



Федеральное государственное бюджетное учреждение науки
Институт автоматики и электрометрии Сибирского отделения Российской академии наук
(ИАиЭ СО РАН)

УТВЕРЖДЕНА

приказом директора ИАиЭ СО РАН
от 23.05.2024 № 241

Рабочая программа раздела
«МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ»
дисциплины **«МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ, ЧИСЛЕННЫЕ
МЕТОДЫ И КОМПЛЕКСЫ ПРОГРАММ»**

Научная специальность: 1.2.2 Математическое моделирование, численные
методы и комплексы программ
Форма обучения: очная

№	Вид деятельности	Семестр
		5
1	Лекции, час.	12
2	Практические занятия (семинары), час.	24
3	Самостоятельная работа, час.	72
4	Всего зачетных единиц	3

Форма аттестации – дифференцированный зачет

Новосибирск – 2024

1. Аннотация к рабочей программе дисциплины

Целью подготовки по дисциплине «Математическое моделирование, численные методы и комплексы программ», раздел «Методы машинного обучения» программы аспирантуры по научной специальности 1.2.2 Математическое моделирование, численные методы и комплексы программ является ознакомление аспирантов с теоретическими основами и основными принципами машинного обучения.

Содержание дисциплины охватывает круг вопросов, связанных с основами интеллектуального анализа данных, включая преобразование и очистку данных, работу с пропущенными значениями, основные способы визуализации данных, корреляционный анализ, различные методы отбора признаков, задачи снижения размерности данных, кластеризации, классификации, регрессии. Аспиранты освоят работу со специализированными программными библиотеками для визуализации и анализа данных и научатся применять полученные знания для решения практических задач, в том числе, загружать данные, сохраненные в разных форматах, выбирать и группировать нужные записи по заданным критериям, строить предсказательные модели и оценивать их качество.

Преподавание раздела дисциплины предусматривает следующие формы организации учебного процесса: лекции, практические занятия (семинары), самостоятельная работа.

Общая трудоемкость составляет 3 зачетных единицы, (108 часов).

2. Место дисциплины в образовательной программе

Рабочая программа раздела «Методы машинного обучения» программы аспирантуры по научной специальности 1.2.2 Математическое моделирование, численные методы и комплексы программ относится к образовательному компоненту и направлена на подготовку к сдаче кандидатского экзамена, является обязательной для освоения обучающимися в 5 семестре (третий год обучения в аспирантуре). Раздел «Методы машинного обучения» является базовым для выполнения научно-исследовательской деятельности и подготовки диссертации на соискание ученой степени кандидата наук.

3. Образовательные технологии

Технология процесса обучения аспирантов включает в себя следующие образовательные мероприятия:

- аудиторные занятия (лекции);
- практические занятия (семинары);
- самостоятельная работа аспирантов.

В процессе изучения дисциплины, как лектором, так и обучающимися используется метод проблемного изложения материала, самостоятельное чтение аспирантами учебной, учебно-методической и справочной литературы, анализ информационных ресурсов в научных библиотеках и сети Internet по актуальным проблемам и последующие свободные дискуссии по освоенному ими материалу.

Аудиторные занятия проводятся с использованием информационно-телекоммуникационных технологий: учебный материал представлен также в виде мультимедийных презентаций. Презентации позволяют четко структурировать материал занятия.

Самостоятельная работа организована в соответствии с технологией проблемного обучения и предполагает следующие формы активности:

- поиск научной информации в открытых источниках с целью ее анализа и выявления ключевых особенностей исследуемых явлений;
- самостоятельная проработка учебно-проблемных задач, выполняемая с привлечением основной и дополнительной литературы;
- решение проблемных задач стимулируют познавательную деятельность и научно-исследовательскую активность аспирантов.

4. Требования к результатам освоения дисциплины

В результате изучения раздела «Методы машинного обучения» программы аспирантуры по научной специальности 1.2.2 Математическое моделирование, численные методы и комплексы программ и в соответствии с программой кандидатского экзамена аспирант должен:

знать:

- методы предварительной обработки данных, методы отбора информативных признаков, методы классификации, методы регрессионного анализа, методы анализа текстовых данных;
- возможности актуальных алгоритмов машинного обучения, которые широко используются на практике, основные сферы их применения;
- инструментальные средства, применяемые для контроля принимаемых проектных решений;
- особенности работы со специализированными программными библиотеками языка программирования Python для анализа данных и решения задач машинного обучения.

уметь

- анализировать многомерные данные и преодолевать вычислительные

проблемы, связанные с высокой размерностью данных применять методы машинного обучения при решении задач в различных прикладных областях;

– использовать библиотеки языка Python для построения моделей машинного обучения.

владеть:

– построения и проверки качества моделей машинного обучения;

– интерпретации полученных результатов в терминах прикладной области с целью получения новых знаний и выводов;

использования библиотек языка Python для построения систем, обучающихся по прецедентам.

5. Структура и содержание раздела учебной дисциплины

Раздел, тема занятия		Виды аудиторных занятий (час.)		
		Лекции	Семинары	Итого
Семестр 5		12	24	36
1.	<i>Введение. Общие понятия. Алгоритмическое обучение и обучение потоком образцов данных. Основные виды машинного обучения (МО) - с учителем, без учителя, с подкреплением. Классы задач МО – регрессия, классификация, кластеризация, уменьшение размерности, поиск аномалий. Основные алгоритмы МО. Место МО в структуре Искусственного интеллекта.</i>	2		2
2.	<i>Цикл машинного обучения Стадии машинного обучения (сбор данных, подготовка данных, обработка данных, анализ данных, обучение модели, тестирование модели, развертывание). Источники данных МО. Типы данных. Подготовка данных. Предварительная обработка данных. Матрица диаграмм рассеяния. Корреляционная матрица и ее тепловая карта. Разбиение датасета. Обучение модели. Тестирование модели.</i>		2	2

3.	<p><i>Регрессионный анализ в машинном обучении</i> <i>Определение регрессионного анализа, суть регрессии. Виды регрессии – линейная, полиномиальная, логистическая. Метод наименьших квадратов. Простая и множественная линейные регрессии. Критерий R2 для оценки линейной регрессии.</i></p>	2	2	4
4.	<p><i>Алгоритмы классификации в машинном обучении.</i> <i>Определение алгоритма классификации. Двоичный и многоклассовый классификаторы. Ленивое и жадное обучение. Линейные и нелинейные модели классификаторов. Оценка модели классификации - логарифм потерь, матрица путаницы, кривая AUC-ROC. Ошибки первого и второго рода.</i></p>		2	2
5.	<p><i>Дерево решений</i> <i>Определение ДР. Терминология дерева решений. Графическое представление ДР. Основной алгоритм дерева решений. Разновидности алгоритмов ДР - ID3, C4.5, CART, CHAID. Показатели выбора основного атрибута – энтропия, прирост информации, индекс Джини, коэффициент усиления, уменьшение дисперсии, критерий Хи-квадрат. Обрезка: получение оптимального дерева решений.</i></p>	2	2	4
6.	<p><i>Ансамблевые методы</i> <i>Процесс ансамблирования. Виды ансамблей - стекинг, бэггинг, бустинг. Случайный лес. Варианты Бустинга – AdaBoost, градиентный бустинг, XGBoost, CatBoost.</i></p>		4	4

7.	<p>Алгоритмы классификации <i>K</i> ближайших соседей (KNN). Логистическая регрессия – уравнение ЛР, логистическая кривая. Классификация данных при помощи нейронных сетей - модель перцептрона, популярные функции активации, обучение перцептрона, метод градиентного спуска и метод обратного распространения ошибки, функция Softmax. Машины опорных векторов (SVM). Наивный Байес.</p>	2	2	4
8.	<p>Кластеризация в машинном обучении Определение кластеризации. Типы кластерных структур. Жесткая и мягкая кластеризация. Виды алгоритмов кластеризации - кластеризация разделением, кластеризация на основе плотности, кластеризация на основе модели распределения, иерархическая кластеризация, нечеткая кластеризация.</p>	2	2	4
9.	<p>Основные алгоритмы кластеризации. Алгоритм <i>K</i>-средних, метод локтя, метод силуэта, улучшенный алгоритм <i>K</i>-средних, алгоритм среднего сдвига, DBSCAN, Агломеративный иерархический алгоритм.</p>		2	2
10.	<p>Алгоритмы сокращения размерности Пороговая обработка дисперсии. Анализ главных компонент – ковариационная матрица, ее собственные вектора и собственные значения). Алгоритм Relief.</p>		2	2
11.	<p>Анализ временных рядов Определение временного ряда. Задачи идентификации и прогноза. Дискретизация и квантования сигнала. Теорема Котельникова. Детерминированные и случайные сигналы. Виды шумов. Операции над временными рядами (сдвиг, изменение шага, скользящее среднее). Разложение временного ряда на компоненты (тренд, сезонность, случайные колебания). Экспоненциальное сглаживание.</p>	2	2	4

12.	<i>Обнаружение аномалий Понятие аномалии. Приложения поиска аномалий. Типы аномалий. Методы обнаружения аномалий: вероятностный подход, линейные методы, метрические методы, гармоническое и полиномиальное оценивание, Isolation Forest (Изолирующий лес).</i>		2	2
-----	---	--	---	---

Самостоятельная работа с литературой.

6. Литература

Крутиков, В.Н. Анализ данных : учебное пособие / В.Н. Крутиков, В.В. Мешечкин ; Кемеровский государственный университет. – Кемерово : Кемеровский государственный университет, 2014. – 138 с. : ил.– URL: <http://biblioclub.ru/index.php?page=book&id=278426>– Библиогр. в кн. – ISBN 978-5-8353-1770-7. – Текст : электронный.

2. Каган, Е.С. Прикладной статистический анализ данных : учебное пособие : [16+] /Е.С. Каган ; Кемеровский государственный университет. – Кемерово : Кемеровский государственный университет, 2018. – 235 с.

3. Шелудько, В.М. Основы программирования на языке высокого уровня Python : учебное пособие / В.М. Шелудько ; Министерство науки и высшего образования РФ, Южный федеральный университет, Инженерно-технологическая академия. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2017. – 147 с. : ил.

4. Глебов, В.И. Практикум по математической статистике: проверка гипотез с использованием Excel, MatCalc, R и Python : [16+] / В.И. Глебов, С.Я. Криволапов ; Финансовый университет при Правительстве Российской Федерации. – Москва : Прометей, 2019. – 87 с. : ил.

7. Электронные энциклопедические издания свободного доступа

1. Machine Learning for Developers - <https://xyclade.github.io/MachineLearning/>

2. Видео лекции по Машинному обучению - <https://yandexdataschool.ru/edu-process/courses/machine-learning>

3. Машинное обучение с примерами - <http://www.uic.unn.ru/~zny/ml/Lectures/>