

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

---

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

**Рабочая программа дисциплины**

Математические методы биоинформатики. Дополнительные главы

---

Уровень высшего образования

Бакалавриат

---

Направление подготовки / специальность

01.03.02 - Прикладная математика и информатика

---

Направленность образовательной программы

Прикладная математика и информатика (общий профиль)

---

Форма обучения

очная

---

г. Нижний Новгород

2024 год начала подготовки

## 1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.ДВ.12.05 Математические методы биоинформатики. Дополнительные главы относятся к части, формируемой участниками образовательных отношений образовательной программы.

## 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	Для текущего контроля успеваемости	Для промежуточной аттестации
ПК-13: Способен участвовать в исследовании математических моделей в естественных науках и технике	ПК-13.1: Знает методы создания, анализа и исследования математических моделей в естественных науках и технике ПК-13.2: Знает математические методы обработки информации, полученной в результате экспериментальных исследований ПК-13.3: Умеет корректно использовать методы создания, анализа и исследования математических моделей, умеет применять численные и аналитические методы решения базовых математических задач и классических задач естествознания в практической деятельности ПК-13.4: Владеет навыками использования математических методов обработки информации, полученной в результате экспериментальных исследований	ПК-13.1: Знает базовые математические методы исследования для решения прикладных задач биоинформатики  ПК-13.2: Умеет определять и профессионально реализовывать необходимые для решения прикладных задач биоинформатики математические методы  ПК-13.3: Умеет применять для исследования пакеты прикладных программ, обрабатывать и анализировать полученные результаты.  ПК-13.4: Имеет практический опыт применения современных пакетов прикладных программ для решения прикладных задач биоинформатики	Задачи Собеседование	Зачёт: Контрольные вопросы

## 3. Структура и содержание дисциплины

### 3.1 Трудоемкость дисциплины

	<b>очная</b>
<b>Общая трудоемкость, з.е.</b>	<b>2</b>
<b>Часов по учебному плану</b>	<b>72</b>
в том числе	
<b>аудиторные занятия (контактная работа):</b>	
- занятия лекционного типа	<b>24</b>
- занятия семинарского типа (практические занятия / лабораторные работы)	<b>24</b>
- КСР	<b>1</b>
<b>самостоятельная работа</b>	<b>23</b>
<b>Промежуточная аттестация</b>	<b>0</b> <b>Зачёт</b>

### 3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование 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 (практические занятия/лабораторные работы), часы	Всего	
	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0
Дальнейшие методы статистического анализа.	21	7	7	14	7
Построение статистических моделей.	21	7	7	14	7
Статистический анализ в среде R.	29	10	10	20	9
Аттестация	0				
КСР	1			1	
Итого	72	24	24	49	23

#### Содержание разделов и тем дисциплины

1. 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).
2. Построение статистических моделей. – Искусственная нейронная сеть. Метод опорных векторов (SVM). Уменьшение размерности. Кросс-валидация. Оценка эффективности классификатора. Точность. Чувствительность. Специфичность. Коэффициент корреляции Мэтьюса. Персептрон. Многоуровневые нейронные сети. Прямое и обратное распространение ошибки. Метод главных компонент (PCA). Сравнение с методом скрытого распределения Дирихле. PCA vs LDA. Обучение без учителя. Метод k-

средних, иерархический кластерный анализ. Алгоритм ближайших соседей и другие алгоритмы кластерного анализа.

3. 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.

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

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

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/> .

#### **5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)**

**5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:**

**5.1.1 Типовые задания (оценочное средство - Задачи) для оценки сформированности компетенции ПК-13:**

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  $p$ -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.

### Критерии оценивания (оценочное средство - Задачи)

Оценка	Критерии оценивания
зачтено	Выполнены все или большая часть этапов решения задачи или задача решена с незначительными недочетами. Результаты работы представлены преподавателю в срок.
не зачтено	Выполнены не все практические задания или выполнены не в полном объеме (представлено не полное описание этапов выполнения заданий, получен неверный ответ, результаты работы не представлены преподавателю).

### 5.1.2 Типовые задания (оценочное средство - Собеседование) для оценки сформированности компетенции ПК-13:

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.

### Критерии оценивания (оценочное средство - Собеседование)

Оценка	Критерии оценивания
зачтено	Студент дал развернутый ответ на все вопросы без существенных ошибок.
не зачтено	При ответе студент допускает грубые ошибки в основном материале.

## 5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

### Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатора достижения 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. Ошибок нет.	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 базовых навыков. 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 оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «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».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

### 5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

#### 5.3.1 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-13

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.

### Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Студент ответил на большую часть вопросов возможно с незначительными недочетами.
не зачтено	При ответе студент допускает грубые ошибки в основном материале и решении стандартных задач.

### 6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Буре В. М. Теория вероятностей и математическая статистика / Буре В. М.,Парилина Е. М. - Санкт-Петербург : Лань, 2022. - 416 с. - Допущено УМО по классическому университетскому образованию в качестве учебника для студентов вузов, обучающихся по направлениям ВПО 010400 — «Прикладная математика и информатика» и 010300 — «Фундаментальная информатика и информационные технологии». - Книга из коллекции Лань - Математика. - ISBN 978-5-8114-1508-3., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=799898&idb=0>.
2. Пугачев В. С. Теория вероятностей и математическая статистика / Пугачев В. С. - 2-е изд. - Москва : ФИЗМАТЛИТ, 2002. - 496 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции ФИЗМАТЛИТ - Математика. - ISBN 5-9221-0254-0., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=695911&idb=0>.

Дополнительная литература:

1. Практикум и индивидуальные задания по курсу теории вероятностей ( типовые расчеты) : учебное пособие для вузов / Болотюк В. А.,Болотюк Л. А.,Гринь А. Г.,Гринь И. П.,Окишев С. В.,Оранская Л. А.,Филимонова Т. А.,Швед Е. А.; Болотюк В. А., Гринь А. Г., Гринь И. П., Окишев С. В., Оранская Л. А., Филимонова Т. А., Швед Е. А. - 2-е изд., испр. - Санкт-Петербург : Лань, 2023. - 288 с. - Допущено НМС по математике Министерства образования и науки РФ в качестве учебного пособия для студентов вузов, обучающихся по специальностям «Мировая экономика», «Финансы и кредит», «Промышленная теплоэнергетика», «Локомотивы», «Стандартизация и сертификация», «Вычислительные машины, комплексы, системы и сети». - Книга из коллекции Лань - Математика. - ISBN 978-5-507-47289-5., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=883780&idb=0>.
2. Ибрагимов И. М. Основы компьютерного моделирования наносистем / Ибрагимов И. М.,Ковшов А. Н.,Назаров Ю. Ф. - Санкт-Петербург : Лань, 2022. - 384 с. - Книга из коллекции Лань - Математика. - ISBN 978-5-8114-1032-3., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=799614&idb=0>.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

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

## **7. Материально-техническое обеспечение дисциплины (модуля)**

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения, компьютерами, специализированным оборудованием: 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>.

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

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки 01.03.02 - Прикладная математика и информатика.

Автор(ы): Заикин Алексей Анатольевич, кандидат физико-математических наук.

Заведующий кафедрой: Иванченко Михаил Васильевич, доктор физико-математических наук.

Программа одобрена на заседании методической комиссии от 13.12.2023, протокол № 3.