

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Механико-математический факультет

УТВЕРЖДАЮ:

Декан

 Л. В. Гензе

« 30 » _____ 06 _____ 20 22 г.

Рабочая программа дисциплины

Искусственные нейронные сети на Python

по направлению подготовки

02.03.01 Математика и компьютерные науки

Направленность (профиль) подготовки :

Основы научно-исследовательской деятельности в области математики и компьютерных наук

Форма обучения

Очная

Квалификация

Бакалавр

Год приема

2022, 2023

Код дисциплины в учебном плане: Б1.В.3.05

СОГЛАСОВАНО:

Руководитель ОП

 Л. В. Гензе

Председатель УМК

 Е. А. Тарасов

Томск – 2022

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-2 Способен находить или создавать, анализировать, реализовывать программно и использовать на практике современный математический аппарат, математические модели и алгоритмы, в том числе с применением современных вычислительных систем в научно-исследовательской и (или) опытно-конструкторской деятельности в различных областях техники, естествознания, экономики и управления.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК 2.1 Использует методы построения и анализа математических моделей в задачах естествознания, техники, экономике и управлении.

ИОПК 2.2 Демонстрирует умение применять на практике математические модели и компьютерные технологии (в том числе с применением многопроцессорных систем) для решения различных задач в области профессиональной деятельности.

ИОПК 2.3 Участвует в разработке математических моделей для решения задач естествознания, техники, экономики и управления под руководством более квалифицированного работника.

2. Задачи освоения дисциплины

- Освоить современные понятия и методы теории искусственных нейронных сетей.
- Сформировать навыки программирования на языке Python.
- Получить практические знания для самостоятельного создания и использования сложных структур данных.
- Научиться разрабатывать архитектуру искусственной нейронной сети под определённые классы задач

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к Блоку 1 «Дисциплина (модули)».

Дисциплина относится к обязательной части образовательной программы.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Седьмой семестр, зачет с оценкой

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: программирование, математическая логика, дискретная математика, теория вероятностей и математическая статистика.

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 32 ч.

-практические занятия: 32 ч.

в том числе практическая подготовка: 32 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Язык программирования Python

Введение в среду программирования Python. Типы данных и составные объекты Python. Управляющие конструкции в Python. Ввод/вывод. Пользовательские функции и основы функционального программирования на Python. Модульное программирование. Стандартные и нестандартные модули Python. Численные методы. Работа с текстом и строками. 4 Домашних и 3 индивидуальных задания.

Тема 2. Искусственные нейронные сети

Введение. Исторический аспект. Основные компоненты ИНС. Обучение нейронной сети. Многослойные нейронные сети прямого распространения. Обучение нейронной сети. Алгоритм обратного распространения ошибки. Основы теории нейронных сетей. Рекуррентные нейронные сети. Сверточные нейронные сети. 1 итоговое индивидуальное задание.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля выполнения домашних и индивидуальных заданий и фиксируется в форме контрольной точки не менее одного раза в семестр.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Результаты зачета с оценкой определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

При оценке выполнения индивидуальных и домашних заданий учитывается правильность, оригинальность и сроки выполнения.

Ближе к концу семестра каждый студент получает итоговое индивидуальное задание. Оно включает в себя некий результирующий итог по освоению материала курса. Работа оформляется в виде отчёта, который студенту необходимо защитить: рассказать о ходе выполнения работы и ответить на дополнительные вопросы по теории нейронных сетей.

По результатам защиты итогового индивидуального задания и по результатам сданной практики определяется оценка.

Примерный перечень теоретических вопросов

1. Структура нейрона
2. Нейронные сети прямого распространения
3. Прямое распространение ошибки
4. Обратное распространение ошибки
5. Минимизация ошибки
6. Почему при оптимизации используются частные производные
7. Как борются с переобучением
8. Какой метод лежит в основе оптимизации нейронных сетей
9. Режимы обучения
10. Функции активации

Пример задачи

В ходе разведки месторождений нефти специалисты производят пробные бурения скважин и осуществляют анализ получаемых в ходе этого технических, геологических и геофизических данных. Целью этого является обнаружение нефтенасыщенных пластов, то есть пластов, содержащих в себе нефть и способных ее отдавать.

Перед вами стоит задача разработать алгоритм интеллектуального анализа реальных данных, позволяющий наиболее качественно определять наличие или отсутствие нефтяных пластов на тех или иных глубинах залегания скважин.

Метрикой качества выступает точность нахождения нефтенасыщенного пласта

$$Accuracy = \frac{samples_true}{samples_all},$$

Где *samples_true* - количество правильных предсказаний наличия/отсутствия нефтяного пласта,

sampels_all - общее количество записей в таблице

Формат ввода

data_train.csv — файл с обучающими табличными данными

X_data_predict.csv — файл с данными, для которых необходимо предсказать целевую переменную

Файл с тренировочными табличными данными содержит информацию по 600 скважинам, для каждой из которых имеется различная техническая, геологическая и геофизическая информация в виде следующих полей:

- **MD** — относительная глубина скважины (относительно поверхности бурения), всегда является положительной величиной, используется для привязки глубин внутри скважины, но не может выступать в роли какого-то признака при прогнозе (по крайней мере с физической точки зрения).
- **TVDSS** — глубина скважины относительно уровня моря, всегда является положительной величиной, может отражать поверхность геологического пласта или уровень водонефтяного контакта.
- **Layer** — название пласта, геологическая принадлежность интервала, качественная характеристика, выдаваемая геологом на основе его понимания геометрических характеристик целевого пласта, служащая для сопоставления пластов из различных скважин между собой.
- **GK** — гамма-каротаж, измеряет естественную радиоактивность пород, различные минералы имеют разное содержание радиоактивных материалов, как правило, чем выше — тем больше глинистая составляющая и меньше песчаная, может измеряться в единицах API или мкр/ч.
- **NNKT_big** — нейтронный каротаж, регистрирует относительное водородосодержание, что может говорить о количестве пор в горных породах (они не могут быть пустыми и всегда содержат какой-то флюид, который в значительном объеме содержит в себе водород). Меньшие значения отвечают за более высокое флюидосодержание.
- **PS** — каротаж естественной поляризации, последняя возникает при фильтрации флюида через породу, уменьшение значений говорит о наличии проницаемого интервала. Единица измерения — милливольты, может иметь совершенно разный масштаб в разных скважинах.
- **IK** — индукционный каротаж, отражает электрическую проводимость горных пород, величину, обратную сопротивлению. Поскольку нефть является диэлектриком, а вода проводником, высокие показания отражают водонасыщенные пласты, а низкие — интервалы, вмещающие нефть. С другой стороны, плотные породы, не содержащие в себе пор, также имеют высокое сопротивление, поскольку не имеют в себе флюида, который способен проводить ток.
- **BK** — боковой зонд, отражает сопротивление горной породы, интерпретируется схожим образом с кривой индукционного каротажа, но уже наоборот, повышенные значения — нефть или плотные породы, пониженные — вода или глина.
- **PZ** — потенциал-зонд, отражает сопротивление горной породы, интерпретируется схожим образом с кривой индукционного каротажа, но уже наоборот, повышенные

значения — нефть или плотные породы, пониженные — вода или глина. Схож с боковым зондом (БК), но имеет другую глубинность исследования.

- **Grad_zond** — другая группа зондов, отвечающих за сопротивление горных пород, в зависимости от числа в названии определяется глубинность метода. При бурении буровой раствор попадает в пласт и может изменить содержание того или иного флюида, поэтому, в теории, пониженные сопротивления в затронутой части пласта и повышенные в глубинной могут быть признаком наличия углеводородов.
- **target_collector** — бинарная характеристика, выдаваемая специалистом по интерпретации каротажных данных, отвечающая за то, является ли тот или иной интервал коллекторским пластом, то есть пластом, способным принимать и отдавать флюид.
- **target_oil** — бинарная характеристика, выдаваемая специалистом по интерпретации каротажных данных, отвечающая за то, является ли тот или иной интервал коллекторским нефтенасыщенным пластом.
- **Well** — номер скважины.

В качестве целевой переменной выступает **target_oil**, которая при значении 1 говорит о наличии нефтенасыщенного пласта, а при значении 0 — о его отсутствии.

Формат вывода

В файл submission.csv необходимо записать одну колонку, в которой для каждой скважины из тестовой выборки стоит классифицирующая ее метка.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=33786>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

в) Самостоятельная работа студентов включает в себя: теоретическое освоение лекционного курса, практическое выполнение заданий и индивидуальных заданий, подготовку к зачету с оценкой. Для выполнения самостоятельной работы обеспечивается доступ к информационным ресурсам курса:

- материалы лекций;
- список вопросов для самостоятельной проверки знаний и подготовки к зачёту.
- список литературы, включающий учебники и книги по изучаемым в курсе вопросам.

Все лабораторные работы и индивидуальные задания подобраны так, чтобы максимально стимулировать психологическую установку студентов-математиков на формирование связи между математической теорией и ее практическим применением. Отчет по каждой лабораторной работе включает теоретическую часть, выполненное практическое задание и анализ полученных результатов.

г) Для успешного освоения материала студентам необходимо посещать занятия, а во время самостоятельной работы пользоваться основной и дополнительной литературой, базами данных и информационно-справочными системами, которые представлены в списке литературы. Самостоятельная работа студентов состоит в повторении материала с практических занятий и самостоятельного изучения дополнительных вопросов, более глубокого анализа темы с помощью литературы.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

1. Сузи Р.А. Язык программирования Python. 2016 – 350с.
2. Васильев А.Н. Python на примерах. С.-Пб.: Наука и Техника, 2017. – 432с.
3. Маккинли У. Python и анализ данных. Саратов.: Профобразование, 2019. –

4. Горожанина Е.И. Нейронные сети. Самара: ПГУТиИ, 2017. 84с.
5. Павлова Информационные технологии: основные положения теории искусственных нейронных сетей. Новосибирск: НГУЭиУ «НИИХ», 2017. – 191с.

б) дополнительная литература:

1. Элбон К. Машинное обучение с использованием Python. Сборник рецептов. СПб: БХВ, 2020 – 384с.

2. Вандер П. Дж. Python для сложных задач: наука о данных и машинное обучение. СПб.: Питер, 2020. – 576с.

в) ресурсы сети Интернет: отсутствуют

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– операционная система Windows 7 или Windows 10 <https://www.microsoft.com/ru-ru/software-download/windows10>

– python (дистрибутив python) <https://www.python.org/?downloads>

– Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office On-eNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);

– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>

– Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>

– ЭБС Лань – <http://e.lanbook.com/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Для проведения лабораторных работ и самостоятельной работы используются аудитории учебно-вычислительной лаборатории ММФ При выполнении индивидуальных заданий, самостоятельных и лабораторных работ используется свободное и лицензионное программное обеспечение:

– офисный пакет Microsoft Office 2010 (составление отчетов);

– IDE для python (программа для организации работы на языке python).

15. Информация о разработчиках

Старченко Александр Васильевич, профессор, доктор физико-математических наук.

Данилкин Евгений Александрович, доцент, кандидат физико-математических наук.