

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

УТВЕРЖДЕНО
Заместитель директора

Ю.О. Соболев

	Рабочая программа дисциплины (модуля)
по дисциплине:	Рекомендательные системы
по направлению:	Прикладная математика и информатика
профиль подготовки:	Науки о данных центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
курс:	2
квалификация:	магистр

Семестр, формы промежуточной аттестации: 4 (весенний) - Дифференцированный зачет

Аудиторных часов: 22 всего, в том числе:

лекции: 2 час.

семинары: 20 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 113 час.

Всего часов: 135, всего зач. ед.: 3

Программу составил: К.А. Лапин, старший методист

Программа обсуждена на заседании центра дополнительного, дополнительного профессионального и
онлайн-образования "Пуск" 01.03.2025

Аннотация

Целью данной дисциплины является дать практические навыки применения языка R, IDE R-studio и основных пакетов для анализа данных. Студент после освоения курса будет понимать основные пакеты программной среды R, основы синтаксиса R.

1. Цели и задачи

Цель дисциплины

Дать практические навыки применения языка R, IDE R-studio и основных пакетов для анализа данных.

Задачи дисциплины

- обучить основам языка программирования R;
- научить пользоваться IDE R-studio;
- ознакомить с основными пакетами для анализа данных с помощью R;
- формирование у студентов основных биоинформатических навыков и приобретение ими практического опыта, необходимого для проведения самостоятельных научных исследований в области системной биологии.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
ПК-7 Способен разрабатывать и реализовывать инновационные технологические проекты, осуществлять технико-экономическое обоснование инновационных проектов	ПК-7.1 Знает теорию и методы управления высокотехнологическими проектами и принципы их технико-экономического обоснования
	ПК-7.2 Умеет управлять высокотехнологическим проектом для эффективного достижения целей проекта в рамках утвержденных заказчиком требований, бюджета и сроков
	ПК-7.3 Владеет методами планирования, организации исполнения, контроля, анализа отклонений и коррекции исполнения

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные пакеты программной среды R;
- основы синтаксиса R.

уметь:

- программировать на языке R;
- имплементировать и отлаживать биоинформатические алгоритмы;
- реализовывать статистический анализ в программной среде R.

владеть:

- навыками работы с большими объемами биологических данных;
- культурой планирования и осуществления многоступенчатого биоинформатического 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	Введение в рекомендательные системы	2	6		20
2	Коллаборативная фильтрация		6		30
3	Контент-базированные методы и гибридные подходы		4		30
4	Глубокое обучение в рекомендательных системах		4		33
Итого часов		2	20		113
Подготовка к экзамену		0 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

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

Семестр: 4 (Весенний)

1. Введение в рекомендательные системы

Основные понятия и задачи рекомендательных систем. Типы рекомендательных систем: коллаборативная фильтрация, контент-базированные методы, гибридные подходы. Обзор современных рекомендательных систем.

2. Коллаборативная фильтрация

Методы коллаборативной фильтрации: user-based и item-based подходы. Матричные разложения (SVD, ALS). Оценка качества рекомендаций: метрики precision, recall, RMSE.

3. Контент-базированные методы и гибридные подходы

Контент-базированные методы: TF-IDF, word2vec. Гибридные подходы: комбинирование коллаборативной фильтрации и контент-базированных методов. Практические примеры гибридных систем.

4. Глубокое обучение в рекомендательных системах

Нейронные сети в рекомендательных системах. Рекуррентные нейронные сети (RNN) и их применение в рекомендательных системах. Примеры использования глубокого обучения в реальных системах.

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

Оборудование, необходимое для семинаров: аудитория, компьютер и мультимедийное оборудование (проектор, звуковая система), UNIX сервер с отдельным аккаунтом для каждого студента

6. Перечень рекомендуемой литературы

Основная литература

1. Python для сложных задач: наука о данных и машинное обучение [Текст] / Дж. Вандер Плас ; [пер. с англ. И. Пальти], СПб., Питер, 2018
2. Введение в методы машинного обучения с подкреплением, учебное пособие /А. И. Панов; Министерство науки и высшего образования Российской Федерации ; Московский физико-технический институт (национальный исследовательский университет). Москва, МФТИ, 2019
3. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвилль . — Москва, ДМК Пресс, 2018.— URL: <https://e.lanbook.com/book/107901> (дата обращения: 29.01.2021). - Полный текст (Режим доступа : из сети МФТИ / Удаленный доступ)

Дополнительная литература

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

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных — <http://www.machinelearning.ru/>
2. Python Numpy Tutorial - <https://cs231n.github.io/python-numpy-tutorial/>
3. Введение в анализ данных с помощью Pandas - <https://habr.com/ru/post/196980/>
4. Learn Git Branching - https://learngitbranching.js.org/?locale=ru_RU
5. https://scikit-learn.org/stable/supervised_learning.html#supervised-learning

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

Доступ в Интернет, UNIX сервер с отдельным аккаунтом для каждого студента. Для части занятий потребуется Zoom. Google Drive 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)

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладная математика и информатика
профиль подготовки:	Науки о данных центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
курс:	<u>2</u>
квалификация:	магистр
Семестр, формы промежуточной аттестации: 4 (весенний) - Дифференцированный зачет	
Разработчик:	К.А. Лапин, старший методист

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
ПК-7 Способен разрабатывать и реализовывать инновационные технологические проекты, осуществлять технико-экономическое обоснование инновационных проектов	ПК-7.1 Знает теорию и методы управления высокотехнологическими проектами и принципы их технико-экономического обоснования
	ПК-7.2 Умеет управлять высокотехнологическим проектом для эффективного достижения целей проекта в рамках утвержденных заказчиком требований, бюджета и сроков
	ПК-7.3 Владеет методами планирования, организации исполнения, контроля, анализа отклонений и коррекции исполнения

2. Показатели оценивания компетенций

В результате изучения дисциплины «Рекомендательные системы» обучающийся должен:

знать:

- основные пакеты программной среды R;
- основы синтаксиса R.

уметь:

- программировать на языке R;
- имплементировать и отлаживать биоинформатические алгоритмы;
- реализовывать статистический анализ в программной среде R.

владеть:

- навыками работы с большими объемами биологических данных;
- культурой планирования и осуществления многоступенчатого биоинформатического анализа.

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

Во время текущего контроля студент должен уметь ответить на следующие вопросы:

1. Сгенерируйте два случайных Пуассоновских набора по 200 чисел, один с средним 0.5, другой с средним 3. Есть ли между ними линейная зависимость?
2. Какая структура подойдёт для хранения значений температуры, измеряемой каждую минуту у пяти больных за сутки. Создайте такую структуру и заполните ее случайными данными. Когда у второго пациента температура была выше 40 градусов?
3. Создайте новую числовую переменную `new_var` в данных `mtcars`, которая содержит единицы в строках, если в машине не меньше четырёх карбюраторов (переменная `"carb"`) или больше шести цилиндров (переменная `"cyl"`). В строках, в которых условие не выполняется, должны стоять нули.
4. В датафрейме `mtcars` создайте новую колонку (переменную) под названием `even_gear`, в которой будут единицы, если значение переменной (`gear`) четное, и нули если количество нечетное.

5. Создайте 3 линейные переменные одинаковой длины с произвольными названиями, содержащие Проверьте, действительно ли сумма первых двух чисел строго больше, чем третье число. Результат сравнения (TRUE или FALSE) сохраните в новую переменную с именем result.
6. Превратите датафрейм mtcars в лист и создайте новый элемент листа под названием even_gear, в которой будут единицы, если значение переменной (gear) четное, и нули если количество нечетное.
7. Создайте структуру для хранения имени, фамилии, возраста и пола трех человек и заполните ее. Как сделать простой поиск по фамилии?
8. Пьяный идет по мосту шириной l шагов. Каждый шаг пьяный смещается случайно на один шаг вправо или влево. Если пьяный переступит через край моста — он падает. Напишите функцию, которая считает, сколько шагов сделает пьяный до падения. Ширина моста и положение пьяного — начальные параметры. Drunken_path(l), вывод: 'Our boozier will fall at n-d step'.
9. Напишите функцию, которая получает на вход две экспериментально померенные зависимости (x1,y1,x2,y2), аппроксимирует их прямыми методом наименьших квадратов и возвращает координаты пересечения.
10. Напишите функцию, которая сравнивает набор слов одинаковой длины и возвращает матрицу попарных расстояний — долю несовпадающих букв.
11. Напишите функцию, которая рисует траекторию луны относительно Солнца. Считать, что земля движется равномерно со скоростью V1 вокруг Солнца по орбите радиуса R1. Луна движется равномерно со скоростью V2 вокруг земли по орбите радиуса R1.

Во время занятий могут проходить интерактивные обсуждения в чатах курса, что будет являться домашним заданием. Возможно выполнение патентного поиска в качестве самостоятельной задачи. Успешное выполнение всех заданий по курсу и выполнение контрольных срезов знаний дает преимущество на дифференцированном зачете.

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

1. Сгенерируйте два случайных Пуассоновских набора по 200 чисел, один с средним 0.5, другой с средним 3. Есть ли между ними линейная зависимость?
2. Какая структура подойдет для хранения значений температуры, измеряемой еже часно у пяти больных за сутки. Создайте такую структуру и заполните ее случайными данными. Когда у второго пациента температура была выше 40 градусов?
3. Создайте новую числовую переменную new_var в данных mtcars, которая содержит единицы в строках, если в машине не меньше четырёх карбюраторов (переменная "carb") или больше шести цилиндров (переменная "cyl"). В строках, в которых условие не выполняется, должны стоять нули.
4. В датафрейме mtcars создайте новую колонку (переменную) под названием even_gear, в которой будут единицы, если значение переменной (gear) четное, и нули если количество нечетное.
5. Создайте 3 линейные переменные одинаковой длины с произвольными названиями, содержащие Проверьте, действительно ли сумма первых двух чисел строго больше, чем третье число. Результат сравнения (TRUE или FALSE) сохраните в новую переменную с именем result.
6. Превратите датафрейм mtcars в лист и создайте новый элемент листа под названием even_gear, в которой будут единицы, если значение переменной (gear) четное, и нули если количество нечетное.
7. Создайте структуру для хранения имени, фамилии, возраста и пола трех человек и заполните ее. Как сделать простой поиск по фамилии?
8. Пьяный идет по мосту шириной l шагов. Каждый шаг пьяный смещается случайно на один шаг вправо или влево. Если пьяный переступит через край моста — он падает. Напишите функцию, которая считает, сколько шагов сделает пьяный до падения. Ширина моста и положение пьяного — начальные параметры. Drunken_path(l), вывод: 'Our boozier will fall at n-d step'.
9. Напишите функцию, которая получает на вход две экспериментально померенные зависимости (x1,y1,x2,y2), аппроксимирует их прямыми методом наименьших квадратов и возвращает координаты пересечения.

10. Напишите функцию, которая сравнивает набор слов одинаковой длины и возвращает матрицу попарных расстояний — долю несовпадающих букв.
11. Напишите функцию, которая рисует траекторию луны относительно Солнца. Считать, что земля движется равномерно со скоростью V_1 вокруг Солнца по орбите радиуса R_1 . Луна движется равномерно со скоростью V_2 вокруг земли по орбите радиуса R_1 .
12. Написать функцию которая рисует заданный набор точек на плоскости и соединяет прямыми те из них, что находятся на расстоянии менее заданного (максимальное расстояние — параметр функции). Размер точек пропорционален числу ребер входящих в нее (максимальный размер — параметр функции). Цвет прямых зависит от расстояния (цвет соответствующий минимальному и максимальному расстоянию — параметры функции).
13. Написать функцию, находящую все вхождения слова (w) в текст (t) с не более чем n ошибок. W , t и n — параметры функции. Функция возвращает позиции начала вхождений.
14. Написать функцию находящую все вхождения слова (w) в текст (t) с не более чем n ошибок. W , t и n — параметры функции. Функция возвращает позиции начала вхождений.
15. Написать функцию рисующую случайную ломанную. Длина шага и поворот определяется (от предыдущего направления) определяется случайно исходя из равномерного распределения. Диапазон значений длин и поворота, равно как и число шагов — параметры функции.
16. Напишите функцию, генерирующую все возможные последовательности данной длины из данного алфавита.
17. Для встроенных в R данных `AirPassengers` рассчитайте скользящее среднее с интервалом сглаживания равным 10. Напечатайте получившийся результат (первым значением в выводе должно быть среднее для элементов 1:10, во втором значении - среднее для элементов 2:11 и т.д., в последнем - среднее для элементов 135 :144). Все полученные значения средних сохраните в переменную `moving_average`.

Примеры билетов:

1. Операторы в R.
2. Реализация функций на языке R, функции первого уровня.
3. Функции n -ного уровня.
4. Анализ главных компонент.
5. Общие принципы реализации PCA.
6. Функции `prcomp` пакета `stats`.
7. Пакет `rca3d`.

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.

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

При проведении дифференцированного зачета обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету не должен превышать одного астрономического часа.