

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

Физико-технический факультет

УТВЕРЖДАЮ:



Декан

Ю.Н. РЫЖИХ

20 23 г.

Рабочая программа дисциплины

Введение в машинное обучение

по направлению подготовки

15.04.06 Мехатроника и робототехника

Направленность (профиль) подготовки :
Моделирование робототехнических систем

Форма обучения

Очная

Квалификация

Магистр

Год приема

2023

Код дисциплины в учебном плане: Б1.В.07

СОГЛАСОВАНО:

Руководитель ОПОП

Г.Р. Шрагер

Председатель УМК

В.А. Скрипняк

Томск – 2023

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-4 Способен использовать современные информационные технологии и программные средства при моделировании технологических процессов;

ПК-2 Способность самостоятельно применять знания на практике по обработке и анализу научно-технической информации и результатов исследований, экспериментов и наблюдений.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК 4.1 Знать современные информационные технологии и программные средства для моделирования технологических процессов

ИОПК 4.2 Уметь применять современные информационные технологии при моделировании технологических процессов

ИОПК 4.3 Иметь навыки использования информационных технологий при моделировании технологических процессов

ИПК 2.1 Знать как осуществить и организовать сбор, анализ и систематизацию информации по проблеме исследования

ИПК 2.2 Уметь анализировать, интерпретировать, оценивать, представлять результаты проектов планов и программ проведения отдельных этапов работ.

ИПК 2.3 Владеть способами подготовки элементов документации и проведением отдельных этапов работ

2. Задачи освоения дисциплины

– Освоить ключевые понятия, цели и задачи использования машинного обучения при моделировании технологических процессов.

– Научиться чтению и анализу академической литературы по построению и оценке качества полученных моделей и технической документации по программным библиотекам и алгоритмам машинного обучения.

– Приобрести навыки разработчика методов и алгоритмов анализа данных.

– Научиться применять специализированные программные библиотеки для анализа и визуализации данных в решении практических задач профессиональной деятельности.

3. Место дисциплины в структуре образовательной программы

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

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, является обязательной для изучения.

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

Четвертый семестр, зачет

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: Программирование контроллеров, системы автоматизированного проектирования и производства, элементы электромашинной автоматики.

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 2 з.е., 72 часов, из которых:

-практические занятия: 28 ч.

в том числе практическая подготовка: 28 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Этапы развития искусственного интеллекта. Постановка задачи в машинном обучении и метрики качества вычислительных моделей.

Основные этапы эволюции искусственного интеллекта. Предмет и задачи машинного обучения и анализа данных. Основные принципы, задачи и подходы, использование в различных областях науки и индустрии. Обучающая, тестовая и валидационная выборка. Точность, полнота и анализ качества с помощью ROC кривой.

Тема 2. Источники данных. Библиотеки предварительной обработки данных.

Источники и основные типы данных. Повторение основ программирования на языке Python 3. Знакомство со специализированными библиотеками NumPy, SciPy, pandas. Матрицы. Индексирование, срезы. Объединение массивов. Запросы к таблицам: выборка строк/столбцов по заданным критериям. Модификация элементов таблицы. Добавление строк/столбцов. Группировка и агрегирование. Объединение таблиц (различные виды join). Построение сводных таблиц (pivot tables).

Тема 3. Алгоритмы обучения с учителем: классификация, регрессия, ранжирование.

Постановка задачи классификации, обзор основных методов ее решения. Бинарная и многоклассовая классификация. Разделение данных на обучающие и тестовые. Линейный регрессионный анализ. Реализация моделей на основе метода k -ближайших соседей. Метод логистической регрессии.

Тема 4. Алгоритмы обучения без учителя: кластеризация и уменьшение размерности.

Алгоритмы кластеризации с фиксированным количеством кластеров. Ансамбли алгоритмов машинного обучения. Методы отбора признаков. Оптимизация гиперпараметров. Метод главных компонент.

Тема 5. Визуализация полученных результатов.

Знакомство с различными методами предобработки данных, описательными статистиками и основными способами визуализации данных. Обзор библиотек matplotlib, seaborn.

9. Текущий контроль по дисциплине

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

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет в четвертом семестре проводится в письменной форме по билетам. Билет содержит теоретический вопрос и две задачи (проверяющих ИОПК-4.1, ИОПК-4.2, ИОПК-4.3, ИПК-2.1, ИПК-2.2, ИПК-2.3). Продолжительность зачета 1,5 часа.

Примерный перечень теоретических вопросов

1. 1. Преобработка. Масштабирование. Нормировка. Полиномиальные признаки. One-hot encoding.
2. Кластеризация. kMeans.
3. Смещение и дисперсия (bias and variance). Понятие средней гипотезы.
4. Ансамблевые методы.
5. Типы обучения: с учителем, без учителя, с подкреплением, с частичным участием учителя, активное обучение.
6. Метод опорных векторов.
7. Ошибка внутри и вне выборки. Валидация и кросс-валидация.
8. Линейная регрессия.
9. Логистическая регрессия.
10. Точность, полнота и ROC кривые. AUC.

Примеры задач:

1. Задача 1. Реализуйте алгоритм kNN классификации по k ближайшим соседям
Дано: исходный набор данных
Требуется: загрузить и предобработать исходный набор данных, реализовать программный модуль на языке Python 3 по решению задачи классификации и оценить качество полученной модели.

Формой промежуточной аттестации является зачет. Результаты сдачи зачета оцениваются отметкой «зачтено» или «не зачтено». В качестве оценочных средств контроля знаний применяются: контрольные вопросы и задачи.

- Оценка «зачтено» - студент знает программный материал, грамотно и по существу его излагает, не допускает существенных неточностей в ответе на вопрос, может правильно применять теоретические положения, правильно решена минимум одна задача.

- Оценка «не зачтено» - студент не знает значительной части программного материала, допускает существенные ошибки, не решена ни одна задача.

11. Учебно-методическое обеспечение

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

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

– Introduction to Machine Learning with Python: A Guide for Data Scientists / Andreas C. Müller, Sarah Guido – O'Reilly, 2017. – 376 с.

б) дополнительная литература:

– 1. Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning Series) / Murphy, Kevin P. MIT Press. 2014

– 2. Encyclopedia of Machine Learning and Data Mining / Sammut. Springer. 2016

– 3 Лутц М. Изучаем Python. Подробный справочник – М.: Символ-плюс, 2015 - 1280 с.

в) ресурсы сети Интернет:

– <http://www.machinelearning.ru> Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.

– <http://kaggle.com> Платформа для проведения конкурсов решения задач машинного обучения.

– <https://scipy.org>, <https://pandas.pydata.org>, <https://numpy.org> Техническая документация к библиотекам NumPy, SciPy, pandas.

13. Перечень информационных технологий

- а) лицензионное и свободно распространяемое программное обеспечение:
- Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office OneNote, 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.).
 - текстовый редактор Notepad++ или любой другой, поддерживающий подсветку синтаксиса, переключение между разными кодировками и поиск с использованием регулярных выражений;
 - набор библиотек для анализа данных Anaconda (<https://anaconda.org>)
 - среда разработки Jupyter Notebook (<https://jupyter.org>);
 - язык программирования Python (<https://www.python.org>) ;

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.

Вычислительные серверы суперкомпьютера Томского государственного университета.

15. Информация о разработчиках

Гойко Вячеслав Леонидович, Центр прикладного анализа больших данных, директор; ассистент, кафедра теоретических основ информатики (ИПМКН)