

Аннотация рабочей программы дисциплины

Б1.В.ДВ.05.02 «Методы искусственного интеллекта в задачах классификации»

Курс 4 Семестр 7 Количество з.е. 2

Объем трудоемкости: 2 зачетных единиц (72 ч., из них – 18,2 час. аудиторной нагрузки: лабораторных работ - 16 ч., 53,8 часов самостоятельной работы, 2 часа КСР, 0,2 часа ИКР.), форма контроля – зачет.

Цель. освоения дисциплины: «Методы искусственного интеллекта в задачах классификации» является формирование у студентов систематизированных знаний, практических умений и навыков применения современных методов искусственного интеллекта, машинного обучения для решения задач классификации данных в различных предметных областях.

Дисциплина направлена на развитие способности выбирать, реализовывать, оценивать и интерпретировать модели классификации.

Задачи дисциплины

1. Изучение теоретических основ задач классификации в машинном обучении;
2. Освоение базовых и современных алгоритмов классификации (логистическая регрессия, SVM, деревья решений, байесовский классификатор, ансамбли, нейронные сети);
3. Приобретение практических навыков предобработки данных, проектирования, обучения, оценки и оптимизации моделей классификации с использованием современных инструментов (R, Python, scikit-learn, PyTorch/TensorFlow/Keras);
4. Развитие умений анализировать результаты классификации, выбирать метрики качества, интерпретировать работу моделей;
5. Формирование навыков применения методов классификации для решения прикладных задач.

Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Методы искусственного интеллекта в задачах классификации» относится к части, формируемой участниками образовательных отношений Блока 1 "Дисциплины (модули) по выбору" учебного плана (Б1.В.ДВ.05.02), является альтернативой для дисциплины Б1.В.ДВ.05.01 «Анализ данных машинного обучения».

Дисциплина изучается в 7-м семестре. Для успешного освоения необходимы знания, полученные в дисциплинах: «Алгебра и введение в тензорный анализ», «Теория вероятностей и математическая статистика», «Многомерный статистический анализ и машинное обучение», «Программирование».

Преподавание ведется в виде лабораторных занятий с использованием интерактивных методов. Лабораторные работы направлены на практическое освоение методов и инструментов классификации на реальных данных.

Дисциплина формирует компетенции, необходимые для выполнения выпускной квалификационной работы и профессиональной деятельности в области вычислительных технологий.

Результаты обучения (знания, умения, опыт, компетенции):

BD-1	Способен осуществлять поиск, сбор, очистку и предварительный анализ данных (П)
BD-1.1	Обосновывает способы и варианты применения методов предварительного анализа данных в задачах ИИ, включая их математическое (алгоритмическое) преобразование и адаптацию к специфике задачи
	Знает методы заполнения пропусков в данных и удаления выбросов в табличных данных (случайные величины) Имеет навыки (умеет) очистки зашумленных временных рядов и изображений. Обнаруживает и устраняет выбросы в данных временных рядов. Владеет подходами к заполнению пропусков в данных временных рядов и изображений.
BD-1.2	Применяет методы анализа данных для проверки разведочных гипотез и подготовки данных к применению современных методов ИИ
	Знает основные методы понижения размерности Умеет применить основные методы понижения размерности и подбирает оптимальную размерность в зависимости от необходимой доли объяснённой дисперсии. Владеет методологией применения существующих библиотек, реализующих методы понижения размерности.
BD-1.3	Применяет методы понижения размерности для первичной интерпретации и визуализации многомерных данных
	Знает и умеет применить основы методов отбора признаков и выбирает оптимальное подмножество признаков. Владеет методологией применения существующих библиотек, реализующих методы отбора признаков.
BD-1.4	Знает и умеет применить методы отбора признаков. Владеет способностью применять методы отбора признаков данных, значимых для исследования.
	Умеет отбирать признаки данных, значимые для исследования, Владеет методами finetuning
BD-2	Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных
BD-2.1	Знает, как сформировать требования для набора данных. Владеет умениями по формированию требований к наборам и качеству данных для решения задач машинного обучения
BD-2.2	Знает приемы и инструменты для сбора данных из разрозненных источников. Умеет работать с данными, в том числе собирает данные из разрозненных источников, проверяет данные на корректность. Владеет языками и инструментами для сбора данных и оценки их корректности.
LLM-2	Способен дообучать, адаптировать и оптимизировать генеративные модели под специфические задачи и условия применения
LLM-2.1	Понимает принципы fine-tune Знает: основные подходы к тонкой настройке: полная настройка всех параметров, поэтапная разморозка слоев, методы эффективной тонкой настройки (P-Tuning, LoRA, QLoRA, Adapter). Гиперпараметры, критически важные для fine-tune: learning rate, scheduler, batch size, и их отличия от обучения с нуля. Умеет: Отличать дообучение от первичного обучения, знает базовые процедуры fine-tune , анализировать задачу и выбирать наиболее подходящий метод fine-tune

	(полная настройка vs. эффективные методы). Владеет: Навыком осознанного выбора стратегии fine-tune под ограничения (вычислительные ресурсы, объем данных, требования к качеству). Применяет fine-tune к предобученным моделям на новых датасетах. • Методами анализа и интерпретации процесса дообучения (использование логов, графиков потерь). • Критическим мышлением для оценки целесообразности применения fine-tune в конкретном сценарии versus использования prompt engineering или RAG.
LLM-2.2	Создаёт обучающие наборы данных. Знает: Требования к данным для fine-tune: релевантность, объем, разнообразие, качество разметки. Форматы данных для популярных фреймворков (Hugging Face, TensorFlow, PyTorch) и структур задач (текст-текст, текст-изображение, инструкции и т.д.). Методы аугментации данных (data augmentation), специфичные для генеративных моделей (e.g., back-translation для текста, модификация промптов). Принципы разбиения данных на обучающую, валидационную и тестовую выборки. Умеет: Выбирать методы с учетом требований к latency и ресурсам. собирать данные из различных источников: API, веб-скрапинг, открытые датасеты, синтетическая генерация. Очищать и предобрабатывать сырые данные: удаление шума, дубликатов, нормализация текста, приведение к единому формату. Размечать данные в соответствии с поставленной задачей (e.g., составлять пары "инструкция-ответ", аннотировать изображения). Применять методы аугментации данных для увеличения размера и разнообразия обучающего набора. Владеет: Навыками работы с библиотеками и инструментами для обработки данных (Pandas, NumPy, Hugging Face Datasets). Методами обеспечения репрезентативности и сбалансированности создаваемого набора данных. Технологиями создания синтетических данных для задач, где реальных данных недостаточно. Полным циклом подготовки данных: от сбора сырых данных до формирования готового для обучения объекта (DataLoader, Dataset)
MF-4	Способен применять статистические методы для анализа данных, валидации моделей машинного обучения и проведения экспериментов в области ИИ.
MF-4.1	Применяет статистические методы анализа и машинного обучения для решения задач анализа данных и проведения экспериментов на данных. Применяет и выбирает методы статистического машинного обучения, учитывая особенности данных и задачи, а также объясняет различия между подходами. Знает основные статистические методы описательного и инференционного анализа, принципы планирования экспериментов (A/B-тесты) и базовые алгоритмы машинного обучения (линейные модели, деревья). Умеет применять статистические методы (проверка гипотез, анализ распределений) и алгоритмы машинного обучения для исследования данных, извлечения инсайтов и проверки рабочих гипотез. Владеет навыком проведения полного цикла анализа данных: от предобработки и разведочного анализа (EDA) до построения, интерпретации результатов и формирования выводов.
MF-4.2	Способен применять статистические методы для построения предсказательных моделей, включая методы для анализа и прогнозирования временных рядов, а также моделирования нестационарных случайных процессов.

	Строит модели динамических систем для многомерных временных рядов и полей. Знает математические основы и предположения регрессионных, прогнозных моделей и методов анализа временных рядов (ARIMA, экспоненциальное сглаживание, подходы к работе с нестационарностью). Умеет строить, обучать и валидировать предсказательные модели (регрессия, классификация, прогнозирование), включая работу с временными рядами и нестационарными процессами. Владеет навыком выбора и настройки модели под конкретную задачу прогнозирования, диагностики её качества и интерпретации результатов прогноза.
MF-4.3	Способен применять статистические методы для оценки качества моделей ИИ, включая метрики и критерии для регрессии, классификации и кластеризации, а также для проведения статистических тестов для сравнения моделей. Оценивает статистические различия моделей и алгоритмов, обучаемых на данных. Знает и применяет модифицированные статистические критерии, A/B тестирование. Применяет оценивание на основе модифицированных доверительных интервалов, использует Байесовские тесты.
MF-7	Способен применять методы дифференциальной геометрии и топологии для формализации, анализа и интерпретации структур данных и признаков пространств, включая задачи отображения, кластеризации, обучения на многообразиях и анализа устойчивости моделей.
MF-7.1	Применяет методы топологического анализа для описания глобальных свойств данных и устойчивости признаков структур. Знает: Узнаёт и интерпретирует базовые топологические характеристики (связность, количество компонент, размерность) в примерах и визуализациях. Умеет: Использовать топологические дескрипторы в качестве новых признаков для модели классификации, характеризующих глобальную форму данных. Оценивать топологическую устойчивость признакового пространства к малым возмущениям в данных. Владеет: Навыком чтения и интуитивной интерпретации персистентных диаграмм для быстрой оценки сложности структуры данных. • Методом использования TDA как инструмента для выявления неочевидных глобальных закономерностей, не улавливаемых традиционными статистическими методами.
ML-2	Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками
ML-2.3	Решает проблемы несбалансированных данных и оценивает качество моделей Знает проблемы несбалансированных данных методы оценивания качества моделей/ Умеет применить на практике основные метрики оценки качества для задач классификации и регрессии (Б) Применяет различные типы кросс-валидации Оценивает качество моделей с учетом bias-variance trade-off (П) Владеет продвинутыми методами работы с несбалансированными данными (SMOTE weighted learning). Настраивает кастомные метрики и функции потерь. Проводит статистический анализ значимости результатов (Э)

Содержание и структура дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.
 Разделы дисциплины, изучаемые в _6_ семестре (очная форма)

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеауди- торная работа
			Л	ПЗ	ЛР	
	Раздел 1. Введение в классификацию					
1.	Задачи классификации в ИИ и МО. Постановка задачи. Типы признаков. Этапы решения задачи классификации.	5			1	4
2.	Основы дифференциальной геометрии и топологии в интеллектуальном анализе данных. Метрики качества классификации. Матрица ошибок, Accuracy, Precision, Recall, F1-score, ROC-AUC. Кросс-валидация. Стратегии работы с дисбалансом классов. Feature Engineering	5			1	4
	Раздел 2. Базовые алгоритмы классификации. Методы визуализации					
3.	Линейные модели: Логистическая регрессия. Метод опорных векторов (SVM). Ядра. Деревья решений. Анализ качества модели. Применение методов топологического анализа	9			1	8
4.	Методы визуализации и понижения размерности как инструмент геометрического исследования. Анализ качества модели. . Применение методов топологического анализа	6			1	5
	Раздел 3. Ансамблевые методы и продвинутые техники					
5.	Бэггинг. Случайный лес. Стратегии работы с дисбалансом классов. Feature Engineering. Стратегии работы с дисбалансом классов. Feature Engineering.	4			1	5
6.	Бустинг (AdaBoost, Gradient Boosting,. XGBoost, LightGBM, CatBoost). Стратегии работы с дисбалансом классов. Feature Engineering.	6			1	5
	Раздел 4. Нейронные сети для классификации					
7.	Многослойные перцептроны. (MLP) для классификации.	6			1	5
8.	Свёрточные нейронные сети (CNN) для классификации изображений.	6			1	5
	Раздел 5. Практика и применение					
9.	Кейсы применения классификации (текст, изображения, звук, табличные данные). Интерпретация моделей (SHAP, LIME).	16,8			8	12,8
	<i>ИТОГО по разделам дисциплины</i>	69,8			16	53,8
	Контроль самостоятельной работы (КСР)	2				
	Промежуточная аттестация (ИКР)	0,2				
	Подготовка к текущему контролю	-				
	Общая трудоемкость по дисциплине	72				

Примечание: Л – лекции, КСР – контрольные и самостоятельные работы, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

Курсовые проекты или работы.

Не предусмотрены учебным планом

Вид аттестации: ЛР, Комплексная итоговая работа, зачет.

Автор Приходько Т.А. – кандидат технических наук, доцент кафедры вычислительных технологий;