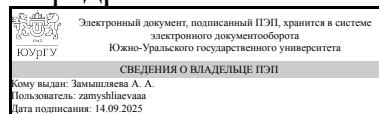


ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:
Заведующий выпускающей
кафедрой



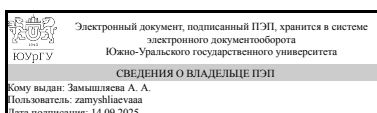
А. А. Замышляева

РАБОЧАЯ ПРОГРАММА

дисциплины 1.Ф.П0.08 Машинное обучение на нестандартных данных
для направления 01.03.02 Прикладная математика и информатика
уровень Бакалавриат
профиль подготовки Искусственный интеллект, глубокое обучение и анализ данных
форма обучения очная
кафедра-разработчик Центр ОП топ-уровня в сфере ИИ "ВиртУм"

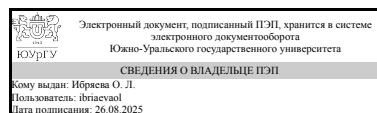
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 01.03.02 Прикладная математика и информатика, утверждённым приказом Минобрнауки от 10.01.2018 № 9

Зав.кафедрой разработчика,
д.физ.-мат.н., проф.



А. А. Замышляева

Разработчик программы,
к.физ.-мат.н., доцент



О. Л. Ибряева

1. Цели и задачи дисциплины

Сформировать у студентов теоретические знания и практические навыки по применению методов машинного обучения в условиях нестандартных данных: малых выборок, несбалансированных распределений, больших объёмов, сдвига данных (drift) и высокого уровня шума. Студенты освоят современные подходы к повышению обобщаемости, устойчивости и эффективности моделей, включая few-shot learning, domain adaptation, федеративное обучение, робастные методы и методы снижения размерности. Особое внимание уделяется обоснованию выбора алгоритмов, анализу trade-off между производительностью, безопасностью и вычислительной стоимостью. Задача курса — подготовить студентов к решению прикладных задач в условиях ограниченных, неоднородных или распределённых данных с использованием современных инструментов машинного обучения.

Краткое содержание дисциплины

В рамках дисциплины изучаются типы нестандартных данных и вызовы, с которыми сталкиваются классические методы машинного обучения. Рассматриваются методы работы с несбалансированными данными (oversampling, undersampling), обучение на малых выборках (few-shot, метаобучение, байесовские методы), робастные подходы (RANSAC), масштабирование на большие данные (mini-batch, online, federated learning), а также методы повышения обобщаемости (domain adaptation, transfer learning, регуляризация). Отдельное внимание уделено сравнительному анализу моделей по эффективности, устойчивости и безопасности. Практические занятия ориентированы на реализацию и адаптацию алгоритмов с использованием Python, Scikit-learn, PyTorch и специализированных библиотек.

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-13 [ML-8] Способен применять алгоритмы обучения на нестандартных объемах данных	Знает: - [И-1, ПУ] основные типы нестандартных данных (малые выборки, несбалансированные данные, большие объёмы); базовые алгоритмы (например, oversampling, undersampling, mini-batch learning) для типовых задач; ограничения стандартных методов при работе с такими данными Умеет: - [И-1, ПУ] обосновывать выбор методов повышения эффективности и обобщаемости (регуляризация, уменьшение размерности модели, domain adaptation, использовать разностных методов типа сиамских сетей, few-shot learning, байесовские методы) [И-2, БУ] проводить сравнительный анализ моделей по эффективности, устойчивости и безопасности; учитывать переобучение, drift и computational cost Имеет практический опыт: - [И-1, ПУ] адекватного выбора и адаптации алгоритмов (transfer learning, few-shot learning, federated

	learning) с учетом специфики нестандартных объемов данных и требований к задаче
--	---

3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Глубокие нейронные сети	Не предусмотрены

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Дисциплина	Требования
Глубокие нейронные сети	Знает: -[И-1, ПУ] TD-методы и методы Монте-Карло для обучения агента, -[И-2, ПУ] стандартные метрики оценки качества (ассигасу, F1, ROC-AUC), -[И-1, ПУ] принцип и алгоритмы градиентного спуска Умеет: -[И-1, ПУ] задавать цель агента с помощью полного вознаграждения, вознаграждения с обесценением, лямбда-дохода, -[И-1, БУ] проводить аппроксимацию функции ценности агента с помощью глубоких нейронных сетей; реализовать поиск по дереву методом Монте-Карло, -[И-2, ПУ] использовать стандартные метрики оценки качества (ассигасу, F1, ROC-AUC) и сопоставлять базовые модели между собой, -[И-1, ПУ] применять регуляризацию и прореживание; выбирать размер пакета для стохастического градиентного спуска [И-2, БУ] применять основные архитектуры глубокого обучения (VGG, ResNet) Имеет практический опыт: -[И-1, ПУ] использования TD-методов и методов Монте-Карло для обучения агента, -[И-2, БУ] создания агентной системы с помощью глубоких нейронных сетей на основе обучения с подкреплением, -[И-1, ПУ] выбора и задания скорости обучения и функции потерь в зависимости от задачи и набора данных

4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 54,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах
		Номер семестра
		7
Общая трудоёмкость дисциплины	108	108
Аудиторные занятия:	48	48

Лекции (Л)	16	16
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	0	0
Лабораторные работы (ЛР)	32	32
Самостоятельная работа (СРС)	53,5	53,5
Изучение материала в пособии "Аугментация данных для решения проблемы обучения на малых данных в технической диагностике"	5	5
Изучение и анализ статей уровня 1 Белого списка (статьи в свободном доступе)	35	35
Подготовка к зачету	13,5	13,5
Консультации и промежуточная аттестация	6,5	6,5
Вид контроля (зачет, диф.зачет, экзамен)	-	диф.зачет

5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Типология нестандартных данных и базовые методы коррекции	12	4	0	8
2	Методы обучения на малых и зашумлённых выборках	14	4	0	10
3	Масштабирование и распределённое обучение	10	4	0	6
4	Повышение обобщаемости, устойчивости и сравнительный анализ	12	4	0	8

5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Введение в нестандартные данные: малые выборки, несбалансированные данные, большие объёмы. Ограничения классических методов.	2
2	1	Методы балансировки данных: oversampling (SMOTE, ADASYN), undersampling. Метрики качества для несбалансированных задач.	2
3	2	Обучение на малых выборках: few-shot learning, метаобучение (MAML), сиамские сети. Байесовские подходы к регуляризации.	2
4	2	Робастные методы: RANSAC, Huber-регрессия, устойчивость к выбросам и сдвигам распределений (drift).	2
5	3	Обучение на больших данных: mini-batch, online learning, методы оптимизации. Ограничения и шумность оценок.	2
6	3	Федеративное обучение: архитектура, методы согласования (FedAvg, FedProx), privacy, работа с non-IID данными.	2
7	4	Методы повышения обобщаемости: transfer learning, domain adaptation (MMD, DANN), регуляризация, снижение размерности.	2
8	4	Сравнительный анализ моделей: эффективность, устойчивость, безопасность. Переобучение, computational cost, adversarial robustness.	2

5.2. Практические занятия, семинары

Не предусмотрены

5.3. Лабораторные работы

№ занятия	№ раздела	Наименование или краткое содержание лабораторной работы	Кол-во часов
1	1	Анализ и визуализация структуры нестандартных данных: выявление дисбаланса классов, малых выборок и наличия выбросов в реальных и синтетических датасетах	4
2	1	Применение методов балансировки несбалансированных данных (SMOTE, RandomUnderSampler) и оценка качества моделей с использованием метрик F1, AUC-PR и balanced accuracy	4
3	2	Реализация и тестирование сиамской нейронной сети для few-shot классификации на датасете Omniglot с использованием метрического обучения и парного сравнения изображений	6
4	2	Применение робастных методов линейной регрессии (RANSAC, Huber-регрессия) на зашумлённых данных и сравнение их устойчивости с классической OLS-регрессией	4
5	3	Симуляция федеративного обучения для диагностики асинхронного двигателя по току статора в условиях распределённых и неоднородных (non-IID) данных (кейс от партнера)	6
6	4	Реализация метода domain adaptation для повышения устойчивости модели диагностики подшипников при изменении нагрузки и скорости вращения (CWRU) (кейс от партнера)	4
7	4	Командный мини-проект: комплексное решение задачи на нестандартных данных с выбором, адаптацией и обоснованием алгоритмов (few-shot, federated, domain adaptation) и защитой решения (кейс от партнера)	4

5.4. Самостоятельная работа студента

Выполнение CPC			
Подвид CPC	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Изучение материала в пособии "Аугментация данных для решения проблемы обучения на малых данных в технической диагностике"	Приложенное методическое пособие	7	5
Изучение и анализ статей уровня 1 Белого списка (статьи в свободном доступе)	Примеры статей (к началу курса список будет обновлен на более современные и актуальные статьи) 1. Yu, F.; Xiu, X.; Li, Y. A Survey on Deep Transfer Learning and Beyond. Mathematics 2022, 10, 3619. https://doi.org/10.3390/math10193619 (статья содержит современный обзор по методам transfer learning) 2. Abreha, H.G.; Hayajneh, M.; Serhani, M.A. Federated Learning in Edge Computing: A Systematic Survey. Sensors 2022, 22, 450. https://doi.org/10.3390/s22020450 (статья посвящена интеграции федеративного обучения с edge computing) 3. Martínez-Otzeta, J.M.; Rodríguez-Moreno, I.; Mendialdua, I.; Sierra, B. RANSAC for Robotic Applications: A Survey. Sensors	7	35

	2023, 23, 327. https://doi.org/10.3390/s23010327 (статья представляет обзор методов RANSAC и их модификаций, ориентированных на робастную оценку параметров в условиях большого количества выбросов.) 4. Ren, J.; Li, C.; An, Y.; Zhang, W.; Sun, C. Few-Shot Fine-Grained Image Classification: A Comprehensive Review. AI 2024, 5, 405-425. https://doi.org/10.3390/ai5010020 (статья фокусируется на задачах few-shot fine-grained классификации изображений.) 5. Walsh, R.; Tardy, M. A Comparison of Techniques for Class Imbalance in Deep Learning Classification of Breast Cancer. Diagnostics 2023, 13, 67. https://doi.org/10.3390/diagnostics13010067 (статья представляет эмпирическое исследование методов борьбы с дисбалансом классов в задаче диагностики рака молочной железы, включая oversampling, undersampling и синтетическую генерацию)		
Подготовка к зачету	Вся основная литература.	7	13,5

6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учитывается в ПА
1	7	Текущий контроль	Лабораторная работа 1	10	10	Загрузка и первичный анализ данных (info, describe) — 1 балл Построение гистограмм и boxplot — 2 балла Количественная оценка дисбаланса классов — 2 балла Выявление выбросов (IQR или Z-score) — 2 балла Вывод о типах нестандартности и возможных методах — 3 балла	дифференцированный зачет
2	7	Текущий контроль	Лабораторная работа 2	10	10	Корректное разделение данных — 1 балл Применение RandomUnderSampler — 2	дифференцированный зачет

						балла Применение SMOTE — 2 балла Обучение модели и расчёт всех метрик — 3 балла Сравнение и обоснованный вывод — 2 балла	
3	7	Текущий контроль	Лабораторная работа 3	15	15	Подготовка пар изображений — 3 балла Корректная архитектура сети (shared weights) — 4 балла Реализация contrastive loss и обучение — 4 балла Тестирование и accuracy $\geq 70\%$ — 2 балла Вывод и анализ — 2 балла	дифференцированный зачет
4	7	Текущий контроль	Лабораторная работа 4	10	10	Генерация данных с выбросами — 2 балла Обучение всех трёх моделей (линейная регрессия (OLS), RANSACRegressor, HuberRegressor) — 3 балла Визуализация линий регрессии — 2 балла Сравнение метрик (таблица) — 2 балла Обоснование выбора робастного метода — 1 балл	дифференцированный зачет
5	7	Текущий контроль	Лабораторная работа 5	15	15	Качественная предобработка сигнала и извлечение признаков — 3 балла Корректное создание non-IID клиентов (разделение по нагрузке) — 2 балла Реализация FedAvg (или использование Flower) — 5 баллов График сходимости и сравнение с централизованным обучением — 3 балла Анализ влияния non-IID и устойчивости — 2 балла	дифференцированный зачет
6	7	Текущий контроль	Лабораторная работа 6	10	10	Подготовка данных — 2 балла Обучение source-only модели и тестирование на целевых данных — 2 балла Реализация DANN или MMD — 4 балла Сравнение результатов и обоснованный вывод — 2 балла	дифференцированный зачет
7	7	Текущий контроль	Командный мини-проект	20	20	Команда (2–3 чел.) выбирает задачу из предложенного списка или предлагает свою,	дифференцированный зачет

					согласованную с преподавателем. Проводит анализ данных: выявляет тип нестандартности (малые выборки, дисбаланс, шум, drift, non-IID и др.). Выбирает и адаптирует один или несколько методов (few-shot learning, domain adaptation, federated learning, регуляризация, байесовские подходы и др.). Реализует решение, оценивает его эффективность, устойчивость и вычислительную стоимость. Подготавливает отчёт (5–7 стр.) и презентацию (5–7 слайдов). Защищает проект. Порядок начисления баллов (макс. 20): Актуальность и сложность задачи — 3 балла Качество анализа данных и выявления нестандартности — 3 балла Обоснованность выбора и адаптации методов — 5 баллов Качество реализации и результатов — 5 балла Качество отчёта и презентации — 2 балла Защита и ответы на вопросы — 2 балла	
8	7	Промежуточная аттестация	Зачетная работа	-	100 Студент получает зачётный билет, содержащий два теоретических вопроса и одну практическую задачу. На подготовку отводится 60 минут. Оценка выставляется по 100-балльной шкале с учётом полноты и правильности ответов на теоретические вопросы (максимум 30 баллов), качества анализа практической задачи — выявления типов нестандартных данных (10 баллов), обоснованности выбора и адаптации методов (25 баллов), оценки эффективности и устойчивости модели (15 баллов), а также	дифференцированный зачет

					ясности изложения, структурированности ответа и грамотного использования терминологии (10 баллов). Дополнительно начисляется до 10 баллов за ответы на уточняющие вопросы (прилагаются).	
--	--	--	--	--	--	--

6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
дифференцированный зачет	Студент получает зачётный билет, содержащий два теоретических вопроса и одну практическую задачу. На подготовку отводится 60 минут. Оценка выставляется по 100-балльной шкале с учётом полноты и правильности ответов на теоретические вопросы (максимум 30 баллов), качества анализа практической задачи — выявления типов нестандартных данных (10 баллов), обоснованности выбора и адаптации методов (25 баллов), оценки эффективности и устойчивости модели (15 баллов), а также ясности изложения, структурированности ответа и грамотного использования терминологии (10 баллов). Дополнительно начисляется до 10 баллов за ответы на уточняющие вопросы (прилагаются).	В соответствии с пп. 2.5, 2.6 Положения

6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ							
		1	2	3	4	5	6	7	8
ПК-13	Знает: - [И-1, ПУ] основные типы нестандартных данных (малые выборки, несбалансированные данные, большие объёмы); базовые алгоритмы (например, oversampling, undersampling, mini-batch learning) для типовых задач; ограничения стандартных методов при работе с такими данными	+	+	+	+	+		+	+
ПК-13	Умеет: - [И-1, ПУ] обосновывать выбор методов повышения эффективности и обобщаемости (регуляризация, уменьшение размерности модели, domain adaptation, использовать разностных методов типа сиамских сетей, few-shot learning, байесовские методы) [И-2, БУ] проводить сравнительный анализ моделей по эффективности, устойчивости и безопасности; учитывать переобучение, drift и computational cost				+			+	+
ПК-13	Имеет практический опыт: - [И-1, ПУ] адекватного выбора и адаптации алгоритмов (transfer learning, few-shot learning, federated learning) с учетом специфики нестандартных объемов данных и требований к задаче		+	+	+	+	+	+	+

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

7. Учебно-методическое и информационное обеспечение дисциплины

Печатная учебно-методическая документация

а) основная литература:

Не предусмотрена

б) дополнительная литература:

1. Демидов А. К. Функциональное и логическое программирование : Учеб. пособие / А. К. Демидов; Юж.-Урал. гос. ун-т, Каф. Приклад. математика; ЮУрГУ. - Челябинск : Издательство ЮУрГУ, 2000. - 58,[1] с.
2. Ясницкий Л. Н. Введение в искусственный интеллект : учеб. пособие / Л. Н. Ясницкий. - 2-е изд., испр.. - М. : Академия, 2008. - 174, [1] с.

в) отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:
Не предусмотрены

г) методические указания для студентов по освоению дисциплины:

1. Аугментация данных для решения проблемы обучения на малых данных в технической диагностике

из них: учебно-методическое обеспечение самостоятельной работы студента:

1. Аугментация данных для решения проблемы обучения на малых данных в технической диагностике

Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в электронной форме	Библиографическое описание
1	Основная литература	ЭБС издательства Лань	Введение в статистическое обучение с примерами на Python : руководство / Г. Джеймс, Д. Уиттон, Т. Хасты, Р. Тибширани ; перевод с английского А. Ю. Гинько. — Москва : ДМК Пресс, 2024. — 846 с. — ISBN 978-5-93700-217-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/464258 (дата обращения: 21.08.2025). — Режим доступа: для авториз. пользователей.
2	Основная литература	ЭБС издательства Лань	Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль ; перевод с английского А. А. Слинкина. — 2-е изд. — Москва : ДМК Пресс, 2018. — 652 с. — ISBN 978-5-97060-618-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/107901 (дата обращения: 21.08.2025). — Режим доступа: для авториз. пользователей.
3	Основная литература	ЭБС издательства Лань	Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Москва : ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/69955 (дата обращения: 21.08.2025). — Режим доступа: для авториз. пользователей.

Перечень используемого программного обеспечения:

1. -Python(бессрочно)

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Лабораторные занятия	333 (36)	Компьютеры, проектор, доска
Дифференцированный зачет	333 (36)	Компьютеры, проектор, доска
Лекции	336 (36)	Компьютер, проектор, доска