

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»



УТВЕРЖДАЮ

Первый проректор

А.В. Толстиков

2022 г.

МОДЕЛИ, МЕТОДЫ И АЛГОРИТМЫ АНАЛИЗА БОЛЬШИХ ДАННЫХ
Рабочая программа
для обучающихся по научной специальности 1.2.2. Математическое моделирование,
численные методы и комплексы программ
форма обучения (очная)

Захарова И.Г. Модели, методы и алгоритмы анализа больших данных. Рабочая программа для обучающихся по научной специальности 1.2.2. Математическое моделирование, численные методы и комплексы программ, форма обучения (очная). Тюмень, 2022.

Рабочая программа составлена в соответствии с требованиями ФГТ, утвержденными приказом Минобрнауки России от 20.10.2021 г. № 951.

Рабочая программа дисциплины Модели, методы и алгоритмы анализа больших данных опубликована на сайте ТюмГУ: [электронный ресурс] / Режим доступа: <https://www.utmn.ru/sveden/education/#>.

1. Цели и задачи освоения дисциплины (модуля)

Цель дисциплины: сформировать у аспирантов знания и навыки практического применения основных подходов и методов глубокого обучения для исследования и решения задач анализа больших данных с применением специализированных алгоритмов, библиотек, экосистем и облачных технологий.

Задачи дисциплины:

- сформировать умение разрабатывать и исследовать нейронные сети различной архитектуры, алгоритмы их обучения и использовать их для решения задач анализа данных;
- развить навыки применения открытых библиотек, экосистем, методов, алгоритмов и технологий для исследования и решения задач глубокого обучения;
- обеспечить готовность использования моделей, методов и алгоритмов анализа больших данных в процессе постановки, исследования и решения задач анализа данных для различных предметных областей.

2. Компетенции обучающегося, формируемые в результате освоения данной дисциплины

ПК-2 - способность к системному анализу проблем, обоснованию выбора, разработке и тестированию адекватных задач, эффективных вычислительных методов и алгоритмов средствами перспективных компьютерных технологий.

3. Структура и объем дисциплины

Таблица 1

Вид учебной работы	Всего часов (академические часы)	Часов в семестре (академические часы)
		4 семестр
Общий объем зач. ед. час	3	3
	72	72
Часы аудиторной работы (всего):	22	22
Лекции	12	12
Практические занятия	10	10
Лабораторные / практические занятия по подгруппам	0	0
Часы внеаудиторной работы, включая самостоятельную работу обучающегося	50	50
Вид промежуточной аттестации	36	Дифференцированный зачет 36

4. Система оценивания

4.1. Система оценивания - пятибалльная. Наличие проходного балла – оценка «Удовлетворительно».

Аспиранты сдают дифференцированный зачет в устной форме. Билеты для экзамена выбираются случайным образом из набора. Зачет проходит в устной форме, аспиранту дается 45 минут на подготовку к вопросу.

Оценка «отлично» выставляется обучающемуся, который дает полный, развернутый ответ на поставленный вопрос, раскрывает основные положения темы; показывает умение выделить существенные и несущественные признаки, причинно-следственные связи; в

ответе прослеживается четкая структура, логическая последовательность, отражающая сущность раскрываемых понятий, теорий, явлений; ответ излагается литературным языком в научных терминах.

Оценка «хорошо» выставляется обучающемуся, который либо дает частичный ответ, либо допускает ошибки, при ответе на поставленный вопрос, допускает неточности при раскрытии основных положений темы, допускает ошибки но показывает умение выделить существенные и несущественные признаки, причинно-следственные связи; в ответе прослеживается четкая структура, логическая последовательность, отражающая сущность раскрываемых понятий, теорий, явлений; ответ излагается литературным языком в научных терминах.

Оценка «удовлетворительно» выставляется обучающемуся, который дает частичный ответ на поставленный вопрос, либо допускает фактические ошибки, не раскрывает основные положения темы; в ответе не в полной мере прослеживается четкая структура, нарушена логическая последовательность, отражающая сущность раскрываемых понятий, теорий, явлений, но ответ излагается логично, литературным языком в научных терминах.

Оценка «не удовлетворительно» выставляется обучающемуся, который дает неполный ответ, представляющий собой разрозненные знания по теме вопроса с существенными ошибками в определениях; в ответе присутствует фрагментарность, нелогичность изложения; обучающийся не осознает связь данного понятия, теории, явления с другими объектами дисциплины; отсутствуют выводы, конкретизация и доказательность изложения; речь неграмотная; дополнительные и уточняющие вопросы преподавателя не приводят к коррекции ответа обучаемого не только на поставленный вопрос, но и на другие вопросы дисциплины, либо обучающийся отказывается от ответа.

5. Содержание дисциплины

5.1. Тематический план дисциплины

Таблица 2

№ п/п	Наименование тем и/или разделов	Объем дисциплины (модуля), час.				
		Всего	Виды аудиторной работы (академические часы)			Иные виды контактной работы
			Лекции	Практические занятия	Лабораторные/практические занятия по подгруппам	
1	2	3	4	5	6	7
1.	Место Deep Learning в научных исследованиях.	2	2	0	0	0
2.	Математические основы DeepLearning.	4	2	2	0	0
3.	Основные архитектуры нейронных сетей.	4	2	2	0	0
4.	Библиотека TensorFlow.	4	2	2	0	0
5.	Сверточные нейронные сети.	4	2	2	0	0
6.	Рекуррентные нейронные сети.	4	2	2	0	0

7.	Дифференцированный зачет	36	0	0	0	36
	Итого (часов)	58	12	10	0	36

5.2. Содержание дисциплины (модуля) по темам

Лекционные занятия

Тема 1. Место Deep Learning в научных исследованиях.

Основные цели и задачи изучения дисциплины. Структура курса. Организация лекционных и практических занятий. Самостоятельная работа. Формы контроля. Среды разработки. Предметная область Deep Learning. Модели глубокого обучения. Основные понятия нейросетей.

Тема 2. Математические основы Deep Learning

Перцептрон и функции активации. Граф вычислений и дифференцирование. Реализация простейшего перцептрона. Алгоритм вычисления весов для задачи классификации. Функции ошибки и регуляризация. Виды функции ошибок для основных задач. Метод градиентного спуска. Стохастический градиентный спуск. Алгоритм adam. Регуляризация. Dropout.

Тема 3. Основные архитектуры нейронных сетей

Основные архитектуры нейронных сетей. Сети ForwardFeed. Инициализация весов. Мини-батчи – нормализация. Облачные вычисления в Colab от Google. Режимы работы. Сеансы.

Тема 4. Библиотека TensorFlow

Введение в Keras и TensorFlow. Общая характеристика. Операции TensorFlow. Виды тензоров. Операции над тензорами. Решение основных задач машинного обучения с помощью нейронных сетей на основе библиотеки TensorFlow. Интеграция TensorFlow с библиотеками SciPy, sklearn. Решение задач бинарной и мультиклассовой классификации.

Тема 5. Сверточные нейронные сети

Понятие свертки. Сверточные сети. Современные сверточные архитектуры. Области применимости сверточных сетей. Создание нейронной сети заданной архитектуры. Реализация классификатора для распознавания рукописных цифр.

Тема 6. Рекуррентные нейронные сети

Рекуррентные нейронные сети. Обработка последовательностей. Распространение ошибки. Виды архитектур рекуррентных нейронных сетей. Нейронные сети для порождения текста. Задача порождения текста. Рекуррентные нейронные сети и синтаксический разбор текста.

Практические занятия

Тема 2. Математические основы Deep Learning

Разработка программного кода для построения моделей на основе простых архитектур нейросетей и исследования алгоритмов их обучения:

- Алгоритм вычисления весов для задачи классификации.
- Функции ошибки и регуляризация.
- Виды функции ошибок для основных задач.
- Метод градиентного спуска.
- Dropout.

Тема 3. Основные архитектуры нейронных сетей

Разработка и тестирование программного кода для реализации основных архитектур нейронных сетей:

- Сети ForwardFeed.
- Инициализация весов.
- Мини-батчи – нормализация.

- Облачные вычисления в Colab от Google.

Тема 4. Библиотека TensorFlow

Разработка программного кода с использованием библиотеки TensorFlow:

- Операции TensorFlow.
- Операции над тензорами.
- Решение основных задач машинного обучения с помощью нейронных сетей на основе библиотеки TensorFlow.
- Решение задач бинарной и мультиклассовой классификации.

Тема 5. Сверточные нейронные сети

Разработка программного кода с использованием библиотеки TensorFlow:

- Создание нейронной сети заданной архитектуры.
- Реализация классификатора для распознавания рукописных цифр.

Тема 6. Рекуррентные нейронные сети

Разработка программного кода с использованием библиотеки TensorFlow:

- Нейронные сети для порождения текста.
- Рекуррентные нейронные сети и синтаксический разбор текста.

6. Учебно-методическое обеспечение и планирование самостоятельной работы обучающихся

Таблица 3

№ темы	Темы	Формы СРС, включая требования к подготовке к занятиям
1	Место Deep Learning в научных исследованиях.	Чтение обязательной и дополнительной литературы, проработка лекций
2	Математические основы DeepLearning.	Чтение обязательной и дополнительной литературы, проработка лекций, разработка программного кода и проведение вычислительного эксперимента
3	Основные архитектуры нейронных сетей.	Чтение обязательной и дополнительной литературы, проработка лекций, разработка программного кода и проведение вычислительного эксперимента
4	Библиотека TensorFlow.	Чтение обязательной и дополнительной литературы, проработка лекций, разработка программного кода и проведение вычислительного эксперимента
5	Сверточные нейронные сети.	Чтение обязательной и дополнительной литературы, проработка лекций, разработка программного кода и проведение вычислительного эксперимента
6	Рекуррентные нейронные сети.	Чтение обязательной и дополнительной литературы, проработка лекций, разработка программного кода и проведение вычислительного эксперимента

Чтение обязательной и дополнительной литературы, предусмотренной рабочей программой дисциплины. Контроль – на практическом занятии в устной или письменной форме при обсуждении теоретических вопросов.

Проработка лекций предполагает присутствие обучаемого на лекционных занятиях и конспектирование материала, подготовка презентаций усвоенного лекционного материала. Контроль – на практическом занятии в устной или письменной форме при обсуждении теоретических вопросов.

Разработка программного кода и проведение вычислительного эксперимента выполняются по материалам лекций и самостоятельных исследований. Контроль – на практическом занятии в устной или письменной форме при обсуждении теоретических вопросов.

7. Промежуточная аттестация по дисциплине (модулю)

7.1 Оценочные материалы для проведения промежуточной аттестации по дисциплине (модулю)

Форма проведения промежуточной аттестации – дифференцированный зачет. Зачет проводится устно в форме собеседования по прилагаемому ниже примерному перечню вопросов.

Вопросы для подготовки к зачету:

1. Основные понятия нейросетей.
2. Перцептрон и функции активации.
3. Граф вычислений и дифференцирование.
4. Алгоритм вычисления весов для задачи классификации.
5. Функции ошибки и регуляризация. Виды функции ошибок для основных задач.
6. Метод градиентного спуска. Стохастический градиентный спуск. Алгоритм adam.
7. Регуляризация. Dropout.
8. Основные архитектуры нейронных сетей. Сети ForwardFeed.
9. TensorFlow. Общая характеристика. Операции TensorFlow.
10. Виды тензоров. Операции над тензорами.
11. Решение основных задач машинного обучения с помощью нейронных сетей на основе библиотеки TensorFlow.
12. Понятие свертки. Сверточные сети.
13. Современные сверточные архитектуры. Области применимости сверточных сетей.
14. Создание нейронной сети заданной архитектуры.
15. Реализация классификатора для распознавания рукописных цифр.
16. Рекуррентные нейронные сети.
17. Обработка последовательностей. Распространение ошибки.
18. Виды архитектур рекуррентных нейронных сетей.
19. Нейронные сети для порождения текста.
20. Рекуррентные нейронные сети и синтаксический разбор текста

8. Учебно-методическое и информационное обеспечение дисциплины (модуля)

8.1 Основная литература:

1. Одинцов, Б.Е. Модели и проблемы интеллектуальных систем [Электронный ресурс]: монография / – Электрон. дан. – М.: НИЦ ИНФРА-М, 2020. – Режим доступа: <https://znanium.com/catalog/document?id=356918> (дата обращения: 20.03.2022)
2. Сулова, Н. Ю. Искусственный интеллект [Электронный ресурс]: монография / – Электрон. дан. – М.: ЮНИТИ-ДАНА, 2021. 408 с. – Режим доступа: <https://www.iprbookshop.ru/123354.html> (дата обращения: 20.03.2022)

8.2 Дополнительная литература:

3. Ивченко, Г. И. Математическая статистика: учебник для студентов вузов, обучающихся по специальности "Прикладная математика"/ Г. И. Ивченко, Ю. И. Медведев. - Москва: Либроком, 2014. - 352 с.;
4. Петрунин, Ю. Ю. Информационные технологии анализа данных. Data Analysis: учебное пособие по дисциплине "Информатика" для студентов вузов, обучающихся по управленческим и экономическим специальностям и направлениям/ Ю. Ю. Петрунин; Моск. гос. ун-т им. М. В. Ломоносова. - 2-е изд. - Москва: КДУ, 2010. - 292 с.
5. Просто о больших данных = Big Data for Dummies: перевод с английского/ Д. Гурвиц. - Москва: Эксмо, 2015. - 400 с.

8.3 Интернет-ресурсы:

1. Сергей Николенко. Deep Learning [Электр. ресурс]. – Режим доступа <https://logic.pdmi.ras.ru/~sergey/teaching/dlharbour2017.html> (дата обращения: 20.03.2022).
2. First Steps with TensorFlow [Электр. ресурс]. – Режим доступа <https://developers.google.com/machine-learning/crash-course/first-steps-with-tensorflow/toolkit> (дата обращения: 20.03.2022).
3. Overview of Colaboratory Features [Электр. ресурс]. – Режим доступа https://colab.research.google.com/notebooks/basic_features_overview.ipynbml (дата обращения: 20.03.2022).
4. TensorFlow Core [Электр. ресурс] Режим доступа - <https://www.tensorflow.org/tutorials> (дата обращения: 20.03.2022).

9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю)

Интернет, доступ в информационно-образовательную среду ТюмГУ, включающую в себя доступ к учебным планам и рабочим программам, к изданиям электронной библиотечной системы и электронным образовательным ресурсам; лицензионное программное обеспечение: MS Windows, MS Office, PowerPoint, MS Teams.

10. Технические средства и материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения занятий лекционного и практического типа, консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы. Для проведения занятий лекционного типа необходимо демонстрационное оборудование.

11. Средства адаптации преподавания дисциплины (модуля) к потребностям лиц с ограниченными возможностями

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) могут предлагаться одни из следующих вариантов восприятия информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); индивидуальные задания и консультации.

- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные задания и консультации.

- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.

12. Методические рекомендации обучающимся по выполнению самостоятельной работы

При подготовке доклада рекомендуется структурированно и сжато представить содержание проделанной работы в соответствии с планом. Выступление обязательно сопровождать наглядностью (презентация в формате pdf). Также необходимо быть готовым к ответам на вопросы, поступающие от аудитории.

При подготовке к ответам на вопросы, дискуссии рекомендуется, кроме актуализации содержания представляемого материала, чтение теоретических источников, анализ содержания материала лекционных и практических занятий по дисциплинам образовательного компонента, содержания дискуссий на посещенных аспирантом заседаниях кафедры, диссертационного совета, конференциях, научных семинарах и иных мероприятиях.

Рекомендуется использовать информационные ресурсы современных электронных библиотек, обращаться к публикациям по проблемам искусственного интеллекта в рейтинговых изданиях, сборниках материалов конференций, записям вебинаров.

Методические рекомендации по подготовке к практическим занятиям

Выполнение практических заданий предполагает наличие у аспиранта уверенных навыков программирования на одном из языков высокого уровня (C/C++, C#, Java, Scala, Python). При этом преимущество отдается владению языком Python. Для выполнения задания аспирантам выдается инструкция по работе с системой программирования. Для разработки и тестирования программного кода на языке Python аспиранты могут использовать облачную систему Google Colaboratory (доступ свободный).

Для успешного выполнения практических заданий необходимо:

- изучить материалы лекций, а также документацию к используемым библиотекам;
- разбирать и комментировать примеры программного кода, предложенного преподавателем для иллюстрации теоретического материала лекций;
- выполнить поиск дополнительной информации в предложенной преподавателем литературе, а также самостоятельно подобранным источникам;
- при необходимости самостоятельно под руководством преподавателя повысить уровень подготовки в области разработки программного кода.

Методические рекомендации по подготовке к дифференцированному зачёту

Подготовка к дифференцированному зачету по дисциплине является традиционной формой проверки знаний, умений, компетенций, сформированных в процессе освоения всего содержания изучаемой дисциплины. В случае проведения дифференцированного зачета аспирант получает баллы, отражающие уровень его знаний. Самостоятельная подготовка к зачету должна осуществляться в течение всего семестра. Подготовка включает следующие действия:

- изучить материалы лекций, а также материалы, которые готовились к практическим занятиям в течение семестра;
- соотнести прочитанную информацию с вопросами, которые даны к зачету;
- выполнить поиск дополнительной информации в предложенной преподавателем литературе, а также самостоятельно подобранным источникам;
- подготовить краткие записи, фиксирующие логическую схему ответа на вопросы.
- накануне зачета повторить ответы, не заглядывая в записи.