

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Нижегородский государственный университет  
им. Н.И. Лобачевского»**

Институт информационных технологий, математики и механики

(факультет / институт / филиал)

УТВЕРЖДЕНО  
президиумом Ученого совета ННГУ  
протокол от  
«14» декабря 2021 г. № 4

**Рабочая программа дисциплины**

**Математические методы биоинформатики.**

**Дополнительные главы**

(наименование дисциплины (модуля))

Уровень высшего образования

**бакалавриат**

(бакалавриат / магистратура / специалитет)

Направление подготовки / специальность

**01.03.02 Прикладная математика и информатика**

(указывается код и наименование направления подготовки / специальности)

Направленность образовательной программы

**Прикладная математика и информатика (общий профиль)**

(указывается профиль / магистерская программа / специализация)

Форма обучения

**очная**

(очная / очно-заочная / заочная)

Нижний Новгород

2022 год приема

## 1. Место дисциплины в структуре ООП

Дисциплина относится к части, формируемой участниками образовательных отношений.

Код дисциплины Б1.В.ДВ.12.05.

№ варианта	Место дисциплины в учебном плане образовательной программы	Стандартный текст для автоматического заполнения в конструкторе РПД
2	Блок 1. Дисциплины (модули) Часть, формируемая участниками образовательных отношений	Дисциплина Б1.В.ДВ.12.05 «Математические методы биоинформатики. Дополнительные главы» относится к части ООП направления подготовки 01.03.02 «Прикладная математика и информатика», формируемой участниками образовательных отношений.

## 2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства
	Индикатор достижения компетенции* (код, содержание индикатора)	Результаты обучения по дисциплине**	
ПК-6. Способен изучать и применять программное обеспечение, проводить расчётные работы и выполнять обработку результатов исследований	ПК-6.1. Знает методы применения современных программных комплексов, пакетов прикладных программ и автоматизированных систем для решения прикладных задач при проведении исследований	<i>Знает базовые математические методы исследования для решения прикладных задач биоинформатики</i>	<i>Собеседование</i>
	ПК-6.2. Умеет самостоятельно проводить расчётные работы, выбирать и применять современные программные комплексы, пакеты прикладных программ и автоматизированные системы, обрабатывать и анализировать полученные	<i>Умеет определять и профессионально реализовывать необходимые для решения прикладных задач биоинформатики математические методы Умеет применять для исследования пакеты прикладных программ, обрабатывать и анализировать полученные результаты.</i>	<i>Практические задания</i>

	результаты		
	ПК-6.3. Имеет практический опыт применения современного программного обеспечения для решения прикладных задач	<i>Имеет практический опыт применения современных пакетов прикладных программ для решения прикладных задач биоинформатики</i>	<i>Практические задания</i>

### 3. Структура и содержание дисциплины

#### 3.1. Трудоемкость дисциплины

	<b>Очная форма обучения</b>
<b>Общая трудоемкость</b>	<b>2 ЗЕТ</b>
<b>Часов по учебному плану</b>	<b>72</b>
<b>в том числе</b>	
<b>аудиторные занятия (контактная работа):</b>	<b>57</b>
- занятия лекционного типа	28
- занятия семинарского типа	28
- текущий контроль (КСР)	1
<b>самостоятельная работа</b>	<b>15</b>
<b>Промежуточная аттестация – зачет</b>	

#### 3.2. Содержание дисциплины

Наименование и краткое содержание разделов и тем дисциплины	Всего (часы)	В том числе				
		Контактная работа (работа во взаимодействии с преподавателем), часы. Из них				Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа	Занятия лабораторного типа	Всего	
Дальнейшие методы статистического анализа. – Моделирование данных. Линейная регрессия. Принцип максимального правдоподобия. Отбор модели. Логистическая регрессия и отношение вероятностей. Скрытые марковские модели в контексте генетического анализа. Классификация. Скрытое распределение Дирихле (LDA). Ближайший центроид. Метод k ближайших соседей (kNN).	21	8	8		16	5
Построение статистических моделей. – Искусственная нейронная сеть. Метод опорных векторов (SVM). Уменьшение размерности. Кросс-валидация. Оценка эффективности классификатора. Точность.	21	8	8		16	5

Чувствительность. Специфичность. Коэффициент корреляции Мэтьюса. Персептрон. Многоуровневые нейронные сети. Прямое и обратное распространение ошибки. Метод главных компонент (РСА). Сравнение с методом скрытого распределения Дирихле. РСА vs LDA. Обучение без учителя. Метод k-средних, иерархический кластерный анализ. Алгоритм ближайших соседей и другие алгоритмы кластерного анализа.						
Статистический анализ в среде R. - Обзор функционала программного пакета R. Регрессионный анализ в среде R: теория и примеры. Тест ANOVA в среде R: теория и примеры. Логистическая регрессия в среде R: теория и примеры.	29	12	12		24	5
Текущий контроль (КСР)	1				1	
Промежуточная аттестация – зачет						
<b>Итого</b>	<b>72</b>	<b>28</b>	<b>28</b>		<b>57</b>	<b>15</b>

Текущий контроль успеваемости реализуется в формах опросов на занятиях семинарского типа.

Промежуточная аттестация проходит в традиционной форме (зачет).

#### **4. Учебно-методическое обеспечение самостоятельной работы обучающихся**

##### **а. Виды самостоятельной работы студентов**

- проработка теоретического материала лекционных занятий;
- подготовка к практическим занятиям;
- подготовка к промежуточной аттестации.

##### **Проработка теоретического материала лекционных занятий**

Выполняется самостоятельно обучающимся с использованием материалов лекций и рекомендованной основной и дополнительной литературы. Контроль проводится в часы контроля самостоятельной работы студентов.

##### **Подготовка к практическим занятиям**

Выполняется самостоятельно обучающимся с использованием лекционных материалов. Контроль выполняется на практических занятиях.

##### **Подготовка к промежуточной аттестации**

Фактором успешного прохождения промежуточной аттестации является систематическая работа студента в течение всего периода изучения дисциплины. В этом

случае подготовка к зачету будет концентрированной систематизацией всех полученных знаний, умений и навыков.

В качестве методических материалов при подготовке к зачету рекомендуется использовать собственные конспекты лекций и рекомендованную основную и дополнительную литературу.

**Выполнение практических заданий на следующие темы:**

1. Моделирование данных. Линейная регрессия. Принцип максимального правдоподобия. Отбор модели.
2. Логистическая регрессия и отношение вероятностей.
3. Скрытые марковские модели в контексте генетического анализа. Классификация
4. Скрытое распределение Дирихле (LDA). Ближайший центроид. Метод k ближайших соседей (kNN).
5. Искусственная нейронная сеть. Метод опорных векторов (SVM). Уменьшение размерности. Кросс-валидация.
6. Оценка эффективности классификатора. Точность. Чувствительность. Специфичность. Коэффициент корреляции Мэтьюса.
7. Персептрон. Многоуровневые нейронные сети. Прямое и обратное распространение ошибки.
8. Метод главных компонент (PCA). Сравнение с методом скрытого распределения Дирихле.
9. PCA vs LDA. Обучение без учителя.
10. Метод k-средних, иерархический кластерный анализ.
11. Алгоритм ближайших соседей и другие алгоритмы кластерного анализа.

**в. Образовательные материалы для самостоятельной работы студентов**

1. Кремер Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов.– М.: ЮНИТИ – ДАНА, 2006.-573 с. (10 экз)
2. Программный пакет R <http://cran.r-project.org/>.

Контрольные вопросы и практические задания для проведения текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведены в п. 5.2.

## 5. Фонд оценочных средств для промежуточной аттестации по дисциплине (модулю), включающий:

### 5.1. Описание шкал оценивания результатов обучения по дисциплине

Уровень сформированности компетенций (индикатора достижения компетенций)	Шкала оценивания сформированности компетенций						
	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	Не зачтено		Зачтено				
<u>Знания</u>	Отсутствие знаний теоретического материала.  Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки.	Минимально допустимый уровень знаний. Допущено много негрубых ошибок.	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки, без ошибок.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения.  Имели место грубые ошибки.	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме.	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме.	Продemonстрированы все основные умения, решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие владения материалом. Невозможность оценить наличие навыков вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы базовые навыки.  Имели место грубые ошибки.	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами.	Продemonстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов.	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов.	Продemonстрирован творческий подход к решению нестандартных задач.

### Шкала оценки при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	Превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно»
	Отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично», при этом хотя бы одна компетенция сформирована на уровне «отлично»
	Очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо», при этом хотя бы одна компетенция сформирована на уровне «очень хорошо»
	Хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо», при этом хотя бы одна компетенция сформирована на уровне «хорошо»
	Удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	Неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно», ни одна из компетенций не сформирована на уровне «плохо»
	Плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

### Критерии оценивания выполнения практического задания (каждое задание оценивается в 1 балл)

Практическое задание выполнено в полном объеме, отчет правильно и аккуратно оформлен	1	Превосходно	Зачтено
		Отлично	
Практическое задание выполнен в полном объеме, но отчет не аккуратно оформлен	0,75	Очень хорошо	
		Хорошо	
Практическое задание выполнено в полном объеме, но не достаточно самостоятельно, отчет оформлен	0,5	Удовлетворительно	Не зачтено
Практическое задание не выполнено	0	Неудовлетворительно	

## 5.2. Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения

### 5.2.1 Контрольные вопросы

вопросы	Код формируемой компетенции
1. Моделирование данных. Линейная регрессия. Принцип максимального правдоподобия. Отбор модели.	ПК-6
2. Логистическая регрессия и отношение вероятностей.	ПК-6

3. Скрытые марковские модели в контексте генетического анализа. Классификация.	ПК-6
4. Скрытое распределение Дирихле (LDA). Ближайший центроид. Метод k ближайших соседей (kNN).	ПК-6
5. Искусственная нейронная сеть. Метод опорных векторов (SVM). Уменьшение размерности. Кросс-валидация.	ПК-6
6. Оценка эффективности классификатора. Точность. Чувствительность. Специфичность. Коэффициент корреляции Мэтьюса.	ПК-6
7. Персептрон. Многоуровневые нейронные сети. Прямое и обратное распространение ошибки.	ПК-6
8. Метод главных компонент (PCA). Сравнение с методом скрытого распределения Дирихле.	ПК-6
9. PCA vs LDA. Обучение без учителя.	ПК-6
10. Метод k-средних, иерархический кластерный анализ.	ПК-6
11. Алгоритм ближайших соседей и другие алгоритмы кластерного анализа.	ПК-6

### 5.2.2. Типовые практические задания для оценки сформированности компетенции ПК-6

#### Задание 1. Регрессионный анализ в среде R.

Дано: Файл данных с двумя колонками значений.

Требуется: Посчитайте линейную регрессию переменной  $Y$  как функции от переменной  $X$ , и найдите два коэффициента регрессии:  $a$  и  $b$ , где  $b$  – это наклон прямой. Вычислите также  $r$ -значение для коэффициента “ $b$ ”.

#### Задание 2. Тест ANOVA в среде R.

Дано: файл данных с двумя колонками значений.

Требуется: Используя тест ANOVA, протестируйте взаимодействие переменной  $Y$  и  $X$ . Соответствующее  $p$ -значение напишите справа.

#### Задание 3. Логистическая регрессия в среде R.

Дано: файл данных с тремя колонками,  $X$ ,  $A$  (переменные) и  $Y$  (результат).

Требуется: Используя логистическую регрессию, найдите отношение шансов для того, чтобы результат  $Y$  был равен 1, при условии, что переменная  $X$  находится в своем пятом квантиле. Откорректируйте результат, учитывая переменную  $A$ . Вычислите четыре величины: отношение шансов, две границы доверительного интервала для отношения шансов и  $p$ -значение.

## 6. Учебно-методическое и информационное обеспечение дисциплины

а) основная литература:

1. Буре, В.М. Теория вероятностей и математическая статистика. [Электронный ресурс] / В.М. Буре, Е.М. Парилина. — Электрон. дан. — СПб.: Лань, 2013. — 416 с. Режим доступа: [https://e.lanbook.com/book/10249?category\\_pk=913#book\\_name](https://e.lanbook.com/book/10249?category_pk=913#book_name)



2. Пугачев, В.С. Теория вероятностей и математическая статистика. [Электронный ресурс] — Электрон. дан. — М.: Физматлит, 2002. — 496 с. — Режим доступа: [https://e.lanbook.com/book/48170?category\\_pk=913#book\\_name](https://e.lanbook.com/book/48170?category_pk=913#book_name)

б) дополнительная литература:

1. Болотюк, В.А. Практикум и индивидуальные задания по курсу теории вероятностей (типовые расчеты). [Электронный ресурс] / В.А. Болотюк, Л.А. Болотюк, А.Г. Гринь, И.П. Гринь. — Электрон. дан. — СПб.: Лань, 2010. — 288 с. — Режим доступа: [https://e.lanbook.com/book/534?category\\_pk=913#book\\_name](https://e.lanbook.com/book/534?category_pk=913#book_name)

2. Ибрагимов, И.М. Основы компьютерного моделирования наносистем. [Электронный ресурс] / И.М. Ибрагимов, А.Н. Ковшов, Ю.Ф. Назаров. — Электрон. дан. — СПб.: Лань, 2010. — 384 с. —

Режим доступа: [https://e.lanbook.com/book/156?category\\_pk=915#authors](https://e.lanbook.com/book/156?category_pk=915#authors)

в) программное обеспечение и Интернет-ресурсы

1. Программный пакет R <http://cran.r-project.org/> (в свободном доступе), лицензия: <https://www.r-project.org/COPYING>.

## **7. Материально-техническое обеспечение дисциплины**

Помещения представляют собой учебные аудитории для проведения учебных занятий, предусмотренных программой, оснащенные оборудованием и техническими средствами обучения: компьютерный класс, проектор, экран.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду ННГУ.

Лицензионное и свободно распространяемое ПО:

- операционные системы семейства Microsoft Windows, – лицензия по подписке Microsoft Imagine;
- программный пакет R, – <http://cran.r-project.org/> (в свободном доступе), лицензия: <https://www.r-project.org/COPYING>.

Программа составлена в соответствии с требованиями ОС ННГУ 01.03.02 Прикладная математика и информатика.

Автор: к.ф.-м.н., проф. каф. прикладной математики Заикин А.А.

Рецензент: д.т.н., профессор НГТУ им. Р.Е. Алексеева Ломакина Л.С.

Зав. каф. прикладной математики: д.ф.-м.н. Иванченко М.В.

Программа одобрена на заседании методической комиссии института информационных технологий, математики и механики от 1 декабря 2021 года, протокол № 2.