

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«07» марта 2024 г.
Протокол № 1

**Рабочая программа дисциплины (модуля)
«Байесовские статистические модели»**

Направление подготовки: 02.04.01 Математика и компьютерные науки

Направленность (профиль) подготовки: Продуктовая аналитика

Квалификация (степень) выпускника: магистр

Форма обучения: очная

Срок освоения программы: 2 года

Год набора: 2024

**Москва
2024**

Содержание

1. Краткая характеристика дисциплины (модуля)	3
2. Перечень планируемых результатов обучения.....	4
3. Тематический план.....	6
4. Содержание дисциплины (модуля).....	6
5. Учебно-методическое обеспечение	7
6. Материально-техническое обеспечение	7
7. Методические и оценочные материалы	9

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Байесовские статистические модели» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по специальности 02.04.01 Математика и компьютерные науки, профиль Продуктовая аналитика, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 810 от 23.08.2017 года.

Изучение дисциплины (модуля) «Байесовские статистические модели» позволяет студентам освоить вероятностный подход к анализу данных, который эффективно учитывает неопределенность и интегрирует априорные знания с новыми наблюдениями, обеспечивая более точные и обоснованные выводы по сравнению с традиционными методами. Это особенно ценно в бизнес-контексте, где такие навыки применяются для оптимизации A/B-тестирования, прогнозирования временных рядов и разработки предиктивных моделей, способствуя принятию стратегических решений и повышению качества аналитических проектов.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 02.04.01 Математика и компьютерные науки, профиль Продуктовая аналитика и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений.

Дисциплина (модуль) изучается на 2 курсе в 3 семестре, доступна для прохождения при условии успешного завершения дисциплины (модуля) «Machine Learning (Машинное обучение)» и общеуниверситетского факультатива «Математика в DS».

Цель изучения дисциплины (модуля): формирование знаний и навыков в области байесовских статистических моделей для применения вероятностного подхода к анализу данных и решению бизнес-задач с учетом неопределенности.

Задачи изучения дисциплины (модуля):

— формирование знаний по темам: основы байесовской статистики и её отличие от частотного подхода; как применять байесовский подход к бизнес-задачам; механизмы ABCDE-тестирования и интерпретации результатов; принципы байесовской линейной регрессии; методы анализа временных рядов с помощью Гауссовских процессов; подходы к построению вероятностных моделей и их применению в реальных проектах;

— освоение умений: формулировать гипотезы и пересчитывать априорные вероятности с учётом новых данных; настраивать и интерпретировать ABCDE-тесты с байесовским выводом; строить и анализировать байесовские регрессионные модели; применять Гауссовские процессы к анализу временных данных; оценивать параметры распределений с помощью приближённых методов; разрабатывать и защищать прикладной аналитический проект;

— формирование навыков применения байесовского подхода к анализу данных в бизнес-контексте; навыков выполнения A/B/n-тестов с использованием байесовской логики; навыков построения предиктивных моделей на основе линейной регрессии; навыков работы с временными рядами и использования Гауссовских процессов для прогноза; навыков представления результатов анализа и составления обоснованных выводов; навыков уверенного использования инструментов байесовского анализа в прикладных проектах.

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-6.	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1.	Знает основные методы самооценки и анализа своей деятельности, а также принципы управления временем и целеполагания
		УК-6.2.	Умеет ставить реалистичные и достижимые цели, определять приоритеты в своей деятельности, а также разрабатывать и внедрять планы по совершенствованию своих навыков и компетенций на основе полученной самооценки
		УК-6.3.	Имеет практический опыт применения методов самооценки в своей профессиональной деятельности, включая участие в тренингах, семинарах и проектах, направленных на развитие личной эффективности и профессионального роста
ОПК-2.	Способен создавать и исследовать новые математические модели в естественных науках, совершенствовать и разрабатывать концепции, теории и методы	ОПК-2.1.	Знает основные математические модели и методы, используемые в естественных науках, включая статистическое моделирование, дифференциальные уравнения и численные методы, а также современные подходы к исследованию и анализу данных
		ОПК-2.2.	Умеет разрабатывать и адаптировать математические модели для решения конкретных проблем в естественных науках, проводить их анализ и верификацию, а также интерпретировать полученные результаты в контексте научных исследований
		ОПК-2.3.	Имеет практический опыт создания и исследования математических моделей в рамках научных проектов или исследований, включая участие в публикациях, конференциях или коллаборациях, где были разработаны и апробированы новые концепции и методы

ПК-3.	Способен решать задачи профессиональной деятельности в области продуктовой аналитики, формулировать результаты анализа и выявлять последствия полученных данных для принятия обоснованных решений и оптимизации продуктов	ПК-3.1.	Знает методы и инструменты продуктовой аналитики
		ПК-3.2.	Умеет применять аналитические инструменты и программное обеспечение для обработки и визуализации данных, а также формулировать выводы на основе проведенного анализа
		ПК-3.3.	Имеет опыт работы над реальными проектами в области продуктовой аналитики, включая анализ пользовательского поведения и оптимизацию продуктов на основе полученных данных
ПК-4.	Способен публично представлять собственные и известные научные результаты	ПК-4.1.	Знает основные принципы эффективного публичного выступления, методы визуализации данных и основные требования к научным презентациям, включая структуру и содержание
		ПК-4.2.	Умеет четко и логично формулировать свои научные результаты, адаптируя их для различных аудиторий, а также использовать визуальные средства для улучшения восприятия информации
		ПК-4.3.	Имеет практический опыт участия в научных конференциях, семинарах или других мероприятиях, где успешно представлял свои и известные научные результаты, получая обратную связь и взаимодействуя с аудиторией

3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы					ТКУ (текущий контроль успеваемости)
		Очная форма					
		Аудиторная работа			Контроль	Самост оятель ная работа	
		Лекции	Семинары (практичес кие занятия)	Консул ьтации			
1	Байесовская статистика в бизнесе и принципы моделирования	2	1	3		19	Домашнее задание
2	Байесовские ABCDE-тесты и восстановление параметров	2	2	3		19	Домашнее задание
3	Байесовская линейная регрессия	2	2	3		19	Домашнее задание
4	Временные ряды и Гауссовские процессы	3	2	3		25	Домашнее задание
	<i>Зачет</i>				4		Проект
	Итого:	9	7	12	4	82	
	Объем дисциплины (модуля) (в ак. ч.)	114					
	Объем дисциплины (модуля) (в зач. ед.)	3					

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Байесовская статистика в бизнесе и принципы моделирования	Основы байесовской статистики Понятие априорного и апостериорного распределений, а также правдоподобия Принципы байесовского моделирования
2	Байесовские ABCDE-тесты и восстановление параметров	Мотивация байесовского А/Б тестирования Отличия от стандартного подхода Основы визуализации результатов байесовского моделирования
3	Байесовская линейная регрессия	Виды линейных регрессий Классический байесовский подход к линейной регрессии Современный подход (R2D2M2)
4	Временные ряды и Гауссовские процессы	Основы и виды гауссовских процессов Прикладное значение априорных распределений Численная аппроксимация гауссовских процессов Построение смесей гауссовских процессов

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Курт, У. Байесовская статистика: Star Wars, LEGO, резиновые уточки и многое другое : практическое руководство / У. Курт. - Санкт-Петербург : Питер, 2021. - 304 с. - (Библиотека программиста). - ISBN 978-5-4461-1655-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1817495>.

2. Вэй Цзи Ма, Байесовские модели восприятия и действия : современный взгляд на принципы работы мозга : практическое руководство / Вэй Цзи Ма, К. Кёрдинг, Д. Голдрайх ; пер. с англ. В. С. Яценкова. – Москва : ДМК Пресс, 2023. - 460 с. – ISBN 978-5-93700-229-7. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2204250>.

Дополнительная литература:

1. Коэн, М. Прикладная линейная алгебра для исследователей данных. От ключевых концепций до приложений с использованием Python : практическое руководство / М. Коэн ; пер. с англ. А. В. Логунова. – Москва : ДМК Пресс, 2023. - 330 с. – ISBN 978-6-01798-945-3. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2205123>.

2. Груздев, А. В. Прогнозирование временных рядов с помощью Facebook Prophet, ETNA, sktime и LinkedIn Greykite : строим, настраиваем, улучшаем модели прогнозирования временных рядов с помощью специальных библиотек : практическое руководство / А. В. Груздев. – Москва : ДМК Пресс, 2023. - 782 с. – ISBN 978-5-93700-212-9. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2204240>.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;
- специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование»	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		

КонсультантПлюс: справочно-правовая система	отечественное	лицензионное
Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Байесовские статистические модели» в рамках текущего контроля успеваемости используются такие виды учебной работы, как лекции, семинары, консультации, домашние задания, проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Лекция – систематическое, последовательное, монологическое изложение преподавателем учебного материала, как правило, теоретического характера.

В процессе лекций рекомендуется вести конспект лекций: кратко и схематично фиксировать основные идеи, выводы и обобщения лекции; выделять важные мысли, ключевые слова и термины. Необходимо отметить вопросы или материалы, которые вызывают затруднения, и попытаться найти ответы в рекомендованной литературе. Если разобраться в материале не удастся, следует сформулировать вопрос и задать его преподавателю на консультации или во время семинарского (практического) занятия.

Семинар — это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Консультации – структурированные встречи, на которых преподаватели предоставляют индивидуальную или групповую помощь в освоении учебного материала, обсуждении вопросов и решении проблем, возникающих в процессе обучения.

Консультации могут включать разъяснение сложных тем, подготовку к экзаменам и

помощь в выполнении проектных работ, что способствует более глубокому пониманию предмета и улучшению академической успеваемости.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

Проект – исследовательская работа по дисциплине (модулю) и презентация результатов.

Для успешной подготовки к проекту рекомендуется: четко определить цели и задачи проекта; составить план работы, разбив проект на этапы с указанием сроков выполнения каждого из них; использовать разнообразные источники информации и инструменты для исследования темы; регулярно проверять прогресс и вносить коррективы в план, если это необходимо.

Бонусные баллы — это оценки, которые студенты могут получить за выполнение дополнительных заданий.

Формат бонусных баллов позволяет студентам улучшить общую оценку по дисциплине (модулю) и стимулирует углубленное изучение материала.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Байесовские статистические модели»

Оценивание уровня учебных достижений обучающихся по дисциплине (модулю) осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме *зачета*, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом:

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
10	Отлично	Зачтено	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину (модуль). Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать,
9	Отлично	Зачтено	
8	Отлично	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
7	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	Зачтено	
5	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	
3	Не сдан	Не зачтено	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			преподаватель задает дополнительные наводящие вопросы.

Дисциплина (модуль) «Байесовские статистические модели» оценивается следующим образом:

Активность	Вес	Описание
Домашние задания	60%	Набор задач по темам недели
Зачет	40%	Защита проекта

В рамках изучения дисциплины (модуля) возможно получение бонусных баллов.

Формула расчёта итоговой оценки по дисциплине (модулю) «Байесовские статистические модели»: « $0,6 \times \text{среднее за домашние задания} + 0,4 \times \text{зачет}$ ».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание: Байесовская статистика в бизнесе и принципы моделирования

1. Используя формулу Байеса, рассчитайте апостериорную вероятность успеха нового продукта на основе априорных данных о рынке (например, 30% шанс успеха) и новых наблюдений (5 успешных продаж из 10 тестов), представив результаты в виде графика.

2. Выберите бизнес-сценарий (например, прогноз продаж), задайте априорное нормальное распределение, обновите его с помощью данных о прибыли за месяц и визуализируйте изменения с помощью Python (matplotlib или seaborn), объяснив влияние новых данных.

3. Для гипотезы о вероятности оттока клиентов (априорная бета-распределение) постройте функцию правдоподобия на основе выборки из 50 клиентов (20 оттоков) и рассчитайте обновленную апостериорную вероятность, используя библиотеку PyMC3.

4. Опишите процесс построения простой байесовской модели для оценки эффективности маркетинговой кампании, включая выбор априорных распределений, интеграцию данных и интерпретацию результатов; реализуйте модель в Python и сравните с частотным подходом.

5. Создайте интерактивный дашборд (с помощью Plotly) для бизнес-метрики (например, конверсия сайта), отображающий априорное распределение, правдоподобие и апостериорное распределение на основе симулированных данных, и объясните, как это помогает в принятии решений.

Домашнее задание: Байесовские ABCDE-тесты и восстановление параметров

1. Проанализируйте результаты A/B-теста для двух версий landing page (конверсии: 15% vs 18% на выборке 1000 пользователей), рассчитайте байесовскую вероятность превосходства одной версии и сравните с p-значением частотного теста, используя PyMC3.

2. Для теста с несколькими вариантами (A, B, C, D, E) восстановите параметры бета-распределения на основе наблюдаемых конверсий (например, A: 120/1000, B: 150/1000), постройте апостериорные распределения и определите наиболее вероятный победитель.

3. Проведите сравнение байесовского и частотного A/B-тестирования на примере данных о CTR рекламы (байесовский вывод с априорными знаниями о рынке), визуализируйте различия в интерпретации и объясните преимущества байесовского метода для раннего останова теста.

4. Используя данные о пользовательском поведении в приложении, постройте байесовскую модель для оценки среднего времени сессии, визуализируйте апостериорные интервалы и плотности распределений с помощью ggplot2 в R, и интерпретируйте результаты для бизнес-рекомендаций.

5. Разработайте сценарий ABCDE-теста для оптимизации цены продукта, задайте априорные распределения для каждого варианта, обновите их на основе симулированных продажных данных и создайте отчет с графиками, показывающими вероятность выбора лучшей цены.

Домашнее задание: Байесовская линейная регрессия

1. Сравните классическую линейную регрессию с байесовской на примере данных о зависимости продаж от цены и рекламы (выборка 200 наблюдений), рассчитайте коэффициенты и интервалы неопределенности, используя PyMC3, и обсудите различия.

2. Постройте байесовскую линейную регрессионную модель для предсказания прибыли компании на основе факторов (инвестиции, рынок), задайте априорные нормальные распределения для параметров, оцените апостериорные распределения и визуализируйте предсказательные интервалы.

3. Реализуйте модель R2D2M2 для байесовской регрессии на данных о временных рядах продаж, интегрируя априорные знания о сезонности, сравните с классическим подходом по точности прогноза и постройте графики остатков.

4. Анализируйте влияние маркетинговых расходов на рост клиентов с помощью байесовской линейной регрессии (данные 500 наблюдений), оцените параметры с учетом априорных убеждений и создайте предсказания для будущих сценариев.

5. Для модели прогнозирования спроса на продукт проведите кросс-валидацию байесовской регрессии, сравните с обычной OLS-регрессией по метрикам (RMSE, MAE), и визуализируйте апостериорные распределения коэффициентов для интерпретации значимости факторов.

Домашнее задание: Временные ряды и Гауссовские процессы

1. Опишите и реализуйте простой гауссовский процесс (с квадратичным ядром) для моделирования временного ряда цен акций (данные за 100 дней), постройте предсказания и визуализируйте неопределенность с помощью GPy или scikit-learn.

2. Выберите временной ряд (например, уровень безработицы), задайте различные априорные гауссовские процессы (с разными ядрами: RBF, Matérn), обновите их на данных и сравните влияние априоров на точность прогноза на 10 шагов вперед.

3. Для большого временного ряда (1000 точек) примените аппроксимацию (например, FITS или sparse GP) к моделированию температуры, рассчитайте вычислительную эффективность и точность предсказаний, используя библиотеку GPflow.

4. Создайте смесь гауссовских процессов для анализа многомодального временного ряда (например, продажи с сезонными пиками), определите компоненты смеси, оцените параметры и визуализируйте предсказания с доверительными интервалами.

5. Моделируйте временной ряд выручки компании с помощью гауссовского процесса, интегрируя априорные знания о трендах, оцените параметры и постройте прогноз на квартал вперед, сравнив с ARIMA-моделью по метрикам точности.

Примерное описание задания и критерии оценивания к проекту

Данные

1. **Небольшой датасет** из научного исследования (или свой)
2. Исследуется, например
 - a. причинно-следственная связь
 - b. Эффект от интервенции (лекарство, политика)
3. Необходимо провести описательный анализ

Исследование

1. Исследование должно быть направлено **на проверку гипотезы, не на качество предсказания** (оно важно лишь во вторую очередь)
2. **Рекомендую брать текущий диплом или бывшие темы дипломов.**
3. Если выбрана макроэкономическая тема, можно брать данные, например из world bank, бирж, для этого есть хорошие пакеты выгрузки
 - a. [World bank Python API](#)
 - b. АПИ Мосбиржи
 - c. Любые другие источники
4. Данные не должны быть сгенерированы
5. Если используются закрытые данные, их надо приложить к проекту

Группы

1. Проект выполняется в малой группе 3-4 человека
2. Защита проектов **в день экзамена**
3. Баллы за проект и презентацию у всех членов группы одинаковые

Критерии

Проект и презентация весят одинаково. 50% + 50%

Содержание проекта

Разбивка первых 50%:

1. Проведен разведочный анализ данных **(5%)**
 - a. Должен включать наглядные демонстрации важных особенностей данных необходимых для проведения байесовского моделирования.
2. Построена формальная байесовская модель для проверки гипотезы. **(10%)**
 - a. Модель должна включать в себя все необходимые уравнения, априорные распределения
 - b. Все распределения имеют адекватный домен
 - i. нет негативных плотностей
 - ii. вероятностей больше единицы
 - c. Модель компилируется, запускается, качество модели и параметризация не принимаются во внимание

3. Проведен анализ априорных распределений, выбор структуры модели обоснован и донесен до читателя **(15%)**
 - a. В Jupyter ноутбуке объяснены все переменные
 - b. Структура модели, возможные репараметризации описаны и понятны
 - c. Выбор каждого априорного распределения обоснован, по необходимости, визуализирован. Пользоваться prior predictive можно и нужно для коррекции своего выбора.
 - d. Проведен prior predictive study, если это возможно (смотрим как выглядит априорное), выбирается и анализируется адекватная статистика модели. За помощью можно обращаться к преподавателю или задав вопрос на КНЧ.
 - e. *Если prior predictive study отсутствует, это нужно согласовать (доказать, что его не получится сделать) с преподавателем МИНИМУМ за неделю до сдачи первой части проекта, иначе будет неполный балл за секцию. Скорее всего найдем альтернативную статистику. Не откладываете*
4. Получено апостериорное распределение **(10%)**
 - a. На эти 10% работает коэффициент, как в первой домашке
 - b. Формула коэффициента ($1 - \text{divergences}/\text{samples}$)
 - c. Если дивергенций более 25% следующий пункт будет со штрафом 50%
5. Анализ апостериорного распределения **(10%)**
 - a. Проверка того, что семплирование прошло хорошо
 - i. Trace Plot
 - ii. Energy Plot
 - iii. ...
 - iv. Достаточно, чтобы убедить, что проблем нет
 - v. Если проблемы есть или были, то их описание и попыток борьбы с ними
 - b. Необходимая интерпретация и графики для проверки гипотезы
 - c. **Выводы**

Презентация

1. Постановка проблемы **(10%)**
2. Презентация важных аспектов данных **(10%)**
3. Мотивация используемой модели **(20%)**
 - a. Структура
 - b. Априорные распределения
4. Интерпретация результатов **(10%)**

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Если Вы оцениваете уровень понимания апостериорного распределения после решения задач, как одним словом называется такой анализ?	самооценка	УК-6
2.	Если Вы определяете порядок освоения тем: регрессия, А/В-тесты, гауссовские процессы, как одним словом называется такой процесс?	приоритизация	УК-6

3.	Если Вы постепенно улучшаете модель через повторные обновления и анализ, как одним словом называется такой подход?	итеративность	УК-6
4.	Как одним словом называется распределение параметра до учета наблюдаемых данных в байесовском моделировании?	априорное	ОПК-2
5.	Как одним словом называется процесс численного получения выборки из сложного апостериорного распределения?	МСМС	ОПК-2
6.	Как одним словом называется функция вероятности наблюдаемых данных при фиксированном параметре?	правдоподобие	ОПК-2
7.	Как одним словом называется распределение параметра после учета данных?	апостериорное	ОПК-2
8.	Если гауссовский процесс полностью задается средней функцией и ковариационной функцией, как одним словом называется такая ковариационная функция?	ядро	ОПК-2
9.	Если апостериорный интервал для параметра полностью расположен по одну сторону от нуля, как одним словом называется управленческое решение о запуске изменения в продукте?	внедрение	ПК-3
10.	Если в байесовской линейной регрессии предполагается распределение ошибок с плотностью колоколообразной формы, как одним словом называется такое распределение?	нормальное	ПК-3
11.	Если модель строится как комбинация нескольких гауссовских процессов, как одним словом называется такая конструкция?	смесь	ПК-3
12.	Если распределение параметра представлено графиком плотности для наглядности, как одним словом называется такой способ представления результатов?	визуализация	ПК-4
13.	Если в начале доклада формулируется основной полученный результат исследования, как одним словом называется эта часть выступления?	вывод	ПК-4
14.	Если в докладе показываются различия между распределением до учета данных и после их учета, как одним словом называется такой формат представления?	сравнение	ПК-4
15.	Если во время защиты исследования отвечают на вопросы аудитории о выборе априорных предположений, как одним словом называется этот этап?	обсуждение	ПК-4
16.	Если результаты исследования оформлены в структурированном документе с таблицами и графиками, как одним словом называется такой документ?	отчет / отчёт	ПК-4