

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

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

Институт биологии и биомедицины
(факультет / институт / филиал)

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

Рабочая программа дисциплины

Научное программирование

(наименование дисциплины (модуля))

Уровень высшего образования

Специалитет

Направление подготовки / специальность

30.05.02 Медицинская биофизика

Направленность образовательной программы

(указывается профиль / магистерская программа / специализация)

Форма обучения

Очная

Нижний Новгород
2022 год

1. Место и цели дисциплины в структуре ОПОП

Дисциплина «Научное программирование» относится к обязательной части Блока 1 ООП по направлению подготовки 30.05.02 "Медицинская биофизика.

Студенты имеют навыки разработки программ и программных модулей на языке программирования Python, а также навыки применения методов математической статистики в различных биологических исследованиях.

Целями освоения дисциплины «Научное программирование» являются: Формирование теоретических знаний по научному программированию и основам машинного обучения на языке Python

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников)

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	
ОПК-6. Способен понимать принципы работы информационных технологий, обеспечивать информационно-технологическую поддержку в области здравоохранения; применять средства информационно-коммуникационных технологий и ресурсы биоинформатики в профессиональной деятельности; выполнять требования информационной безопасности	ОПК-6.1. Знает биоинформационные технологии. ОПК-6.2. Умеет применять информационные технологии в области здравоохранения, в профессиональной деятельности. ОПК-6.3. Владеет требованиями информационной безопасности.	Знать основные способы визуализации научных данных; основы математической статистики; специфику машинного обучения; основные типы задач в машинном обучении. Уметь обрабатывать, визуализировать научные данные; применять на практике технологии Numpy, Pandas, Matplotlib, Scipy, scikit-learn для разработки программ для решения задач научной деятельности. Владеть навыками решения задач с Помощью языка программирования Python.	Контрольные вопросы, тесты

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная форма обучения
Общая трудоемкость	7 ЗЕТ
Часов по учебному плану	252
в том числе	
аудиторные занятия (контактная работа):	144
- занятия лекционного типа	64

- занятия практического типа	80
самостоятельная работа	69
КСР	36
Промежуточная аттестация –зачет, экзамен	

3.2. Содержание дисциплины

Наименование и краткое содержание разделов и тем дисциплины (модуля)	Все-го (часы)	В том числе					
		Контактная работа (работа во взаимодействии с преподавателем), часы					Самостоятельная работа обучающегося, часы
		из них					
		Занятия лекционного типа	Занятия семинарского типа	Занятия лабораторного типа		Все-го	
Введение							
Различные способы визуализации научных данных с использованием Matplotlib.							
Описательные статистики							
Построение диаграмм размаха							
Построение гистограмм							
Функции плотности вероятности							
Построение планок погрешностей							
Корреляционный анализ							
Проверка статистических гипотез.							
Решение дифференциальных уравнений с помощью библиотек SciPy и NumPy.							
Решение систем линейных уравнений с помощью библиотек SciPy и NumPy.							
Методы Монте-Карло. Численное моделирование динамики процесса Морана нейтрального дрейфа.							
Итого	71	32	32			74	7

Текущий контроль успеваемости реализуется в рамках практических занятий и в виде проведения устных опросов. Промежуточный контроль осуществляется при проведении экзамена.

4. Учебно-методическое обеспечение самостоятельной работы обучающихся

Виды самостоятельной работы студентов в рамках освоения дисциплины:

- изучение понятийного аппарата и проработка тем дисциплины;
- работа с основной и дополнительной литературой дома и в библиотеке;
- изучение сайтов по темам дисциплины в сети Интернет
- подготовка к устному опросу на семинарских занятиях;
- подготовка к тестам;
- подготовка к зачету и экзамену.

Методические указания по подготовке студентов к текущему и промежуточному контролю по дисциплине «Научное программирование»

Подготовка к устному опросу, тестированию

Все перечисленные виды самостоятельной работы представляют собой систему заданий, позволяющих оценить уровень знаний по основным разделам, темам, проблемам дисциплины, а также умений обучающегося синтезировать материал предшествующих дисциплин. При подготовке к ним студенту необходимо:

- 1) ознакомиться с соответствующей темой программы изучаемой дисциплины;
- 2) изучить рекомендованную учебно-методическую литературу по данной теме;
- 3) тщательно изучить лекционный материал;
- 4) повторить материалы предшествующих дисциплин.

Подготовка к зачету и экзамену

Промежуточная аттестация по итогам освоения дисциплины проходит в форме **зачета и экзамена**. Подготовка к зачету и экзамену является концентрированной систематизацией всех полученных знаний по дисциплине «Научное программирование».

В начале семестра рекомендуется внимательно изучить перечень вопросов к зачету и экзамену по данной дисциплине, а также использовать в процессе обучения программу, другие методические материалы, разработанные кафедрой по данной дисциплине. Это позволит в процессе изучения тем сформировать более правильное и обобщенное видение студентом сущности того или иного вопроса за счет:

- а) уточняющих вопросов преподавателю;
- б) подготовки докладов по отдельным темам;
- в) самостоятельного уточнения вопросов на смежных дисциплинах;
- г) углубленного изучения вопросов темы по учебным пособиям.

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

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 незначительных 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 небольшими 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 оце-	При решении стандартных задач не продемонстрированы базовые	Имеется минимальный набор навыков для решения стан-	Продemonстрированы базовые навыки при решении	Продemonстрированы базовые навыки при решении стандарт-	Продemonстрированы навыки при решении нестандарт-	Продemonстрирован творческий подход к решению

	нить наличие навыков вследствие отказа обучающегося от ответа	вые навыки. Имели место грубые ошибки.	дартных задач с некоторыми недочетами	стандартных задач с некоторыми недочетами	ных задач без ошибок и недочетов.	ных задач без ошибок и недочетов.	нестандартных задач
--	---	--	---------------------------------------	---	-----------------------------------	-----------------------------------	---------------------

Шкала оценки при промежуточной аттестации

Оценка		Уровень подготовки
	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
зачтено	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»

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

5.2.1. Примеры вопросов к устным опросам для оценки знаний компетенции

1. Какие основные типы графиков используются для визуализации научных данных?
2. Какие типы переменных встречаются при анализе научных данных?
3. Какая выборка называется репрезентативной?
4. Что такое статистические выбросы?
5. Какие существуют основные описательные статистики?
6. Что представляет собой диаграмма размаха («ящик с усами», boxplot)?
7. Что такое гистограмма и функция плотности вероятности?
8. Для чего на графиках строятся планки погрешностей (error bar)?
9. Что такое корреляция и какие типы корреляции бывают?
10. Какой график используется для графического представления корреляционной связи?
11. Что такое ковариация?
12. В чём заключается разница между коэффициентами корреляции Пирсона и Спирмена?
13. Какие критерии нормальности вам известны?
14. В чём заключается суть дисперсионного анализа (ANOVA)?
15. Что такое FDR-коррекция?
16. Какие задачи машинного обучения являются задачами обучения с учителем?

17. Что такое нормализация данных?
18. Что такое переобучение?
19. В чём заключается суть метода k ближайших соседей?
20. Для чего используются обучающие и тестовые выборки?
21. В чём заключается суть метода перекрёстного контроля?
22. В чём заключается суть метода главных компонент?
23. Какие существуют эвристические подходы для выбора количества главных компонент?
24. Для чего используется метод наименьших квадратов?
25. Что такое коэффициент детерминации?
26. Какие существуют причины переобучения в задаче восстановления регрессии?
27. Какие предположения задаются в линейном дискриминантном анализе?
28. Что такое логистическая регрессия?
29. Что такое оптимальная разделяющая гиперплоскость?
30. В чём заключается суть использования дерева решений?
31. Какие существуют алгоритмы построения деревьев решений и в чём их отличие?
32. Что такое энтропия?
33. В чём заключается суть баггинга?
34. Чем отличаются экстремально случайные деревья от случайного леса?
35. В чём заключается суть бустинга?
36. Что такое нейронная сеть?
37. Что такое метод обратного распространения ошибки?
38. Что такое глубокое обучение?
39. Какая задача является задачей обучения без учителя?
40. В чём заключается суть метода центров тяжести?
41. Чем метод медоидов отличается от метода центров тяжести?
42. Что такое иерархическая кластеризация?

5.2.2. Примеры вопросов к экзамену

1. Способы визуализации научных данных с использованием Matplotlib. Иерархическая структура рисунка в Matplotlib. Основные элементы графика. Основные типы графиков.
2. Основы статистики в Python. Описательные статистики. Построение диаграмм размаха, гистограмм, функций плотности вероятности, планок погрешностей.
3. Основы статистики в Python. Корреляционный анализ. Проверка статистических гипотез. FDR-коррекция.
4. Численное решение дифференциальных уравнений, решение систем линейных уравнений с помощью библиотек SciPy и NumPy. Методы Монте-Карло.
5. Постановка задачи машинного обучения. Основные классы задач в машинном обучении. Примеры практических задач.
6. Вероятностная постановка задачи обучения с учителем. Функция потерь. Средний риск. Эмпирический риск.
7. Экспериментальная оценка качества обучения и выбор параметров модели. Обучающая, проверочная и тестовая выборки. Метод перекрестного (скользящего) контроля. Переобучение и недообучение.
8. Наивный байесовский классификатор. Сглаживание Лапласа. Ошибки 1-го и 2-го рода. Чувствительность, специфичность, точность, полнота. ROC-кривая. Использование наивного байесовского классификатора для количественных признаков.
9. Метод k ближайших соседей в задачах классификации и восстановления регрессии.
10. Методы предобработки данных. Методы понижения размерности. Метод главных компонент. Сингулярное разложение.
11. Линейная регрессия. Метод наименьших квадратов. Проверка статистической значимости модели. Коэффициент детерминации.
12. Методы борьбы с переобучением в задаче восстановления регрессии. Отбор признаков. Регуляризация.
13. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ.
14. Логистическая регрессия. Логистическая функция и softmax.
15. Машина опорных векторов. Оптимальная разделяющая гиперплоскость. Опорные векторы. Случаи линейно-разделимых и неразделимых классов. Ядра и спрямляющие пространства.

16. Деревья решений. Алгоритм CART.
17. Ансамбли решающих правил. Баггинг. Случайный лес. Экстремально случайные деревья.
18. Ансамбли решающих правил. Бустинг. AdaBoost. Градиентный бустинг деревьев решений.
19. Нейронные сети. Персептрон Розенблатта. Сети прямого распространения. Алгоритм обратного распространения ошибки.
20. Глубокое обучение. Сверточные нейронные сети.
21. Обучение без учителя. Задача кластеризации. Метод центров тяжести. Метод медоидов.
22. Алгоритмы кластеризации: EM-алгоритм, алгоритм DBSCAN. Алгоритмы иерархической кластеризации.

5.2.3. Примеры практических заданий

Загрузите из файла heart.csv (<https://www.kaggle.com/ronitf/heart-disease-uci?select=heart.csv>) данные о сердечных заболеваниях. Выполните следующие задания:

- 1) Сколько образцов (объектов) содержит данный датасет?
- 2) Сколько атрибутов (признаков) содержит данный датасет? Подробно опишите значение каждого признака.
- 3) Опишите тип каждого признака (числовой / дискретный / непрерывный / категориальный / номинальный / бинарный / ординальный)?
- 4) Вычислите, сколько мужчин/женщин в датасете?
- 5) Вычислите описательные статистики для количественных признаков (среднее значение, медиана, мода, размах, дисперсия, среднее квадратичное отклонение, 1й/2й/3й квартили, межквартильный размах).
- 6) Постройте гистограммы для признаков age, trestbps, chol, thalach, oldpeak. Расположите гистограммы на одном графике в одну линию. Подпишите оси каждой гистограммы.
- 7) Постройте диаграммы размаха для признаков age, trestbps, chol, thalach, oldpeak.
- 8) Постройте на одном графике две кривые PDF (probability density function) для признака chol. Одна PDF для мужчин, другая - для женщин. Подпишите оси, добавьте легенду.
- 9) Для признаков, которые не были указаны в п.6-7, постройте полигоны частот (подграфики расположите в нескольких строках и столбцах). Подпишите оси.
- 10) Сгруппируйте данные по полу и вычислите для каждой группы среднее значение признака chol, применив функцию агрегации. Изобразите результаты в виде столбчатой диаграммы, где столбцы должны соответствовать полу, а высота столбцов - соответствующим средним значениям признака chol. Добавьте к каждому столбцу планку погрешности, отражающую среднее квадратичное отклонение.
- 11) Постройте следующие диаграммы рассеяния:
 - trestbps от age
 - chol от age
 - thalach от age
 - oldpeak от age

Изобразите точки на диаграммах разными цветами в зависимости от пола. Подпишите оси и добавьте легенду. Попробуйте визуально определить, коррелируют ли рассматриваемые переменные с возрастом. Проверьте свои предположения, вычислив коэффициенты корреляции Спирмена. Сделайте выводы.

- 12) Проверьте признаки age, trestbps, chol, thalach, oldpeak на нормальность с помощью критерия Шапиро-Уилка.

Загрузите датасет ISOLET (<https://archive.ics.uci.edu/ml/datasets/ISOLET>). Выполните следующие задания:

- 1) Опишите рассматриваемый датасет.
- 2) Выполните нормализацию признаков.
- 3) Разбейте данные на обучающую и тестовую выборку.
- 4) Вызовите метод ближайших соседей. Постройте графики зависимости ошибки этого метода на обучающей и тестовой выборках от . Сделайте вывод.

- 5) Выполните процедуру перекрестного контроля (5-fold, 10-fold, LOO) с обучающей выборкой. Постройте графики зависимости CV-ошибки от числа используемых соседей в методе ближайших соседей. Выберите наилучшую модель и проверьте ее качество на тестовой выборке.
- 6) Примените к рассматриваемым данным
 - Линейный дискриминантный анализ
 - Квадратичный дискриминантный анализ
 - Логистическую регрессиюДля каждого метода вычислите ошибки на обучающей и тестовой выборках.
- 7) На рассматриваемых данных обучите следующие классификаторы:
 - Random Forest
 - Extremely Random TreesПостройте графики зависимости ошибки на обучающей и тестовой выборке от количества используемых деревьев.
- 8) Натренируйте на рассматриваемых данных нейронную сеть с одним скрытым слоем ("vanilla") с 200 нейронами в нем (`hidden_layer_sizes = (200,)`). В качестве функции активации используйте положительную срезку (`activation = 'relu'`). Постройте графики зависимости ошибки на обучающей и тестовой выборке от параметра `alpha` (`weight decay`).

6. Методические материалы, определяющие процедуры оценивания

Положение «О проведении текущего контроля успеваемости и промежуточной аттестации обучающихся в ННГУ», утвержденное приказом ректора ННГУ от 13.02.2014 г. №55-ОД.
Положение о фонде оценочных средств, утвержденное приказом ректора ННГУ от 10.06.2015 №247-ОД.

6.1. Учебно-методическое и информационное обеспечение дисциплины (модуля)

а) основная литература:

1. Langtangen H.P. A Primer on Scientific Programming with Python. - Springer, 2009. — ISBN: 978-3-642-02474-0.
2. Tosi S. Matplotlib for Python Developers. — Packt Publishing, 2009. — ISBN: 978-1-847197-90-0.
3. Johansson R. Numerical Python: Scientific Computing and Data Science Applications with Numpy, SciPy and Matplotlib. – Apress, 2018. - ISBN: 9781484242469.
4. Мхитарян В.С. Анализ данных: учебник для академического бакалавриата. – М.: Юрайт, 2018. – 490 с. – ISBN: 978-5-534-00616-2.
5. Маккинли У. Python и анализ данных. – М.: ДМК Пресс, 2015. - 481 с. – ISBN: 978-5-9706031-5-4.
6. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. – Из-во Вильямс, 2017. - 480 с. - ISBN: 978-5-9908910-8-1.

б) дополнительная литература:

1. Доусон М. Програмируем на Python. – 3-е изд. – СПб: Питер, 2014. – 416 с. – ISBN 978-1435455009.
2. Плас Дж.В. Python для сложных задач: наука о данных и машинное обучение. - СПб.: Питер, 2018. — 576 с.: ил. — ISBN: 978-5-496-03068-7.
3. Неделько В.М. Основы статистических методов машинного обучения. - Новосибирск: Изд-во НГТУ, 2010. – 72 с. - ISBN 978-5-7782-1385-2.

в) Интернет ресурсы

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. Режим доступа: <http://www.machinelearning.ru>.

2. Документация библиотеки Matplotlib. Режим доступа: <https://matplotlib.org/stable/index.html>.
3. Документация для основных проектов библиотеки SciPy. Режим доступа: <https://scipy.org/docs.html>.
4. Документация библиотеки scikit-learn. Режим доступа: http://scikit-learn.org/stable/user_guide.html.

г) программное обеспечение

1. Anaconda3 (дистрибутив Python)
2. Jupyter Notebook (интерактивная оболочка Python)

7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения занятий лекционного и практического типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации, укомплектованы учебной мебелью. Для проведения занятий имеется демонстрационное оборудование (доска, мультимедийное оборудование (проектор, экран)). Практические занятия проводятся в аудитории, оборудованной персональными компьютерами с установленным лицензионным программным обеспечением: операционная система Windows 10, дистрибутив языка Python – Anaconda3, интерактивная оболочка Python – Jupyter Notebook.

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

Программа составлена в соответствии с требованиями ОС ННГУ.

Автор(ы) _____ м.н.с., О.С. Вершинина

Рецензент _____

Директор ИББМ _____ д.б.н., доц. М.В. Ведунова

Программа одобрена на заседании Методической комиссии Института биологии и биомедицины от 06.12.2021 года, протокол №3.