетевых моделей и методов	ПК-3.1. Руководит проектами по разработке, систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов со стороны заказчика	<i>Знания:</i> <ul style="list-style-type: none">– Знает принципы построения моделей глубоких нейронных сетей и глубокого машинного обучения.– Знает подходы к применению моделей на основе нечеткой логики в системах искусственного интеллекта. <i>Умения:</i> <ul style="list-style-type: none">– Умеет руководить выполнением коллективной проектной деятельности для создания, поддержки и использования систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов.

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоёмкость дисциплины составляет 7 зачётных единиц, 252 часа,
в том числе 1 зачётная единица, 36 часов на экзамен

Форма промежуточной аттестации: экзамен

4.1. Содержание дисциплины, структурированное по темам

№ п/п	Темы дисциплины	Семестр	Виды учебной работы и их трудоёмкость, часы (в том числе с использованием онлайн-курсов)				Наименования оценочных средств
			Контактная работа			Самостоя- тельная работа	
			Лекции	Практические занятия	Лабораторные занятия		
Модуль 1. Математические методы машинного обучения для анализа больших данных							
1	Метрические алгоритмы. Байесовский подход к обучению	2	2	2	-	20	Работа на практических занятиях (вопросы для собеседования);
2	Линейные алгоритмы классификации. Метод опорных векторов	2	2	-	-	20	Работа на практических занятиях (вопросы для собеседования);
3	Методы восстановления регрессии. Логические методы классификации. Композиции алгоритмов	2	-	2	-	20	Работа на практических занятиях (вопросы для собеседования);
4	Ранжирование и рекомендательные системы. Тематическое моделирование. Кластеризация.	2	2	2		20	Работа на практических занятиях (вопросы для собеседования);

Модуль 2. Искусственные нейронные сети							
5	Введение в нейронные сети	2	2	2	-	16	Практикум
6	Искусственные нейронные сети	2	2	-	-	16	Практикум
7	Однослойные и многослойные перцептроны	2	-	2	-	16	Практикум
8	Сети на основе радиальных базисных функций.	2	2	2		16	Практикум
9	Анализ главных компонент	2	2	-		16	Практикум
10	Карты самоорганизации Кохонена	2	-	2		12	Практикум
11	Нейродинамические модели	2	2	2		12	Практикум
Промежуточная аттестация (для дисциплин с экзаменом)		2	-	-	-	36	Экзаменационные вопросы и билеты
Итого часов			16	16	-	220	-

4.2. План внеаудиторной самостоятельной работы

№ п/п	Темы дисциплины	Семестр	Вид самостоятельной работы	Сроки выполнения (нед.)	Затраты времени (часы)	Учебно-методическое обеспечение
Модуль 1. Математические методы машинного обучения для анализа больших данных						
1	Метрические алгоритмы. Байесовский подход к обучению	2	– проработка и повторение материала лекционных занятий; – подготовка к практическим занятиям	1–2	20	[1]-[3]
2	Линейные алгоритмы классификации. Метод опорных векторов	2	– проработка и повторение материала лекционных занятий; – подготовка к практическим занятиям	3-4	20	[1]-[3]
3	Методы восстановления регрессии. Логические методы классификации. Композиции алгоритмов	2	– проработка и повторение материала лекционных занятий; – подготовка к практическим занятиям	5-6	20	[1]-[3]
4	Ранжирование и рекомендательные системы. Тематическое моделирование. Кластеризация.	2	– проработка и повторение материала лекционных занятий; – подготовка к практическим занятиям – подготовка к контрольной работе №1	7-10	20	[1]-[3]

Модуль 2. Искусственные нейронные сети

5	Введение в нейронные сети	2	– проработка и повторение материала лекционных занятий; – выполнение практикума	11	16	[1]-[3]
6	Искусственные нейронные сети	2	– проработка и повторение материала лекционных занятий; – выполнение практикума	12	16	[1]-[3]
7	Однослойные и многослойные перцептроны	2	– проработка и повторение материала лекционных занятий; – выполнение практикума	13	16	[1]-[3]
8	Сети на основе радиальных базисных функций.	2	– проработка и повторение материала лекционных занятий; – выполнение практикума	14	16	[1]-[3]
9	Анализ главных компонентов	2	– проработка и повторение материала лекционных занятий; – выполнение практикума	15	16	[1]-[3]
10	Карты самоорганизации Кохонена	2	– проработка и повторение материала лекционных занятий; – выполнение практикума	16	12	[1]-[3]
11	Нейродинамические модели	2	– проработка и повторение материала лекционных занятий; – выполнение практикума	17-18	12	[1]-[3]
Подготовка к экзамену (для дисциплин с экзаменом)					36	[1]-[3]
Общая трудоёмкость самостоятельной работы по дисциплине					220	–

4.3. Содержание учебного материала

Модуль 1 «Математические методы машинного обучения для анализа больших данных»

Название раздела	Содержание раздела
Метрические алгоритмы. Байесовский подход к обучению	Обобщенный алгоритм. Примеры частных алгоритмов. Метод ближайших соседей. Метод окна Парзена. Понятие выступа объекта. Алгоритм отбора эталонов. Проклятие размерности. Выбор метрики. Применение формула Байеса к задаче машинного обучения. Построение приближенной плотности распределения: одномерный случай (дискретное, кусочно-постоянное, приближение с помощью ядра, приближение нормальным законом распределения), многомерный случай (наивный байесовский классификатор, приближение с помощью ядра, многомерное нормальное распределение). Логистическая регрессия. Бинаризация признаков. Скоринг. Смеси распределений. EM-алгоритм восстановления смеси.
Линейные алгоритмы классификации. Метод опорных векторов	Общая формула линейного классификатора. Метод стохастического градиента. Частные случаи. Обоснование метода СГ. Выступ объекта для линейного классификатора. ROC и AUC. Случаи линейно разделимой и неразделимой выборки. Двойственная задача. Типы объектов. Нелинейное обобщение SVM. SVM-регрессия. L1 регуляризация
Методы восстановления регрессии. Логические методы классификации. Композиции алгоритмов	Метод наименьших квадратов. Геометрический смысл. Регуляризация. Сингулярное разложение. Непараметрическая регрессия. Понятие закономерности. Критерий качества закономерностей. Поиск закономерностей. Алгоритмы классификации на основе логических закономерностей. AdaBoost. AnyBoost. Градиентный бустинг. Бэггинг и метод случайных подпространств
Ранжирование и рекомендательные системы. Тематическое моделирование. Кластеризация.	Оценки качества. Подходы к решению задачи: поточечный, попарный, списочный. Латентный семантический анализ (LSA). Вероятностный LSA (PLSA). Латентное размещение Дирихле (LDA). EM-алгоритм. Метод k- средних. DBSCAN

Модуль 2 «Искусственные нейронные сети»

1. Введение

Биологические предпосылки возникновения искусственных нейронных сетей. Структура человеческого мозга. Организация памяти в коре человеческого мозга. Ритмы колебаний больших нейронных ансамблей. Биологически правдоподобные модели нейронов. Модели

визуального восприятия.

2. Искусственные нейронные сети

Типы функций активации нейронов. Представление нейронных сетей с помощью направленных графов. Архитектура сетей. Сети прямого распространения. Рекуррентные сети. Обучение нейронных сетей. Обучение, основанное на коррекции ошибок. Обучение на основе памяти. Обучение Хебба. Математические модели предложенного Хеббом механизма модификации синаптической связи. Конкурентное обучение. Обучение Больцмана. Обучение с учителем. Обучение с подкреплением. Обучение без учителя.

3. Однослойные и многослойные перцептроны

Однослойный перцептрон. Обучение перцептрона. Методы безусловной оптимизации. Метод наискорейшего спуска. Метод Ньютона. Метод Гаусса-Ньютона. Взаимосвязь перцептрона и байесовского классификатора. Многослойный перцептрон. Алгоритм обратного распространения ошибки. Извлечение признаков. Линейный дискриминант Фишера. Сети свертки.

4. Сети на основе радиальных базисных функций.

Теорема Ковера о разделимости множеств. Разделяющая способность поверхности. Задача интерполяции. Теория регуляризации. Функция Грина. Решение задачи регуляризации. Многомерные функции Гаусса. Обобщенные сети на основе радиальных базисных функций. Свойства аппроксимации сетей RBF. Сравнение сетей RBF и многослойных перцептронов.

5. Анализ главных компонент

Анализ признаков на основе самоорганизации. Структура анализа главных компонент. Представление данных. Сокращение размерности. Фильтр Хебба для выделения максимальных собственных значений. Анализ главных компонент на основе правила Хебба.

6. Карты самоорганизации Кохонена

Модели отображения признаков. Карты самоорганизации. Процессы конкуренции, кооперации и адаптации. Варианты самоорганизующихся карт. Адаптивные тензорные веса. Самоорганизующиеся карты для символьных строк. Самоорганизующиеся карты с эволюционным обучением. Пакеты программ, реализующие самоорганизующиеся карты.

7. Нейродинамические модели

Динамические системы. Пространство состояний. Условие Лившица. Теорема о дивергенции. Устойчивость состояний равновесия. Теоремы Ляпунова. Гиперболические аттракторы. Аддитивная и связанная нейродинамические модели. Модель Хопфилда. Теорема Коэна-Гроссберга.

Перечень тем практических занятий

№ п/п	Тема практического занятия	Количество часов
Модуль 1. Математические методы машинного обучения для анализа больших данных		
1	Метрические алгоритмы. Байесовский подход к обучению	2
2	Методы восстановления регрессии. Логические методы классификации. Композиции алгоритмов	2
3	Ранжирование и рекомендательные системы. Тематическое моделирование. Кластеризация.	2
Модуль 2. Искусственные нейронные сети		

№ п/п	Тема практического занятия	Количество часов
4	Введение в нейронные сети	2
5	Однослойные и многослойные перцептроны	2
6	Сети на основе радиальных базисных функций.	2
7	Карты самоорганизации Кохонена	2
8	Нейродинамические модели	2
Всего часов		16

V. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Наряду с традиционными образовательными технологиями, для реализации дисциплины будут использоваться технологии электронного обучения и дистанционные образовательные технологий в электронной информационно-образовательной среде Южного федерального университета.

Дисциплина может быть реализована с использованием электронного обучения и дистанционных образовательных технологий (ЭО и ДОТ) в ЭИОС Университета, включая систему электронного обучения ИКТИБ (lms.sfedu.ru). Аудиторные занятия и другие формы контактной работы обучающихся с преподавателем могут проводиться с использованием платформ Microsoft Teams, Cisco, Skype, Google Classroom, Zoom и др., что позволяет обеспечить онлайн и офлайн взаимодействие преподавателя с обучающимися в рамках дисциплины

Основными методами текущего контроля являются электронный учёт и контроль учебных достижений студентов (использование средств сервиса балльно-рейтинговой системы; ведение электронного журнала успеваемости, проведение электронного тестирования и применение других средств контроля с использованием системы электронного обучения).

VI. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Основная литература

1. Келлехер, Д. Наука о данных: базовый курс : [16+] / Д. Келлехер, Б. Тирни ; науч. ред. З. Мамедьяров ; пер. с англ. М. Белоголового. – Москва : Альпина Паблишер, 2020. – 224 с. : схем., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=598235>

2. Трофимова Е.А. Нейронные сети в прикладной экономике : учебное пособие / Трофимова Е.А., Мазуров Вл.Д., Гилёв Д.В.. — Екатеринбург : Издательство Уральского университета, 2017. — 96 с. — ISBN 978-5-7996-2018-9. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/106462.html>

3. Пролубников, А. В. Математические методы распознавания образов : учебное пособие : [16+] / А. В. Пролубников. – Омск : Омский государственный университет им. Ф.М. Достоевского, 2020. – 110 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=614061>

6.2. Дополнительная литература

4. Крутиков В. Н. Анализ данных / В.Н. Крутиков; В.В. Мешечкин - Кемерово: Кемеровский государственный университет, 2014. – 138 с. <http://biblioclub.ru/index.php?page=book&id=278426>

5. Балдин, К. В. Теория вероятностей и математическая статистика : учебник / К. В. Балдин, В. Н. Башлыков, А. В. Рукосуев. – 3-е изд., стер. – Москва : Дашков и К°, 2020. – 472 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=573173>

6.3. Периодические издания

- IEEE Spectrum <https://spectrum.ieee.org/>
- Научный журнал «Машинное обучение и анализ данных» <http://jmla.org/ru/journal>

6.4. Перечень ресурсов сети Интернет

- ЭБС IPR Books <http://www.iprbookshop.ru/>
- ЭБС «Университетская библиотека онлайн» <http://biblioclub.ru>.
- Образовательная платформа Юрайт <https://urait.ru/>
- IBM Academic Initiative http://ictis.sfedu.ru/ibm_academic_initiative/ (учебные материалы)
- <http://github.com/>
- <http://habr.com/>
- <http://www.kdnuggets.com/>
- Python, Свободное ПО, <https://www.python.org/>
- <https://www.jetbrains.com/pycharm/>

VII. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

При реализации дисциплины используются следующие помещения, оборудование и программное обеспечение:

Лаборатория машинного обучения и технологий больших данных

Персональные компьютеры (8 шт.), проектор, экран. Windows 10, Microsoft Office 365, Adobe Acrobat Reader (Бесплатное проприетарное ПО, <https://acrobat.adobe.com/ru/ru/acrobat/pdf-reader/volume-distribution.html>), Google Chrome (Свободное ПО, <https://google.com/chrome/browser/>), Mozilla Firefox, Бесплатное ПО (GNU GPL), <https://firefox.com/>, Foxit (Бесплатное проприетарное ПО, <https://www.foxitsoftware.com/ru/>), i2 Analyst's Notebook (Бесплатная лицензия для образовательных целей), <https://developer.ibm.com/academic/>), Notepad++, Бесплатное ПО (GNU GPL 2), <https://notepad-plus-plus.org/>, Total Commander 7.x, WinRAR, XAMPP, Бесплатное ПО (GNU GPL), <http://www.apachefriends.org/en/xampp.html>, Team Foundation Server 2015, Visual Studio 2015, Android Studio, Операционная система на базе Linux; Офисный пакет Open Office, актуальные версии браузеров Google Chrome (Свободное ПО, <https://google.com/chrome/browser/>), Mozilla Firefox, Бесплатное ПО (GNU GPL), <https://firefox.com/>, Edge, Safari с поддержкой протокола WebRTC, PyCharm 2017.1.2 <https://www.jetbrains.com/pycharm/> Свободное ПО, <https://www.python.org/>, Evolus Pencil, Свободное ПО (GNU GPL 2), <https://pencil.evolus.vn/>

VIII. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Дисциплина включает в себя лекционные и практические занятия, а также самостоятельную работу обучающихся.

Организация образовательного процесса по дисциплине осуществляется с использованием системы электронного обучения.

Все лекционные занятия проводятся с визуализацией учебного материала в форме презентаций лекционного материала, которые доступны в системе электронного обучения.

Лекционная часть курса включает следующие компоненты системы знаний учебной дисциплины: понятийный аппарат (тезаурус курса), теоретические утверждения, разъяснения и комментарии; междисциплинарные точки зрения; описание рассматриваемых разделов; ретроспективный и перспективный взгляды на изучаемую проблематику.

Практические занятия по всем модулям дисциплины требуют предварительной теоретической подготовки по соответствующим темам: проработка лекционного материала, ознакомление и изучение отдельных источников основной и дополнительной литературы.

Лекционные и практические занятия могут проводиться с применением дистанционных образовательных технологий с использованием платформ Microsoft Teams, Cisco, Moodle (BigBlueButton) и др.

Проведение лекционных и практических занятий осуществляется с постановкой проблемных вопросов, допускающих возникновение дискуссий, что предполагает активное включение студентов в образовательный процесс.

В организации процесса обучения используются как традиционные, характерные лекционно-семинарской форме обучения, так и инновационные (интерактивные, имитационные, проектные) технологии.

Используемые технологии обеспечивают:

- формирование компетенций, осознанное усвоение знаний, качественное освоение умений их применять и формирование заинтересованного отношения к изучаемым объектам в единстве;

- продуктивность познавательной деятельности, научный поиск, создание субъективно и объективно новых знаний или других продуктов;

- ориентацию на студентов, стимулирование их активности, самостоятельности, инициативы и ответственности;

- контекстный характер обучения, то есть привязку к реальным профессиональным задачам;

- вовлеченность студентов в выполняемую деятельность, возможность проявить и развить свой интеллектуальный, творческий, личностный, деловой потенциал.

Самостоятельная работа направлена на повышение качества обучения, углубление и закрепление знаний студента, развитие аналитических навыков по проблематике учебной дисциплины, активизацию учебно-познавательной деятельности студентов и снижение аудиторной нагрузки.

Максимальное количество баллов по каждому виду контрольных мероприятий указано в учебной карте дисциплины.

IX. УЧЕБНАЯ КАРТА ДИСЦИПЛИНЫ

Курс 1, семестр 2, очная форма обучения

№ п/п	Виды контрольных мероприятий (наименования оценочных средств)	Количество баллов	
		Текущий контроль	Рубежный контроль
Модуль 1. Математические методы машинного обучения для анализа больших данных			
1	Работа на практических занятиях (вопросы для собеседования)	30	–
Модуль 2. Искусственные нейронные сети			
2	Практикум	30	–
Всего		60	–
Бонусные баллы		Не предусмотрены	
Промежуточная аттестация в форме экзамена		<p style="text-align: center;">40 баллов</p> <p>Экзамен считается сданным при получении не менее 22 баллов, для допуска к экзамену необходимо набрать не менее 38 баллов по сумме текущего и рубежного контроля.</p> <p>Оценка по дисциплине выставляется по сумме баллов за текущий контроль, рубежный контроль и экзамен:</p> <ul style="list-style-type: none"> – 85–100 баллов – оценка «отлично»; – 71–84 балла – оценка «хорошо»; – 60–70 баллов – оценка «удовлетворительно»; – менее 60 баллов – оценка «неудовлетворительно» 	

Х. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

10.1. Паспорт фонда оценочных средств

№ п/п	Индикатор достижения компетенции	Наименование оценочного средства
1	ПК-1.1. Ставит задачи по адаптации или совершенствованию методов и алгоритмов для решения комплекса задач предметной области	– вопросы для собеседования; – практикум;
2	ПК-3.1. Руководит проектами по разработке, систем искусственного интеллекта на основе моделей глубоких нейронных сетей и нечетких моделей и методов со стороны заказчика	– вопросы для собеседования; – практикум; – экзаменационные вопросы и билеты.

10.2. Вопросы для собеседования

Модуль 1. Математические методы машинного обучения для анализа больших данных

1. Основные определения: прецедент, обучающая выборка, признаки объектов, виды признаков, матрица объектов-признаков. Модель алгоритмов, метод обучения, функционал качества алгоритма.

2. Вероятностная постановка задачи обучения. Принцип максимума правдоподобия. Связь максимизации правдоподобия и минимизации эмпирического риска.

3. Проблема переобучения и обобщающая способность алгоритма. Состоятельные методы обучения. Эмпирические оценки обобщающей способности.

4. Выбор алгоритма для вероятностной постановки задачи. Функционал среднего риска.

5. Метрические алгоритмы классификации. Обобщенный метрический классификатор. Виды и особенности частных случаев: методы ближайшего соседа, k ближайших соседей, взвешенных соседей, парзеновского окна постоянной и переменной ширины.

6. Классификация объектов по значению отступа. Алгоритм STOLP отбора эталонных объектов. Выбор метрики и проклятие размерности.

7. Приближенное вычисление плотности распределения. Наивный байесовский классификатор. Одномерный случай. Многомерный случай. Смеси распределений. EM-алгоритм разделения смеси. Смеси многомерных нормальных распределений.

8. Линейные алгоритмы классификации. Модель Мак Каллока-Питтса, алгоритм стохастического градиента для минимизации функционала среднего риска. Эвристики для улучшения сходимости и обобщающей способности.

9. Логистическая регрессия. Случайные величины с экспонентным законом распределения. Теорема о линейности байесовского классификатора (с доказательством).

10. Метод опорных векторов (SVM). Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Функция Лагранжа. Классификация объектов в зависимости от значений множителей Лагранжа. Двойственная задача. Обучение SVM.

11. Кривая ошибок и AUC. Формула вычисления AUC. Примеры.

12. Алгоритмы восстановления регрессии. Метод наименьших квадратов. Непараметрическая регрессия. Многомерная линейная регрессия. Подход с использованием SVD-разложения матрицы. Гребневая регрессия. Метод главных компонент PCA.

13. Логические методы классификации. Понятие информативности предиката: эвристическое, вероятностное, энтропийное. Поиск информативных закономерностей. Алгоритмы для номинальных и порядковых признаков. Градиентный алгоритм синтеза конъюнкций. Построение решающего списка и решающего дерева. Редукция деревьев.

14. Композиции алгоритмов. AdaBoost. AnyBoost. Градиентный бустинг. Бэггинг, метод случайных подпространств.

15. Ранжирование и рекомендательные системы. Оценки качества. Алгоритмы построения ранжирующих систем: поточечный, попарный и списочный.
16. Тематическое моделирование. Векторная модель, LSA, PLSA, LDA.
17. Кластеризация. EM-алгоритм, метод k-средних.

Критерии оценивания:

28-30 баллов, если:

- изученный материал изложен полно, определения даны верно;
- ответ показывает понимание материала;
- обучающийся может обосновать свои суждения, применить знания на практике, привести необходимые примеры, не только по учебнику и конспекту, но и самостоятельно составленные.

22-27 баллов, если:

- изученный материал изложен достаточно полно;
- при ответе допускаются ошибки, заминки, которые обучающийся в состоянии исправить самостоятельно при наводящих вопросах;
- обучающийся затрудняется с ответами на 1-2 дополнительных вопроса.

16-21 балл, если:

- материал изложен неполно, с неточностями в определении понятий или формулировке определений;
- материал излагается непоследовательно;
- обучающийся не может достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры;
- на 50% дополнительных вопросов даны неверные ответы.

0 баллов, если:

- при ответе обнаруживается полное незнание и непонимание изучаемого материала;
- материал излагается неуверенно, беспорядочно;
- даны неверные ответы более чем на 50% дополнительных вопросов.

10.3. Практикум

Модуль 2. Искусственные нейронные сети

Цель работы: реализация алгоритмов распознавания образов с помощью нейронных сетей на языке программирования и изучение их свойств.

Последовательность проведения работы:

1. Изучить устройство и работу однослойной, двухслойной нейронных сетей, сети встречного распространения.
2. Выбрать исходный алфавит классов (цифры, буквы русского алфавита, буквы английского алфавита, буквы греческого алфавита, знаки препинания и арифметические символы, радиоэлементы, геометрические фигуры, узоры, дорожные знаки, топологические элементы, картографические обозначения) по согласованию с преподавателем.
3. Составить программу, реализующую однослойную, двухслойную нейронные сети и сеть встречного распространения для задачи распознавания. Обучить нейронные сети.
4. Программное обеспечение должно позволять просматривать эталоны (классы) изображений, а также распознаваемые изображения; записывать эталоны изображений в библиотеку на диск, записывать входной образ на диск; позволять редактировать входной образ, сохранять веса нейронов при обучении сети.
5. Испытать программное обеспечение для различных входных данных.

6. Произвести оценку качества распознавания для различных случаев, систематизировав полученные результаты в таблицы. Построить графики выявленных зависимостей, сделать выводы.
7. Результаты работы оформить в виде отчета в текстовом редакторе.

Рекомендации по созданию программного обеспечения

После изучения теоретического материала, и выбора исходных изображений, системы признаков, необходимо определить количество нейронов в слоях. Максимальное количество нейронов первого слоя может совпадать с количеством признаков, хотя теоретически избыточность допускается, но не является рациональным подходом. Желательно, при создании программного обеспечения использовать объектно-ориентированный язык, чтобы была возможность динамически варьировать количество нейронов в слоях – порождать соответствующее количество экземпляров класса и количество слоев.

Рекомендации по исследованиям

При исследовании нейронных сетей необходимо проделать следующие эксперименты по выявлению качества распознавания с помощью созданного программного обеспечения.

Эксперимент №1. Исследование влияния отклонения изображения от эталона (в разных точках изображения) на качество распознавания. В качестве исходного распознаваемого образа берется один из эталонов и подправляется таким образом, чтобы отсутствовал один, два и т. д. пикселей в изображении эталона. Эксперимент проводится для нескольких случаев (отсутствие пикселей в разных участках изображения) для всех эталонов. При этом сравниваются различные нейронные сети.

Эксперимент №2. Исследование влияния отклонений в виде шума одного, двух, трех и т. д. пикселей в изображении на качество распознавания. В качестве исходного распознаваемого образа берется один из эталонов и добавляется один или несколько пикселей шума. Эксперимент проводится для различного расположения шума и для различных эталонов. В ходе эксперимента также сравниваются различные нейронные сети.

Эксперимент №3. Исследование влияния наличия шума и отклонений в изображении в виде одного, двух, трех и т. д. пикселей в изображении на качество распознавания. В качестве исходного распознаваемого образа берется один из эталонов. В изображение распознаваемого образа вносится шум в виде нескольких пикселей и удаляется несколько пикселей в изображении символа. Эксперимент повторяется для различного расположения шума и отклонений, и для разных эталонов на различных типах нейронных сетей. Данный эксперимент является комбинацией первых двух.

Эксперимент №4. Исследование влияния наличия черной строки или столбца в изображении (как помеха в образе) на качество распознавания. В качестве исходного распознаваемого образа берется один из эталонов. В изображение вносится черная строка или столбец. Эксперимент повторяется для различного положения строки или столбца в изображении и для различных эталонов. В ходе эксперимента сравниваются различные нейронные сети.

Эксперимент №5. Исследование влияния наличия белой строки или столбца в изображении (как помеха в образе) на качество распознавания. В качестве исходного распознаваемого образа берется один из эталонов. В изображение вносится белая строка или столбец. Эксперимент повторяется для различного положения строки или столбца в изображении и для различных эталонов на различных нейронных сетях.

Эксперимент №6. Исследование влияния количества нейронов в слоях на качество распознавания. В качестве исходного распознаваемого образа берется один из эталонов. Количество нейронов в слое варьируется от двух до равного числу признаков (для двухслойных сетей слои варьируются последовательно). Эксперимент повторяется на различных нейронных сетях.

Эксперимент №7. Исследование влияния количества нейронов в слоях и количества эталонов на скорость обучения сети. В ходе проведения эксперимента количество нейронов в слое варьируется от двух до равного числу признаков (для двухслойных сетей слои варьируются

последовательно), также варьируется количество эталонов (в пределах выбранных по согласованию с преподавателем). В каждом случае фиксируется число итераций и время обучения. Эксперимент повторяется на различных нейронных сетях.

Эксперимент №8. Исследование влияния начертания входных образов на качество распознавания. В качестве исходного распознаваемого образа берется один из эталонов. Эталон модифицируется (делается жирное или наклонное, или подчеркнутое начертание). Эксперимент повторяется на различных нейронных сетях для различных видоизменений.

Примечание. При использовании цвета в изображении эксперименты №4 и №5 следует проводить для цвета фона и цвета образа. В эксперименте №4 вместо черной строки или столбца берется столбец или строка цвета образа. В эксперименте №5 вместо белой строки или столбца берется строка или столбец фоновых цвета.

Рекомендации по оформлению отчета

Результаты экспериментов следует оформить в виде таблиц и графиков с пояснениями и подписями. На графиках следует показать зависимость ошибки распознавания для различных нейронных сетей от количества и вида помех, вносимых в изображение, а также от варьируемых параметров.

В качестве таблицы значений можно использовать таблицу следующего вида

Изображение входного образа	Значение СКО	Нейронная сеть		
		Однослойная	Двухслойная	Кохонена-Гроссберга

Заполняйте две копии этой таблицы – одну для эталонов, другую для предъявляемых изображений, записывая в колонки значения выхода нейронной сети.

В отчете также следует привести данные об обучении нейронной сети: количество итераций, первоначальные и конечные значения весов.

Контрольные вопросы для собеседования

1. Каковы основные понятия теории распознавания?
2. Дайте определение класса образов.
3. Что такое алфавит классов?
4. Дайте определение объекта класса образов.
5. Дайте определение признака класса образов.
6. Какие типы признаков вы знаете? Приведите примеры.
7. Что такое нейронная сеть?
8. Что такое синапс?
9. Что такое аксон?
10. Что определяет уровень активации нейрона?
11. Дайте определение активационной функции.
12. Какие типы активационных функций Вам известны?
13. Что такое персептрон?
14. В чем преимущество сигмоидальной функции?
15. В чем заключается проблема функции «исключающее или»?
16. В чем заключается цель обучения нейронной сети?
17. Что такое обучающая пара?
18. Что такое обучающее множество?
19. Расскажите алгоритм обучения персептрона.
20. Что такое дельта-правило?
21. Перечислите шаги процедуры обратного распространения.
22. Какие действия выполняются при проходе вперед?

23. Какие действия выполняются при обратном проходе?
24. Какие недостатки есть у процедуры обратного распространения?
25. Опишите устройство сети встречного распространения.
26. Как устроен и работает слой Кохонена?
27. Как устроен и работает слой Гроссберга?
28. В чем заключается проблема выбора начальных значений весовых векторов?
29. Как решают проблему выбора начальных значений весовых векторов?

Критерии оценивания:

26-30 баллов выставляется студенту, если он своевременно выполнил все задачи, предусмотренные практикумом, подготовил отчет в соответствии с требованиями преподавателя и в процессе защиты продемонстрировал наличие теоретических знаний в объеме содержания учебной дисциплины, относящейся к практикуму. Сумел ответить на дополнительные вопросы, связанные не только с процессом выполнения работы, но и с пониманием совершенных действий и решенных задач.

21-25 баллов выставляется студенту, если он выполнил от 70% задач, предусмотренных практикумом, подготовил отчет в соответствии с требованиями преподавателя и в процессе защиты продемонстрировал наличие теоретических знаний в объеме содержания учебной дисциплины, относящейся к практикуму. Сумел ответить на вопросы, связанные с процессом выполнения работы.

18-20 баллов выставляется студенту, если он более чем на 50% выполнил поставленные в практикуме задачи, способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к практикуму.

0 баллов выставляется студенту, если он более чем на 50% не выполнил поставленные в практикуме задачи, не способен ответить на вопросы, касающиеся теоретической составляющей в объеме содержания учебной дисциплины, относящейся к практикуму.

10.4. Экзаменационные вопросы и билеты

Вопросы к экзамену

Модуль 1. Математические методы машинного обучения для анализа больших данных

1. Основные определения: прецедент, обучающая выборка, признаки объектов, виды признаков, матрица объектов-признаков. Модель алгоритмов, метод обучения, функционал качества алгоритма.

2. Вероятностная постановка задачи обучения. Принцип максимума правдоподобия. Связь максимизации правдоподобия и минимизации эмпирического риска.

3. Проблема переобучения и обобщающая способность алгоритма. Состоятельные методы обучения. Эмпирические оценки обобщающей способности.

4. Выбор алгоритма для вероятностной постановки задачи. Функционал среднего риска.

5. Метрические алгоритмы классификации. Обобщенный метрический классификатор. Виды и особенности частных случаев: методы ближайшего соседа, k ближайших соседей, взвешенных соседей, парзеновского окна постоянной и переменной ширины.

6. Классификация объектов по значению отступа. Алгоритм STOLP отбора эталонных объектов. Выбор метрики и проклятие размерности.

7. Приближенное вычисление плотности распределения. Наивный байесовский классификатор. Одномерный случай. Многомерный случай. Смеси распределений. EM-алгоритм разделения смеси. Смеси многомерных нормальных распределений.

8. Линейные алгоритмы классификации. Модель Мак Каллока-Питтса, алгоритм стохастического градиента для минимизации функционала среднего риска. Эвристики для улучшения сходимости и обобщающей способности.

9. Логистическая регрессия. Случайные величины с экспонентным законом распределения. Теорема о линейности байесовского классификатора (с доказательством).
10. Метод опорных векторов (SVM). Случай линейно разделимой выборки. Случай линейно неразделимой выборки. Функция Лагранжа. Классификация объектов в зависимости от значений множителей Лагранжа. Двойственная задача. Обучение SVM.
11. Кривая ошибок и AUC. Формула вычисления AUC. Примеры.
12. Алгоритмы восстановления регрессии. Метод наименьших квадратов. Непараметрическая регрессия. Многомерная линейная регрессия. Подход с использованием SVD-разложения матрицы. Гребневая регрессия. Метод главных компонент PCA.
13. Логические методы классификации. Понятие информативности предиката: эвристическое, вероятностное, энтропийное. Поиск информативных закономерностей. Алгоритмы для номинальных и порядковых признаков. Градиентный алгоритм синтеза конъюнкций. Построение решающего списка и решающего дерева. Редукция деревьев.
14. Композиции алгоритмов. AdaBoost. AnyBoost. Градиентный бустинг. Бэггинг, метод случайных подпространств.
15. Ранжирование и рекомендательные системы. Оценки качества. Алгоритмы построения ранжирующих систем: поточечный, попарный и списочный.
16. Тематическое моделирование. Векторная модель, LSA, PLSA, LDA.
17. Кластеризация. EM-алгоритм, метод k-средних.

Модуль 2. Искусственные нейронные сети

1. Основные архитектуры и виды нейронных сетей: слоистые, полносвязные, сигмоидные, монотонные; нейросети с учителем и без учителя, Хопфилда, Кохонена
2. Элементы нейросетей: синапс или линейная связь, нелинейный элемент или функция активации, точка ветвления, сумматоры - простой, адаптивный, неоднородный, квадратичный
3. Биологический нейрон
4. Режимы работы нейросетей (операции с нейросетями)
5. Типы нелинейных функций
6. Входные и выходные сигналы, функционирование, обучение, тестирование, оценивание
7. Обучение и оптимизация. Методы обучения: градиентный, случайный, партан и др. квазиньютоновский и сопряженных градиентов; одномерная оптимизация
8. Обучаемые нейросети. Обучение по примерам, страницам, по всему задачку (обучающей выборке); преимущества, проблемы и особенности обучения по страницам
9. Значимость параметров и сигналов
10. Контрастирование
11. Предобработка, ее виды: Перемасштабирование, Нормализация, Стандартизация
12. Задачи для нейросетей: задачи математические и прикладные
13. Оценка работы сети
14. Архитектуры нейроимитаторов: элементы нейрокомпьютера или нейроимитатора
15. Постановка задачи для обучения НС; методика сбора и организации данных
16. Аппроксимация и основные теоремы: Вейерштрасса, Стоуна, Обобщенная

ПРИМЕР ЭКЗАМЕНАЦИОННОГО БИЛЕТА

Направление/специальность 09.04.03 Прикладная информатика

1. Проблема переобучения и обобщающая способность алгоритма. Состоятельные методы обучения. Эмпирические оценки обобщающей способности.
2. Архитектуры нейроимитаторов: элементы нейрокомпьютера или нейроимитатора

Критерии оценки:

40-34 балла выставляется студенту, если:

- ответ на вопросы экзаменационного билета дан в полном объеме;
- ответ характеризуется связностью, логичностью, достаточным объёмом, с приведением примеров и объяснением алгоритмов выполнения операций;
- допустимы единичные ошибки, которые исправлялись в процессе ответа самим экзаменуемым.

33-28 баллов выставляется студенту, если:

- экзаменационное задание в основном выполнено;
- ответ характеризуется достаточной связностью и логичностью, но небольшим объёмом изложенного материала;
- допустимы единичные ошибки, которые исправлялись экзаменаторами в процессе ответа и отсутствие практических примеров.

27-22 балла выставляется студенту, если:

- экзаменационное задание частично выполнено.
- ответ характеризуется недостаточной связностью, логичностью, небольшим объёмом изложенного материала;
- при ответе студент допускает ошибки, которые не в состоянии самостоятельно исправить.

0 баллов выставляется студенту, если:

- экзаменационное задание выполнено фрагментарно;
- в ответе отсутствует связность и логичность;
- при ответе студент допускает грубые ошибки, которые не в состоянии исправить самостоятельно.