

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Б1.В.11 Теория вероятностей и математическая статистика

Объем трудоемкости: 2 з.е.

Цели и задачи изучения дисциплины (модуля)

Цель освоения дисциплины

Формирование у студентов системного понимания фундаментальных понятий и методов теории вероятностей и математической статистики, их роли в анализе данных и искусственном интеллекте, а также развитие навыков применения вероятностно-статистических подходов для решения практических задач в области ИИ.

Задачи дисциплины

- Изучение основ теории вероятностей, включая случайные события, случайные величины и их распределения.
- Освоение ключевых теорем (включая теорему Байеса) и их приложений в анализе данных и машинном обучении.
- Приобретение навыков работы с основными вероятностными распределениями и их параметрами.
- Изучение методов математической статистики: описательная статистика, оценка параметров, проверка статистических гипотез.
- Развитие умения применять статистические методы для анализа реальных данных и валидации моделей машинного обучения.
- Формирование понимания различий между байесовским и частотным подходами в статистике.
- Приобретение практических навыков использования Python (с библиотеками NumPy, SciPy, Pandas, Matplotlib, StatsModels) для вероятностно-статистического анализа данных.

Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Теория вероятностей и математическая статистика» относится к части учебного плана Блока 1. Дисциплины (модули), формируемой участниками образовательных отношений.

Предшествующие дисциплины (пререквизиты) Математический анализ, Алгебра и аналитическая геометрия, Дискретная математика.

Последующие дисциплины, для которых данная дисциплина является предшествующей Машинное обучение, Глубокое обучение, Нейросетевые технологии, Технологии обработки больших данных, Технологии обработки языка, Генеративный искусственный интеллект.

Постреквизиты: Основы прикладной статистики (MF-1.1 (C)), Подготовка данных машинного обучения (MF-4.1 (C)), Data Centric Machine Learning (MF-4.1 (C)).

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Роль 1: Data Engineer (Инженер по данным)

Задачи:

1. Проектирование и построение ETL-процессов
2. Создание и оптимизация хранилищ данных
3. Обеспечение качества и доступности данных
4. Настройка инфраструктуры для обработки больших данных

5. Интеграция разрозненных источников данных
6. Работа с данными в области природопользования, медицины, связи и телекоммуникаций

Роль 2: ML Engineer (Инженер МО)

Задачи:

1. Реализация ML-моделей в продуктивных системах
2. Оптимизация производительности и масштабирование моделей
3. Разработка ML-пайплайнов и автоматизация процессов
4. Мониторинг качества моделей в продуктиве
5. Интеграция ML-решений с бизнес-приложениями

Роль 3: MLOps (Специалист по эксплуатации ИИ)

Задачи:

1. Автоматизация процессов обучения и развертывания моделей
2. Мониторинг производительности ML-систем
3. Управление версиями моделей и данных
4. Обеспечение CI/CD для ML-проектов
5. Оптимизация вычислительных ресурсов

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

Код и наименование индикатора	Результаты обучения по дисциплине
MF-1 Способен применять современную теоретическую математику для разработки новых алгоритмов и формулирования перспективных задач ИИ	
MF-1.2 Применяет аппарат теории вероятностей, матстатистики и теории информации для формулирования и анализа задач искусственного интеллекта.	Уровень Б (базовый) Знать: основные понятия теории вероятностей (случайные события, вероятность, случайные величины, распределения), основы математической статистики (выборка, оценки параметров), базовые понятия теории информации (энтропия, условная энтропия). Уметь: оценивать эмпирические распределения, вычислять выборочные характеристики, строить гистограммы, рассчитывать простые статистические зависимости (корреляцию). Владеть: методами точечного оценивания (метод моментов, ММП), базовыми навыками работы с распределениями (нормальное, биномиальное, пуассоновское), интерпретацией доверительных интервалов. Уровень С (средний) Знать: продвинутые разделы теории вероятностей (условные распределения, многомерные случайные величины), критерии независимости, информационные меры (взаимная информация, расхождение Кульбака–Лейблера). Уметь: применять методы и модели ИИ (логистическая регрессия, наивный байесовский классификатор, деревья решений), анализировать потребности задачи и адаптировать модели (выбор функции потерь, настройка гиперпараметров). Владеть: методами оценки качества моделей (кросс-валидация, AUC-ROC, матрица ошибок), способами повышения точности и эффективности (регуляризация, feature engineering).

Код и наименование индикатора	Результаты обучения по дисциплине
MF-1.3 Применяет аппарат теории вероятностей для исследования методов и моделей машинного обучения	Уровень Б (базовый) Знать: основные понятия проверки статистических гипотез (нулевая и альтернативная гипотезы, уровень значимости, p-value), популярные критерии (t-критерий, критерий χ^2, критерий Колмогорова–Смирнова). Уметь: применять t-критерий для сравнения средних, χ^2 для проверки независимости признаков, корректно интерпретировать p-value и принимать/отвергать гипотезы. Владеть: терминологией проверки гипотез (ошибки I и II рода, мощность критерия), использованием стандартных статистических функций в Python (scipy.stats) или аналогичных инструментах. Уровень С (средний) Знать: различия в постановке задач для параметрических и непараметрических критериев, теоретические основы специализированных критериев (критерий Манна–Уитни, критерий Уилкоксона, ANOVA, критерий Шапиро–Уилка). Уметь: формулировать задачу проверки гипотезы в терминах ML (например, сравнение качества двух моделей), выбирать критерий в зависимости от распределения данных и типа задачи. Владеть: применением специализированных критериев для анализа ML-экспериментов (сравнение алгоритмов, оценка значимости признаков), интерпретацией результатов в контексте обучения моделей.
MF-2 Способен применять байесовский подход для построения вероятностных моделей, анализа неопределенности и создания адаптивных систем ИИ	
MF-2.1 Использует теорему Байеса и её следствия, понимает отличия байесовской статистики от частотного подхода.	Уровень Б (базовый) Знать: формулировку теоремы Байеса, понятия априорной, правдоподобия и апостериорной вероятности, ключевые различия между байесовским и частотным подходами (параметр как случайная величина vs. фиксированная константа). Уметь: применять теорему Байеса для решения простых задач (например, диагностические тесты, классификация с наивным байесовским классификатором), объяснять различия на примерах. Владеть: базовыми принципами байесовского вывода (вычисление апостериорного распределения), пониманием роли априорного распределения.
MF-2.3 Анализирует и применяет байесовский и частотный подходы, свободно использует теорему Байеса и её следствия для решения сложных статистических задач	Уровень Б (базовый) Знать: теорему Байеса и методы частотного подхода (ММП, доверительные интервалы), основные различия (интерпретация вероятности, обработка неопределенности). Уметь: применять оба подхода к стандартным статистическим задачам (оценка параметров, проверка гипотез), анализировать, в каких случаях байесовский подход предпочтительнее. Владеть: расчетом апостериорных распределений для простых моделей (нормальная, биномиальная), использованием частотных методов (p-value, доверительные интервалы) на реальных данных.

Код и наименование индикатора	Результаты обучения по дисциплине
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	Уровень Б (базовый) Знать: отличия статистического обучения от нестатистического (вероятностная постановка vs. детерминированная), классификацию методов статистического МО (регрессия, классификация, кластеризация с вероятностной точки зрения). Уметь: понимать, когда задача требует статистического подхода (наличие шума, выборочные данные), различать параметрические и непараметрические методы. Владеть: описанием базовых моделей (линейная регрессия с нормальными ошибками, логистическая регрессия) как примеров статистического обучения. Уровень С (средний) Знать: особенности данных (гетероскедастичность, мультиколлинеарность, несбалансированные классы) и их влияние на выбор статистических методов МО. Уметь: выбирать подходящий статистический метод МО в зависимости от типа данных и задачи (например, квантильная регрессия для несимметричных распределений, байесовские методы для малых выборок), объяснять различия между подходами (например, bagging vs. байесовское усреднение моделей). Владеть: методикой проведения вычислительных экспериментов с использованием статистических тестов для сравнения моделей (например, тест Макнемара, кросс-валидация с t-тестом), интерпретацией результатов в терминах статистической значимости.

Структура дисциплины

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 3 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	СРС
1	2	3	4	5	6	7
1.	Тема 1. Классическое определение вероятности	6	2		2	2
2.	Тема 2. Аксиоматическое построение теории вероятностей. Формула Байеса	6	2		2	2
3.	Тема 3. Случайные величины	6	2		2	2
4.	Тема 4. Распределение дискретных случайных величин	6	2		2	2
5.	Тема 5. Распределение непрерывных случайных величин	6	2		2	2

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
6.	Тема 6. Основные непрерывные распределения	6	2		2	2
7.	Тема 7. Многомерные случайные величины. Функция от случайной величины. Формула свертки. Распределения Пирсона, Стьюдента, Фишера	6	2		2	2
8.	Тема 8. Математическое ожидание. Дисперсия случайной величины	6	2		2	2
9.	Тема 9. Характеристики взаимосвязи случайных величин	6	2		2	2
10.	Тема 10. Закон больших чисел. Предельные теоремы теории вероятностей	6	2		2	2
11.	Тема 11. Основные понятия математической статистики	6	2		2	2
12.	Тема 12. Выборочные средние и дисперсии	6	2		2	2
13.	Тема 13. Оценка параметров генеральной совокупности	6	2		2	2
14.	Тема 14. Точечные оценки параметров	6	2		2	2
15.	Тема 15. Гипотезы о равенстве средних, дисперсий. Гипотеза о соответствии законов распределения	6	2		2	2
16.	Тема 16. Элементы регрессионного анализа. Множественный корреляционный анализ	8	2		2	4
17.	Тема 17. Применение изученных методов в задачах машинного обучения: оценка качества моделей, данных	6	2		2	2
ИТОГО по разделам дисциплины		104	34		34	36
Контроль самостоятельной работы (КСР)		4				
Промежуточная аттестация (ИКР)		0,3				
Подготовка к текущему контролю		35,7				
Общая трудоемкость по дисциплине		144				

Примечание: Л – лекции, ПЗ – практические занятия/семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

Форма проведения аттестации по дисциплине: экзамен

Автор: Калайдина Г.В. – к. ф.-м. н., доцент КАДИИ