

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО
Проректор по учебной работе

В.М. Говорун

	Рабочая программа дисциплины (модуля)
по дисциплине:	Моделирование и анализ данных
по направлению:	Прикладные математика и физика
профиль подготовки:	Системная и синтетическая биология Физтех-школа Биологической и Медицинской Физики кафедра системной и синтетической биологии
курс:	4
квалификация:	бакалавр

Семестры, формы промежуточной аттестации:

7 (осенний) - Зачет

8 (весенний) - Дифференцированный зачет

Аудиторных часов: 75 всего, в том числе:

лекции: 30 час.

семинары: 45 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 60 час.

Всего часов: 135, всего зач. ед.: 3

Программу составил: В.М. Говорун, д-р биол. наук, профессор

Программа обсуждена на заседании кафедры системной и синтетической биологии 10.03.2025

Аннотация

Дисциплина направлена на предоставление учащимся целостного понимания современных методов и инструментов анализа экспериментальных результатов, с особым акцентом на статистическое и байесовское моделирование. Курс нацелен на развитие критического мышления и навыков решения сложных аналитических задач, которые необходимы для эффективной работы с экспериментальными данными и сложными наборами данных в различных областях науки.

1. Цели и задачи

Цель дисциплины

- является получение студентами целостного представления о методах современного анализа экспериментальных результатов.

Задачи дисциплины

- овладеть базовыми понятиями статистического анализа данных и основными способами представления результатов анализа;
- освоить байесовский подход в анализе данных, включая понимание его философии и технических аспектов;
- научиться применить преобразования данных для борьбы с композиционностью в экспериментальных данных;
- освоить принципы математического анализа и моделирования генно-регуляторных сетей;
- освоить принципы анализа нуклеотидных последовательностей при помощи современных подходов, включая большие языковые модели.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	УК-6.1 Определяет приоритеты профессиональной деятельности и способы ее совершенствования на основе самооценки
	УК-6.2 Способен планировать самостоятельную деятельность в решении профессиональных задач; подвергать критическому анализу проделанную работу; находить и творчески использовать имеющийся опыт в соответствии с задачами саморазвития
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.3 Знает основные требования информационной безопасности

ПК-1 Способен планировать и проводить научные эксперименты (в избранной предметной области) и (или) теоретические (аналитические и имитационные) исследования	ПК-1.1 Владеет фундаментальными понятиями, законами и теориями современной физики
	ПК-1.3 Владеет культурой постановки научной задачи и моделирования естественнонаучных объектов и систем
	ПК-1.5 Владеет навыками безопасной работы с современными научными приборами и другим экспериментальным оборудованием
ПК-2 Способен анализировать полученные в ходе научно-исследовательской работы данные и делать научные выводы (заключения)	ПК-2.1 Владеет методами статистической обработки и анализа научных данных
	ПК-2.2 Умеет находить ключевые параметры, определяющие изучаемое явление, и производить численные оценки по порядку величины
	ПК-2.3 Способен представлять научные утверждения, их обоснования и доказательства, научные проблемы и их решения ясно и точно в терминах, понятных для профессиональной аудитории, в письменной и устной форме
ПК-3 Способен выбирать и применять подходящее оборудование, инструменты и методы исследований для решения задач в избранной предметной области	ПК-3.1 Знает принципы работы и диапазоны рабочих параметров используемого научного оборудования
	ПК-3.2 Знает области и критерии применимости используемых теоретических подходов и умение оценивать точность приближенных аналитических методов вычислений
	ПК-3.3 Умеет производить оценку точности численных методов, используемых на ЭВМ, вычислительной сложности используемых алгоритмов и объема требуемых вычислительных ресурсов
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области
	ПК-4.2 Знает источники происхождения и умеет производить оценку погрешности измерений и достоверности экспериментальных результатов
	ПК-4.3 Способен обосновать причинно-следственные отношения используемых понятий и моделей

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- базовые понятия классической статистики (ноль-гипотеза, p-value, поправка на множественное сравнение) и их ограничения;
- теорию Байесовского анализа и способы проведения анализа на практике;
- основы синтаксиса языка описания вероятностных моделей STAN;
- особенности работы с композиционными данными;
- фундаментальные идеи в компьютерном моделировании биологических систем.

уметь:

- ставить цели и задачи, понимать поставленные цели и задачи при создании аналитических систем в медико-биологических исследованиях;
- использовать свои знания для решения задач и достижения конечного результата «выхода»;
- оценивать корректность постановок задач и строить алгоритмы достижения их оптимального решения в различных (в том числе меняющихся) условиях.

владеть:

- постройки эффективных схем анализа данных;
- написания программ для моделирования данных и их статистического анализа.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Основные используемые статистики. Принцип работы статистических тестов	2			1
2	Параметрические и непараметрические статистические методы.	2			2
3	Генерация данных и проведение статистического анализа. Экспериментальное исследование мощности t-теста и теста Манна-Уитни.	2			
4	Проверка равенства распределений. Тесты Колмогорова-Смирнова и Андерсона-Дарлинга.	2			
5	Байесовский статистический анализ. Оценка неопределённости.	2			2
6	Практические основы байесовских оценок. Реализация подхода на сетке параметров.	2			2
7	Практические основы байесовских оценок. Алгоритм Метрополиса.	2			
8	Метод Hamiltonian Monte Carlo. Язык Stan для построения вероятностных моделей.	2			
9	Анализ смешанных распределений	2			2
10	Как не обмануть себя при анализе зависимостей. Confounding, Collider и M-bias.	2			
11	Источники артефактных зависимостей. Парадокс Симпсона. Ошибка отбора после воздействия (post-treatment bias).	2			2
12	Понятие энтропии распределений случайных величин. Кросс-энтропия, дивергенция Кульбака-Лейблера.	4			2
13	Информационные критерии сравнения моделей: BIC, WAIC, LOO-CV. Сравнение моделей на практике.	2			2
14	Моделирование данных с линейной зависимостью, анализ методом наименьших квадратов. Анализ при помощи байесовского подхода.	2			
15	Логистическая регрессия. Вероятностные модели для целочисленных случайных величин. Биномиальное и отрицательное биномиальное распределение. Применение для анализа омиксных данных.		6		8

16	Композиционные данные в омиксных исследованиях, их анализ. Преобразования CLR, ILR, ALR.		8		6
17	Работа с неполными данными		8		8
18	Компьютерное моделирование биологических систем. Самоорганизация и коллективное поведение		9		8
19	Анализ и моделирование генно-регуляторных сетей. Сетевые мотивы и их функциональная роль		8		9
20	Анализ и моделирование нуклеотидных последовательностей. Скрытые марковские модели, рекуррентные нейронные сети, применение больших языковых моделей для анализа биологических данных		6		6
Итого часов		30	45		60
Подготовка к экзамену		0 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 7 (Осенний)

1. Основные используемые статистики. Принцип работы статистических тестов

Вспоминаем статистики (среднее, медиану, перцентили, дисперсию, стандартное отклонение). Несмещенные оценки. Распределение случайной величины и плотность распределения. Разбираем как устроен статистический тест: нулевая гипотеза, подсчет статистики, расчет p-value. Как распределено p-value при верности нулевой гипотезы.

2. Параметрические и непараметрические статистические методы.

Различие параметрических и непараметрических статистических методов. Откуда берутся предположения для параметрических методов. Центральная предельная теорема. Принцип максимизации энтропии.

3. Генерация данных и проведение статистического анализа. Экспериментальное исследование мощности t-теста и теста Манна-Уитни.

Нормальное распределение, распределение Стьюдента, распределение хи-квадрат, лог-нормальное распределение. Мощность статистических тестов. Схема генерации случайных величин с нормальным распределением; добавляем выбросы. Сравниваем поведение t-теста и теста Манна-Уитни на данных, содержащих и не содержащих выбросы. Сравниваем мощность тестов на данных, соответствующих и не соответствующих нулевой гипотезе.

4. Проверка равенства распределений. Тесты Колмогорова-Смирнова и Андерсона-Дарлингга.

Различие в нулевых гипотезах статистических тестов, равенство центральных тенденций, равенство распределений. Статистики, используемые в тестах Колмогорова-Смирнова и Андерсона-Дарлингга. Проверка их работы на смоделированных данных.

5. Байесовский статистический анализ. Оценка неопределённости.

Формула условной вероятности и формула Байеса, пример с тестом на редкое заболевание. Разные интерпретации вероятности. Обновление оценок распределений параметров после получения новых данных. Важность оценки неопределённости.

6. Практические основы байесовских оценок. Реализация подхода на сетке параметров.

Задание априорных распределений. Простой подход для анализа вероятностного пространства при помощи сетки параметров. Маргинальные распределения. Доверительные интервалы и оценка неопределенности.

7. Практические основы байесовских оценок. Алгоритм Метрополиса.

Проблема размерности при полном переборе на сетке параметров. Алгоритм Метрополиса как способ ускорения вычислений. Визуализация проб из вероятностного пространства.

8. Метод Hamiltonian Monte Carlo. Язык Stan для построения вероятностных моделей.

Метод Hamiltonian Monte Carlo как способ эффективного семплирования из вероятностного пространства. Знакомство с синтаксисом языка Stan. Блоки описания данных, параметров и вероятностной модели. Реализация линейной модели средствами языка Stan

9. Анализ смешанных распределений

Реализация на языке Stan модели, описывающей смесь нескольких распределений.

10. Как не обмануть себя при анализе зависимостей. Confounding, Collider и M-bias.

Принципы анализа причинно-следственных зависимостей, представление в виде ациклического графа (DAG). Сценарии возникновения кажущихся зависимостей там, где их нет. Методы борьбы с ними. Практическая реализация сценариев: confounding, collider и M-bias.

11. Источники артефактных зависимостей. Парадокс Симпсона. Ошибка отбора после воздействия (post-treatment bias).

Типы возможных связей между тремя переменными. Условия возникновения парадокса Симпсона. Необходимость стратификации выборки. Примеры схем анализа, в которых возникает ошибка отбора после воздействия, способы ее избежать.

12. Понятие энтропии распределений случайных величин. Кросс-энтропия, дивергенция Кульбака-Лейблера.

Формула энтропии. Расчет энтропии исходов для честной и нечестной монеты, испытаний Бернулли с различным количеством исходов. Способы сравнения истинного распределения случайной величины и его оценки.

13. Информационные критерии сравнения моделей: BIC, WAIC, LOO-CV. Сравнение моделей на практике.

Байесовский информационный критерий. Широко применимый информационный критерий. Перекрестная проверка с исключением по одному.

14. Моделирование данных с линейной зависимостью, анализ методом наименьших квадратов. Анализ при помощи байесовского подхода.

Регрессия к среднему, пример с изучением роста. Анализ методом наименьших квадратов. Почему не используются перпендикуляры к линии. Анализ при помощи байесовского подхода. Как неопределенность параметров зависит от количества данных, от зашумленности исходных данных. Устойчивость к выбросам.

Семестр: 8 (Весенний)

15. Логистическая регрессия. Вероятностные модели для целочисленных случайных величин. Биномиальное и отрицательное биномиальное распределение. Применение для анализа омиксных данных.

Логистическая регрессия как метод для моделирования вероятности бинарных исходов. Применение отрицательного биномиального распределения для учета переменной дисперсии. Интерпретация коэффициентов модели и их связь с вероятностью.

16. Композиционные данные в омиксных исследованиях, их анализ. Преобразования CLR, ILR, ALR.

Композиционные данные: данные, представленные в виде пропорций или процентов, сумма которых равна 1. Преобразование CLR (Centered Log-Ratio). Преобразование ALR (Additive Log-Ratio). Преобразование ILR (Isometric Log-Ratio). Применение преобразований для анализа составов в омиксных данных

17. Работа с неполными данными

Проблемы, связанные с недостающими данными и их влияние на анализ. Методы обработки неполных данных: удаление, импутация, максимизация правдоподобия. Применение байесовских методов для работы с недостающими данными.

18. Компьютерное моделирование биологических систем. Самоорганизация и коллективное поведение

Принципы самоорганизации: как локальные взаимодействия приводят к глобальным структурам. Игра Жизнь, модель боидов, swarm chemistry, самоорганизация Dycistellium для образования плодовых тел. Системы реакции-диффузии.

19. Анализ и моделирование генно-регуляторных сетей. Сетевые мотивы и их функциональная роль

Генно-регуляторные сети: взаимодействие генов, белков и других молекул для регуляции экспрессии генов. Применение методов анализа графов для выявления ключевых элементов и структур в сетях. Моделирование генно-регуляторных сетей с помощью дифференциальных уравнений

20. Анализ и моделирование нуклеотидных последовательностей. Скрытые марковские модели, рекуррентные нейронные сети, применение больших языковых моделей для анализа биологических данных

Скрытые марковские модели (HMM) для анализа последовательностей ДНК. Рекуррентные нейронные сети (RNN) для работы с длинными последовательностями, моделирование зависимостей между нуклеотидами. Архитектура сетей кодировщиков-декодировщиков, трансформеры. Применение трансформеров для анализа последовательностей ДНК: поиск регуляторных последовательностей, функциональная аннотация, таксономическая классификация.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебные аудитории, оснащенные мультимедийным оборудованием (экран, проектор, аудио и видеоаппаратура, ноутбук с подключением к сети «Интернет», микрофоны). Для проведения семинарских занятий требуется аудитория оснащенная достаточным количеством компьютеров для самостоятельной работы каждого студента или групп по 2-3 человека.

6. Перечень рекомендуемой литературы

Основная литература

Основная литература

Основная литература предоставлена базовой кафедрой:

McElreath, Richard. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC, 2018.

Alon, Uri. An introduction to systems biology: design principles of biological circuits. Chapman and Hall/CRC, 2019.

Дополнительная литература

Дополнительная литература

Литература предоставлена базовой кафедрой:

Петри А., Сэбин К., Наглядная медицинская статистика ISBN 978-5-9704-3373-7

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Стентон Гланц Медико-биологическая статистика, электронная книга

<https://medstatistic.ru/articles/glantz.pdf>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

Для части занятий потребуется Яндекс телемост. Яндекс диск для доступа к материалам курса. Приветствуется наличие во время занятий смартфонов/ноутбуков для участия в интерактивных упражнениях.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Успешное освоение курса требует напряжённой самостоятельной работы студента. В программе курса приведено минимально необходимое время для работы студента над темой.

Самостоятельная работа включает в себя:

- проработку учебного материала (по конспектам лекций, учебной и научной литературе);
- подготовку ответов на вопросы, предназначенные для самостоятельного изучения, доказательство отдельных утверждений, свойств.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению: Прикладные математика и физика
профиль подготовки: Системная и синтетическая биология
Физтех-школа Биологической и Медицинской Физики
кафедра системной и синтетической биологии
курс: 4
квалификация: бакалавр

Семестры, формы промежуточной аттестации:

7 (осенний) - Зачет

8 (весенний) - Дифференцированный зачет

Разработчик: В.М. Говорун, д-р биол. наук, профессор

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	УК-6.1 Определяет приоритеты профессиональной деятельности и способы ее совершенствования на основе самооценки
	УК-6.2 Способен планировать самостоятельную деятельность в решении профессиональных задач; подвергать критическому анализу проделанную работу; находить и творчески использовать имеющийся опыт в соответствии с задачами саморазвития
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
	ОПК-2.2 Знает и умеет применять численные математические методы и прикладное программное обеспечение для решения научных задач в профессиональной области
	ОПК-2.3 Знает основные требования информационной безопасности
ПК-1 Способен планировать и проводить научные эксперименты (в избранной предметной области) и (или) теоретические (аналитические и имитационные) исследования	ПК-1.1 Владеет фундаментальными понятиями, законами и теориями современной физики
	ПК-1.3 Владеет культурой постановки научной задачи и моделирования естественнонаучных объектов и систем
	ПК-1.5 Владеет навыками безопасной работы с современными научными приборами и другим экспериментальным оборудованием
ПК-2 Способен анализировать полученные в ходе научно-исследовательской работы данные и делать научные выводы (заключения)	ПК-2.1 Владеет методами статистической обработки и анализа научных данных
	ПК-2.2 Умеет находить ключевые параметры, определяющие изучаемое явление, и производить численные оценки по порядку величины
	ПК-2.3 Способен представлять научные утверждения, их обоснования и доказательства, научные проблемы и их решения ясно и точно в терминах, понятных для профессиональной аудитории, в письменной и устной форме
ПК-3 Способен выбирать и применять подходящее оборудование, инструменты и методы исследований для решения задач в избранной предметной области	ПК-3.1 Знает принципы работы и диапазоны рабочих параметров используемого научного оборудования
	ПК-3.2 Знает области и критерии применимости используемых теоретических подходов и умение оценивать точность приближенных аналитических методов вычислений
	ПК-3.3 Умеет производить оценку точности численных методов, используемых на ЭВМ, вычислительной сложности используемых алгоритмов и объема требуемых вычислительных ресурсов

ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области
	ПК-4.2 Знает источники происхождения и умеет производить оценку погрешности измерений и достоверности экспериментальных результатов
	ПК-4.3 Способен обосновать причинно-следственные отношения используемых понятий и моделей

2. Показатели оценивания компетенций

В результате изучения дисциплины «Моделирование и анализ данных» обучающийся должен:

знать:

- базовые понятия классической статистики (ноль-гипотеза, p-value, поправка на множественное сравнение) и их ограничения;
- теорию Байесовского анализа и способы проведения анализа на практике;
- основы синтаксиса языка описания вероятностных моделей STAN;
- особенности работы с композиционными данными;
- фундаментальные идеи в компьютерном моделировании биологических систем.

уметь:

- ставить цели и задачи, понимать поставленные цели и задачи при создании аналитических систем в медико-биологических исследованиях;
- использовать свои знания для решения задач и достижения конечного результата «выхода»;
- оценивать корректность постановок задач и строить алгоритмы достижения их оптимального решения в различных (в том числе меняющихся) условиях.

владеть:

- постройки эффективных схем анализа данных;
- написания программ для моделирования данных и их статистического анализа.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

1. Какие методы создания данных вы знаете?
2. Какие артефактные зависимости могут возникать при анализе данных?
3. В чем суть информационных критериев BIC и AIC?
4. Как сравнивать модели с использованием критериев BIC и AIC?
5. Какие методы поиска ассоциаций в омиксных данных являются наиболее надежными?
6. В чем заключается анализ дифференциальной представленности?
7. Какие современные методы используются для анализа омиксных данных?
8. Что такое композиционные данные и как с ними работать?
9. В чем суть преобразований CLR, ILR и ALR?
10. Как провести анализ композиционных данных с использованием CLR-преобразований?
11. Какие методы оценки уровня неопределенности в моделях машинного обучения существуют?
12. Что такое метод конформного предсказания и как его использовать на практике?
13. Как применяются большие языковые модели для анализа биологических данных?
14. Какие задачи решаются при помощи сетевого анализа?
15. Как визуализировать данные омиксного анализа?
16. Какие проблемы могут возникать при преобразовании данных и как их решать?

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Типовые вопросы для зачета

1. Объясните основные задачи теории вероятностей и их применение в научных исследованиях.

2. Опишите процесс настройки вычислительной среды для работы с языком R. Какие шаги необходимо выполнить для начальной настройки?
3. Какие синтаксические особенности языка программирования R являются наиболее важными для начинающих? Приведите примеры.
4. Какие существуют основные методы визуализации данных в R и каковы их преимущества и недостатки?
5. Объясните различия между параметрическими и непараметрическими методами статистического анализа. Приведите примеры применения каждого из них.
6. Опишите процесс генерации данных для статистического анализа и проведите сравнение мощности t-теста и теста Манна-Уитни.
7. В чем заключаются основные принципы байесовского подхода к анализу данных? Какие преимущества он имеет перед частотным подходом?
8. Объясните метод наименьших квадратов для моделирования данных с линейной зависимостью. Каковы основные этапы этого метода?
9. Приведите пример реализации парадокса Монти Холла в R и объясните теоретические аспекты этого парадокса.
10. В чем заключается применение байесовского анализа для оценки уверенности в статистических оценках? Какие подходы используются для реализации brute-force анализа в байесовских моделях?
11. Объясните понятие confounding, collider и M-bias. Как они могут влиять на результаты анализа данных?
12. Каковы основные ошибки, которые могут возникать при анализе зависимостей в данных? Как их избежать?
13. Какие методы визуализации данных наиболее эффективны для выявления зависимостей и артефактов в данных?
14. Объясните роль и применение линейных моделей в анализе данных. Какие альтернативные методы существуют?
15. Какое значение имеет байесовский подход в моделировании данных с учетом неопределенности?

Типовые вопросы для дифференцированного зачета

1. Какие методы создания данных вы знаете?
2. Какие артефактные зависимости могут возникать при анализе данных?
3. В чем суть информационных критериев BIC и AIC?
4. Как сравнивать модели с использованием критериев BIC и AIC?
5. Какие методы поиска ассоциаций в омиксных данных являются наиболее надежными?
6. В чем заключается анализ дифференциальной представленности?
7. Какие современные методы используются для анализа омиксных данных?
8. Что такое композиционные данные и как с ними работать?
9. В чем суть преобразований CLR, ILR и ALR?
10. Как провести анализ композиционных данных с использованием CLR-преобразований?
11. Какие методы оценки уровня неопределенности в моделях машинного обучения существуют?
12. Что такое метод конформного предсказания и как его использовать на практике?
13. Как применяются большие языковые модели для анализа биологических данных?
14. Какие задачи решаются при помощи сетевого анализа?
15. Как визуализировать данные омиксного анализа?
16. Какие проблемы могут возникать при преобразовании данных и как их решать?

Критерии оценивания

Оценка отлично (10 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (9 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (8 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо (7 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо (6 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо (5 баллов) - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно (4 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно (3 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно (2 балла) - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно (1 балл) - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

Оценка «зачтено» выставляется студенту, если он показал всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка «не зачтено» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

При проведении зачета и дифференцированного зачета обучающемуся предоставляется 45 минут на подготовку. Опрос обучающегося не должен превышать одного астрономического часа.