

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ

Ректор Белорусского
государственного университета

А.Д.Король

27 июня 2025 г.

Регистрационный № 3604/б.



ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ

Учебная программа учреждения образования по учебной дисциплине для
специальности:

6-05-0533-10 Информатика

Профилизация: Системный анализ

2025 г.

Учебная программа составлена на основе ОСВО 6-05-0533-10-2023, учебного плана БГУ: № 6-5.3-58/01 от 15.05.2023.

СОСТАВИТЕЛИ:

М.М.Васьковский, заведующий кафедрой фундаментальной математики и интеллектуальных систем факультета прикладной математики и информатики Белорусского государственного университета, доктор физико-математических наук, доцент;

А.О.Задорожнюк, доцент кафедры фундаментальной математики и интеллектуальных систем факультета прикладной математики и информатики Белорусского государственного университета, кандидат физико-математических наук;

Д.П.Сушко, старший преподаватель кафедры фундаментальной математики и интеллектуальных систем факультета прикладной математики и информатики Белорусского государственного университета

РЕЦЕНЗЕНТЫ:

К.В.Лыков, заведующий отделом функционального анализа и динамических систем ГНУ «Институт математики» НАН Беларуси, доктор физико-математических наук, профессор

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

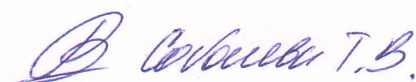
Кафедрой фундаментальной математики и интеллектуальных систем БГУ (протокол № 10 от 27.05.2025);

Научно-методическим советом БГУ (протокол № 11 от 26.06.2025)

Заведующий кафедрой



М.М.Васьковский



ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины «Основы машинного обучения»: формирование у студентов теоретических знаний и практических навыков в области машинного обучения, позволяющих решать реальные задачи анализа данных, строить и оценивать прогностические модели, а также грамотно интерпретировать полученные результаты.

Задачи учебной дисциплины:

1. Сформировать системное понимание основных понятий, задач и методов машинного обучения.
2. Обучить студентов полному циклу анализа данных, от исследования и предобработки до построения, валидации и интерпретации моделей.
3. Привить практические навыки работы с данными и инструментами машинного обучения (Python, Jupyter Notebook, библиотеки scikit-learn, pandas, numpy и др.).
4. Развить умение выбирать адекватные методы машинного обучения в зависимости от постановки задачи и характеристик данных.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием.

Учебная дисциплина относится к дисциплинам профилизации «Системный анализ» компонента учреждения образования.

Программа составлена с учетом межпредметных связей с учебными дисциплинами. Основой для изучения дисциплины является учебная дисциплина государственного компонента «Математический анализ» модуля «Математический анализ», «Методы оптимизации» модуля «Математические методы принятия решений»

Знания, полученные в учебной дисциплине, используются при изучении дисциплины «Нейронные сети» модуля «Системный анализ».

Требования к компетенциям

Освоение учебной дисциплины «Основы машинного обучения» должно обеспечить формирование следующих компетенций:

Универсальные компетенции:

Решать стандартные задачи профессиональной деятельности на основе применения информационно-коммуникационных технологий.

Специализированные компетенции:

Знать классические методы и алгоритмы машинного обучения, уметь адаптировать известные алгоритмы для решения прикладных задач регрессии, классификации, распознавания.

В результате освоения учебной дисциплины студент должен:

знать:

- Основные типы задач машинного обучения (классификация, регрессия, кластеризация) и подходы к их решению.
- Методы разведочного анализа данных (EDA) и визуализации.

- Принципы валидации моделей, методы борьбы с переобучением.
- Основные алгоритмы машинного обучения (линейные модели, решающие деревья, ансамбли, метрические методы, методы кластеризации и др.).
- Методы предобработки и конструирования признаков.
- Основные подходы к работе с временными рядами и построению рекомендательных систем.

уметь:

- Формулировать задачу в терминах машинного обучения.
- Проводить первичный анализ и очистку данных.
- Строить и интерпретировать визуализации.
- Обучать, валидировать и сравнивать модели машинного обучения.
- Анализировать ошибки модели и предлагать пути их уменьшения.

иметь навыки:

- Владения библиотеками Python для анализа данных и машинного обучения (pandas, numpy, matplotlib, seaborn, scikit-learn).
- Реализации полного цикла проекта машинного обучения от сбора данных до внедрения модели.
- Работы в среде Jupyter Notebook.

Структура учебной дисциплины

Дисциплина изучается в 5 семестре. В соответствии с учебным планом всего на изучение учебной дисциплины «Основы машинного обучения» отведено 108 часов, в том числе 68 аудиторных часа: лекции – 34 часа, лабораторные занятия – 34 часа. Из них:

Лекции – 34 часа, практические занятия – 30 часов, управляемая самостоятельная работа – 4 часа.

Трудоёмкость учебной дисциплины составляет 3 зачётных единицы.

Форма промежуточной аттестации – экзамен.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Введение в машинное обучение и подготовка данных

Тема 1.1. Введение в машинное обучение

Что такое машинное обучение и зачем оно используется. Типы обучения (с учителем, без учителя, с подкреплением). Постановки задач машинного обучения (классификация, регрессия, кластеризация, обработка картинок/текста, RL). Типы данных и признаки (структурированные/неструктурированные, числовые, категориальные, временные). Инструменты машинного обучения (SQL, Python, Jupyter Notebook).

Тема 1.2. Методология CRISP-DM и разведочный анализ данных (EDA)

Стандарт CRISP-DM (понимание бизнес-задачи, понимание данных, подготовка данных, моделирование, оценка, внедрение). Разведочный анализ данных (EDA): анализ распределений, поиск аномалий и выбросов, анализ взаимосвязей между переменными.

Тема 1.3. Визуализация данных

Продолжение EDA. Различные виды графиков (гистограмма, box-plot, scatter plot, heatmap, pairplot). Принципы построения наглядных и информативных графиков. Библиотеки визуализации (matplotlib, seaborn, plotly).

Тема 1.4. Валидация и тестирование моделей

Разбиение данных на обучающую, валидационную и тестовую выборки. Кросс-валидация (k-fold, Leave-One-Out). Time-aware кросс-валидация. Функция потерь и метрики качества, их различия и выбор. Понятие переобучения (overfitting) и недообучения (underfitting).

Тема 1.5. Подготовка данных

Масштабирование и нормализация признаков. Обработка категориальных признаков (One-Hot Encoding, Label Encoding). Методы обработки пропущенных значений. Методы обработки выбросов.

Раздел 2. Базовые алгоритмы машинного обучения

Тема 2.1. Линейная регрессия

Постановка задачи регрессии. Метрики регрессии (MSE, MAE, R^2). Способы решения: аналитический (нормальное уравнение) и градиентный спуск. Вероятностное обоснование и метод максимального правдоподобия (ММП). Проблемы и решения: мультиколлинеарность, гетероскедастичность, нелинейность, ненормальность ошибок. Регуляризация (Ridge, Lasso) и ее вероятностный смысл. Плюсы и минусы линейной регрессии.

Тема 2.2. Логистическая регрессия

Постановка задачи классификации. Метрики классификации (Accuracy, Precision, Recall, F1-score, ROC-AUC). Логистическая регрессия, градиентный спуск. Вероятностное обоснование и ММП. Регуляризация. Многоклассовая и многометковая классификация.

Тема 2.3. Оценивание вероятностей. Калибровка. Наивный Байес

Калибровка вероятностей в задаче классификации (Platt Scaling, Isotonic Regression). Наивный байесовский классификатор, его предположения и применение.

Тема 2.4. Решающие деревья

Строение решающего дерева. Критерии информативности (энтропия, Джини, дисперсия). Алгоритмы построения (ID3, CART). Работа с категориальными признаками. Преимущества и недостатки деревьев.

Тема 2.5. Ансамбли моделей

Базовые принципы: бэггинг, бустинг, стекинг. Случайный лес (Random Forest). Метод случайных подпространств (Random Subspaces). Градиентный бустинг (обзор XGBoost, LightGBM, CatBoost).

Раздел 3. Расширенные методы и специальные задачи

Тема 3.1. Продвинутая обработка признаков

Отбор признаков (фильтры, методы-обертки, встроенные методы). Конструирование признаков (полиномиальные признаки, биннинг, извлечение признаков из дат).

Тема 3.2. Метрические методы

Метод k-ближайших соседей (k-NN). Метод Парзенковского окна. Проклятие размерности.

Тема 3.3. Методы кластеризации

Постановка задачи кластеризации. Алгоритм K-Means. Gaussian Mixture Models (GMM) и EM-алгоритм. Иерархическая кластеризация. Алгоритм DBSCAN. Метрики качества кластеризации.

Тема 3.4. Понижение размерности

Метод главных компонент (PCA). t-SNE. UMAP.

Тема 3.5. Временные ряды

Основные компоненты временного ряда: тренд, сезонность, шум. Стационарность, автокорреляция. Модель ARIMA. Библиотеки Prophet и NeuralProphet. Применение решающих деревьев и ансамблей к временным рядам. Проблема "холодного старта".

Тема 3.6. Рекомендательные системы

Постановка задачи. Коллаборативная фильтрация (user-based, item-based). Матричные разложения (SVD). Проблема "холодного старта".

ТРЕБОВАНИЯ К КУРСОВОМУ ПРОЕКТУ (КУРСОВОЙ РАБОТЕ)

Цель курсового проекта (курсовой работы) - формирование у студентов целостного представления о полном цикле проекта в области машинного обучения, от формулировки проблемы и анализа предметной области до реализации, тестирования и оценки эффективности решения. Выполнение курсового проекта позволяет закрепить теоретические знания, полученные в ходе изучения дисциплины "Основы машинного обучения", и развить практические навыки решения реальных задач анализа данных.

Объем задания включает:

- Обзор предметной области и анализ существующих подходов к решению задачи
- Подбор и предобработку наборов данных для экспериментов
- Проектирование и реализацию ML-модели (или pipeline обработки данных)
- Проведение экспериментов и сравнительный анализ эффективности предложенного решения
- Оформление пояснительной записки объемом 25-35 страниц
- Подготовку презентации и защиту результатов работы

Темы курсовых проектов (работ):

1. Разделение монофонической речи на отдельные исходные дорожки из дорожки со смешанными голосами

Руководитель: Сушко Дмитрий Петрович

Основные направления исследования:

- Анализ методов сепарации аудиосигналов (спектральное маскирование, глубокое обучение)
- Сравнение архитектур нейросетей для задач разделения дорожек
- Разработка и обучение модели для разделения перекрывающихся голосовых дорожек
- Оценка качества разделения с использованием объективных метрик (SDR, SI-SNR, WER)
- Практическое применение в системах автоматической расшифровки речи

2. Большие языковые модели в задаче разметки данных для обучения ML-моделей

Основные направления исследования:

- Сравнительный анализ возможностей LLM (GPT-4, Claude, LLaMA, Qwen) для автоматической аннотации данных в задачах, принадлежащих различным доменным областям.

- Разработка конвейера полуавтоматической разметки текстовых/мультимодальных данных
- Сравнительный анализ качества автоматической и ручной разметки
- Оценка экономической эффективности использования LLM для разметки наборов данных
- Применение в реальных сценариях подготовки данных для ML-моделей

Количество часов на выполнение курсового проекта составляет 72 часа в соответствии с учебным планом учреждения образования, включая консультации с руководителем, самостоятельную работу над проектом и подготовку к защите.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Очная (дневная) форма получения высшего образования с применением дистанционных образовательных технологий
(ДОТ)

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСР	Форма контроля
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8	9
1	Введение в машинное обучение и подготовка данных	12			10		1	
1.1	Введение в машинное обучение	2			2			Отчеты по практическим заданиям. Устный опрос.
1.2	Методология CRISP-DM и разведочный анализ данных (EDA)	2			2			Отчеты по практическим заданиям. Собеседование.
1.3	Визуализация данных	2			2			Отчеты по практическим заданиям. Защита графиков.
1.4	Валидация и тестирование моделей	3			2			Отчеты по практическим заданиям. Контрольная работа №1.
1.5	Подготовка данных	3			2		1	Отчеты по практическим заданиям. Собеседование.
2	Базовые алгоритмы машинного обучения	14			12		2	
2.1	Линейная регрессия	4			4			Отчеты по практическим заданиям. Устный опрос.

2.2	Логистическая регрессия	3			2			Отчеты по практическим заданиям. Контрольная работа №2.
2.3	Оценивание вероятностей. Калибровка. Наивный Байес	2			2		1	Отчеты по практическим заданиям. Собеседование.
2.4	Решающие деревья	2			2			Отчеты по практическим заданиям.
2.5	Ансамбли моделей	3			2		1	Отчеты по практическим заданиям. Собеседование.
3	Расширенные методы и специальные задачи	8			8		1	
3.1	Продвинутая обработка признаков	1			1			Отчеты по практическим заданиям.
3.2	Метрические методы	1			1			Отчеты по практическим заданиям.
3.3	Методы кластеризации	2			2			Отчеты по практическим заданиям. Контрольная работа №3.
3.4	Понижение размерности	1			1			Отчеты по практическим заданиям.
3.5	Временные ряды	2			2		1	Отчеты по практическим заданиям. Собеседование.
3.6	Рекомендательные системы	1			1			Отчеты по практическим заданиям.

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Основная литература

1. Уатт, Дж. Машинное обучение: основы, алгоритмы и практика применения / Джереми Уатт, Реза Борхани, Аггелос Катсагелос ; [пер. с англ. Андрея Логунова]. - Санкт-Петербург : БХВ-Петербург, 2022. - 612 с.
2. Баланов, А. Н. Машинное обучение и искусственный интеллект : учебное пособие для вузов / А. Н. Баланов. - Санкт-Петербург ; Москва ; Краснодар : Лань, 2024. - 168 с. - URL: <https://e.lanbook.com/book/414920>.
3. Аминян, А. System Design. Машинное обучение. Подготовка к сложному интервью / Али Аминян, Алекс Сюй ; [пер. с англ. Е. Матвеев]. - Санкт-Петербург ; Москва ; Минск : Питер, 2024. - 317 с.
4. Хливненко, Л. В. Практика нейросетевого моделирования : учебное пособие [для вузов] / Л. В. Хливненко, Ф. А. Пятакович. - Изд. 3-е, стер. - Санкт-Петербург ; Москва ; Краснодар : Лань, 2023. - 196 с. - URL: <https://e.lanbook.com/book/310190>.
5. Николенко, С. Машинное обучение : основы / С. Николенко. - Санкт-Петербург ; Москва ; Минск : Питер, 2025. - 608 с.

Дополнительная литература

1. Aupélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. – O'Reilly Media, 2022.
2. Кристофер М. Бишоп. Pattern Recognition and Machine Learning. – Springer, 2006.
3. Иэн Гудфеллоу, Йошуа Бенджио, Аарон Курвилль. Глубокое обучение. – М.: ДМК Пресс, 2022.
4. Вандер Плас Дж. Python для сложных задач: наука о данных и машинное обучение. – СПб.: Питер, 2020.
5. Орельен Жерон. Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow. – СПб.: Питер, 2022.
6. Кевин Мерфи. Вероятностное машинное обучение: введение. – М.: ДИАЛЕКТИКА, 2023.
7. Тревор Хастис, Роберт Тибширани, Джером Фридман. The Elements of Statistical Learning. – Springer, 2017.

Перечень рекомендуемых средств диагностики и методика формирования итоговой отметки

Для диагностики компетенций могут использоваться следующие средства текущей аттестации: устные опросы, собеседования, защита практических заданий и отчетов по ним, контрольные работы.

Формой промежуточной аттестации по дисциплине «Основы машинного обучения» учебным планом предусмотрен экзамен.

Для формирования итоговой отметки по учебной дисциплине используется модульно-рейтинговая система оценки знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая система предусматривает использование весовых коэффициентов для текущей и промежуточной аттестации студентов по учебной дисциплине.

Формирование итоговой отметки в ходе проведения контрольных мероприятий текущей аттестации (примерные весовые коэффициенты, определяющие вклад текущей аттестации в отметку при прохождении промежуточной аттестации):

- лабораторные занятия – 100 %;

Итоговая отметка по дисциплине рассчитывается на основе итоговой отметки текущей аттестации (модульно-рейтинговой системы оценки знаний) 50 % и экзаменационной отметки 50 %.

Примерный перечень лабораторных занятий

1. Настройка окружения. Знакомство с Jupyter Notebook, pandas, numpy. Первичный анализ датасета.
2. Визуализация данных с использованием matplotlib и seaborn.
3. Разбиение выборок и кросс-валидация. Расчет метрик качества.
4. Реализация линейной регрессии и градиентного спуска.
5. Решение проблем линейной регрессии: регуляризация, анализ остатков.
6. Логистическая регрессия для бинарной и многоклассовой классификации.
7. Наивный байесовский классификатор и калибровка вероятностей.
8. Построение и визуализация решающих деревьев.
9. Ансамбли: Random Forest и Gradient Boosting.
10. Отбор и конструирование признаков.
11. Кластеризация методами K-Means и DBSCAN.
12. Визуализация данных методами PCA и t-SNE.
13. Прогнозирование временных рядов.
14. Построение простой рекомендательной системы.

Примерный перечень заданий для управляемой самостоятельной работы

Тема 1.5. Подготовка данных (1 ч.).

Анализ и предобработка нового набора данных с пропусками, выбросами и категориальными признаками.

Форма контроля - отчёты по практическим заданиям, устный опрос.

Тема 2.3. Оценивание вероятностей (1 ч.).

Исследование калибровки моделей логистической регрессии и наивного Байеса на различных наборах данных.

Форма контроля - отчёты по практическим заданиям, устный опрос.

Тема 2.5. Ансамбли моделей (1 ч.).

Сравнение эффективности Random Forest и Gradient Boosting на соревновательном датасете (например, с Kaggle).

Форма контроля - отчёты по практическим заданиям, устный опрос.

Тема 3.5. Временные ряды (1 ч.).

Прогнозирование временного ряда с использованием ARIMA и Prophet, сравнение качества.

Форма контроля - отчёты по практическим заданиям, устный опрос.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используются следующие методы:

- ***Метод учебной дискуссии***, который предполагает участие студентов в целенаправленном обмене мнениями, идеями для предъявления и/или согласования существующих позиций по определенной проблеме. Использование метода обеспечивает появление нового уровня понимания изучаемой темы, применение знаний (теорий, концепций) при решении проблем, определение способов их решения. Примеры: дискуссии о выборе метрики для конкретной бизнес-задачи, обсуждение этических аспектов применения ML, сравнение эффективности разных алгоритмов.

- ***Метод проектного обучения (Project-Based Learning)***, который представляет собой форму организации учебно-познавательной деятельности обучающихся, предполагающую выполнение сквозного проекта на протяжении курса. Студенты работают над решением реальной или максимально приближенной к реальности задачи (например, "Прогнозирование оттока клиентов", "Классификация обращений в поддержку"), проходя все этапы CRISP-DM: от формулировки задачи и сбора данных до построения, оценки и интерпретации финальной модели.

- ***Метод группового обучения (Collaborative Learning)***, при котором студенты объединяются в малые группы для работы над общими и специфическими учебными заданиями, такими как "аналитические спарринги" по разбору EDA, совместное участие в соревнованиях по машинному обучению (например, на платформе Kaggle InClass) или peer-review кода и выводов друг друга.

- ***Проблемно-ориентированное обучение (Case Study)***, где студенты анализируют конкретные кейсы успешного или неудачного применения машинного обучения в индустрии, что позволяет сформировать понимание практических ограничений и требований к ML-специалистам.

Используются **методы и приемы развития критического мышления**, которые представляют собой систему, формирующую навыки работы с

информацией; понимании информации как отправного, а не конечного пункта критического мышления. Студентов учат не просто применять алгоритмы, но и критически оценивать качество данных, обоснованность выбранного подхода, интерпретировать результаты и анализировать ошибки модели.

При организации образовательного процесса используется **практико-ориентированный и индустриально-релевантный подход**, который предполагает освоение содержания дисциплины через выполнение практических заданий, имеющих элементы учебно-исследовательской деятельности и максимально приближенных к задачам, решаемым в IT-компаниях и аналитических отделах.

Методические рекомендации по организации самостоятельной работы

Для организации самостоятельной работы студентов по учебной дисциплине следует использовать современные информационные ресурсы: разместить на образовательном портале комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, методические указания к практическим занятиям, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательного стандарта высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к экзамену, задания, тесты, вопросы для самоконтроля, список рекомендуемой литературы, информационных ресурсов и др.).

В качестве технических средств для организации работы в рамках учебной дисциплины рекомендуется использовать **Образовательный портал БГУ (<https://edufpmi.bsu.by>)** – инструмент с эффективной функциональностью контроля, тренинга и самостоятельной работы. Для выполнения практических заданий и проектов используются облачные среды (Google Colab, JupyterHub), что обеспечивает единую среду выполнения для всех студентов и исключает проблемы с настройкой локального окружения.

Рекомендуется организовать самостоятельную работу вокруг следующих видов деятельности:

1. **Работа с литературой и видеоматериалами:** изучение документации к библиотекам (scikit-learn, pandas), просмотр записей выступлений с конференций (Data Science Meetups, Conférence on Neural Information Processing Systems).

2. **Выполнение интерактивных тренажеров:** использование онлайн-курсов и платформ (Kaggle Learn, Coursera) для отработки конкретных тем.

3. **Решение задач и соревнований:** регулярное решение задач на платформах (Kaggle, AICrowd) для формирования портфолио.

4. **Самостоятельное углубленное изучение тем:** исследование и подготовка краткого обзора по одной из продвинутых тем (например,

"Применение UMAP для визуализации высокоразмерных данных", "Сравнение библиотек градиентного бустинга").

Примерный перечень вопросов к экзамену

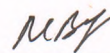
1. Машинное обучение: основные понятия, цели и области применения.
2. Типы обучения: с учителем, без учителя, с подкреплением. Примеры задач для каждого типа.
3. Основные постановки задач машинного обучения: классификация, регрессия, кластеризация.
4. Типы данных: структурированные и неструктурированные данные. Числовые, категориальные, временные признаки.
5. Жизненный цикл проекта машинного обучения (CRISP-DM). Основные этапы.
6. Разведочный анализ данных (EDA): цели, основные методы и визуализации.
7. Валидация моделей: разбиение на обучающую и тестовую выборку. Принцип и причины.
8. Кросс-валидация: сущность, виды (k-fold, LOOCV), преимущества и недостатки.
9. Переобучение (overfitting): причины возникновения и методы борьбы.
10. Функция потерь и метрики качества. Различия и принципы выбора.
11. Метрики качества в задачах регрессии (MSE, MAE, R^2).
12. Метрики качества в задачах бинарной классификации (Accuracy, Precision, Recall, F1-Score, ROC-AUC).
13. Линейная регрессия: постановка задачи, математическая формулировка.
14. Градиентный спуск: алгоритм, виды, гиперпараметры.
15. Вероятностная интерпретация линейной регрессии и метод максимального правдоподобия.
16. Проблема мультиколлинеарности: диагностика и методы решения.
17. Регуляризация в линейной регрессии: Ridge и Lasso. Различия и вероятностный смысл.
18. Логистическая регрессия: постановка задачи, функция потерь, алгоритм обучения.
19. Вероятностная интерпретация и метод максимального правдоподобия для логистической регрессии.
20. Многоклассовая классификация в логистической регрессии (One-vs-Rest, Softmax).
21. Наивный байесовский классификатор: основные предположения, алгоритм, преимущества и недостатки.
22. Калибровка вероятностей: цели и основные методы (Platt Scaling, Isotonic Regression).
23. Решающие деревья: структура, алгоритм построения, критерии информативности (энтропия, Джини).

24. Достоинства и недостатки решающих деревьев.
25. Бэггинг: основная идея, алгоритм, преимущества.
26. Случайный лес (Random Forest): принцип работы, основные гиперпараметры.
27. Бустинг: основная идея, отличия от бэггинга. Обзор градиентного бустинга.
28. Метод k-ближайших соседей (k-NN): алгоритм, выбор метрики и гиперпараметров.
29. "Проклятие размерности" и его влияние на метрические алгоритмы.
30. Кластеризация: постановка задачи. Алгоритм K-Means, его достоинства и недостатки.
31. Алгоритм кластеризации DBSCAN: принцип работы, преимущества перед K-Means.
32. Метод главных компонент (PCA): цель применения, геометрическая и алгебраическая интерпретация.
33. Временные ряды: основные компоненты (тренд, сезонность, шум). Понятие стационарности.
34. Модель ARIMA: основные компоненты и идея применения.
35. Рекомендательные системы: постановка задачи. Коллаборативная фильтрация (user-based, item-based).
36. Проблема "холодного старта" в рекомендательных системах.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Учебная дисциплина не требует согласования			

Заведующий кафедрой фундаментальной математики и интеллектуальных систем
доктор физико-математических наук
профессор



М.М.Васьковский

27.05.2025

ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ УО

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 202_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ
Декан факультета
