

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное  
образовательное учреждение высшего образования  
«Национальный исследовательский Нижегородский государственный  
университет им. Н.И. Лобачевского»

Институт биологии и биомедицины

---

УТВЕРЖДЕНО  
решением ученого совета ННГУ  
протокол от

«16» июня 2021 г. №8

**Рабочая программа дисциплины**

**Машинное обучение**

---

(наименование дисциплины (модуля))

Уровень высшего образования

**Специалитет**

Направление подготовки / специальность

**30.05.03 Медицинская кибернетика**

Направленность образовательной программы

---

(указывается профиль / магистерская программа / специализация)

Форма обучения

**Очная**

Нижегород  
2021 год

## 1. Место и цели дисциплины в структуре ОПОП

Дисциплина «Машинное обучение» относится к базовым дисциплинам цикла Б.1 (профессиональный) ОПОП по специальности 30.05.03 Медицинская кибернетика. Дисциплина обязательна для освоения в 7 семестре.

Студенты к моменту освоения дисциплины «Машинное обучение» демонстрируют базовые представления о принципах программирования, синтаксисе и конструкциях языка программирования Python, основных методах статистического анализа данных, полученные при изучении таких дисциплин специалитета как «Научное программирование».

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 и машинное обучение» являются:**

1. Формирование теоретических знаний по научному программированию и основам машинного обучения на языке Python;
2. Выработка умений и навыков по практическому применению методов обработки, статистического анализа и визуализации научных данных, методов машинного обучения при решении задач в различных прикладных областях;
3. Выработка умений и навыков использования различных библиотек языка программирования Python для анализа и визуализации данных, а так же для применения алгоритмов машинного обучения.

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

Формируемые компетенции  (код компетенции, уровень освоения – при наличии в карте компетенции)	Планируемые результаты обучения по дисциплине (модулю), характеризующие этапы формирования компетенций
<b>ПК-1</b> - способность творчески использовать в научной и производственно-технологической деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность (профиль) программы магистратуры ( <i>Этап формирования - базовый</i> )	<b>З (ПК-1)</b> – <i>знать</i> основные способы визуализации научных данных; основы математической статистики; специфику машинного обучения; основные типы задач в машинном обучении; некоторые методы решения задач классификации и восстановления регрессии (метод k ближайших соседей, линейная регрессия). <b>У (ПК-1)</b> - <i>уметь</i> обрабатывать, анализировать и визуализировать научные данные; применять на практике технологии Numpy, Pandas, Matplotlib, Scipy, scikit-learn для разработки программ для решения задач научной деятельности. <b>В (ПК-1)</b> – <i>владеть</i> навыками решения научных задач с помощью языка программирования Python.

<p><b>ПК-3</b> – способность применять методические основы проектирования, выполнения полевых и лабораторных биологических, экологических исследований, использовать современную аппаратуру и вычислительные комплексы (в соответствии с направленностью (профилем) программы магистратуры) (Этап формирования - начальный)</p>	<p><b>3 (ПК-1)</b> – <i>знать</i> методические основы проектирования моделей машинного обучения; основные алгоритмы решения задачи восстановления регрессии; основные алгоритмы решения задачи классификации; алгоритмы кластеризации; экспериментальные методы оценки качества обучения.</p> <p><b>У (ПК-1)</b> – <i>уметь</i> использовать имеющиеся знания для решения практических задач машинного обучения с помощью Python; оценивать на практике качество обучения используемых моделей.</p> <p><b>В (ПК-1)</b> – <i>владеть</i> навыками решения научных задач с использованием методов машинного обучения; навыками использования библиотеки scikit-learn для решения задач машинного обучения.</p>
---	--

### 3. Структура и содержание дисциплины (модуля)

Объем дисциплины (модуля) составляет 3 зачетных единицы, всего 108 часов, из которых 62 составляет контактная работа обучающегося с преподавателем (32 часа занятия лекционного типа, 48 часов практические занятия, 1 часа мероприятия промежуточной аттестации), 27 часов составляет самостоятельная работа обучающегося.

Наименование и краткое содержание разделов и тем дисциплины (модуля),  форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	в том числе			
		контактная работа (работа во взаимодействии с преподавателем), часы			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия практического типа	Всего	
<i>Тема 1.</i> Различные способы визуализации научных данных с использованием Matplotlib.	6	2	3	5	1
<i>Тема 2.</i> Основы статистики в Python. Описательная статистика. Построение диаграмм размаха, гистограмм, функций плотности вероятности, планок погрешностей.	9.5	3	6	9	0.5
<i>Тема 3.</i> Основы статистики в Python. Корреляционный анализ. Проверка статистических гипотез.	9.5	3	6	9	0.5

<i>Тема 4.</i> Численное решение дифференциальных уравнений, решение систем линейных уравнений с помощью библиотек SciPy и NumPy. Методы Монте-Карло. Численное моделирование динамики процесса Морана в модели нейтрального дрейфа.	8.5	3	3	6	2.5
<i>Тема 5.</i> Введение в машинное обучение. Постановка основных классов задач в машинном обучении: обучение с учителем, обучение без учителя. Примеры практических задач.	12.5	3	7	10	2.5
<i>Тема 6.</i> Вероятностная постановка задачи обучения с учителем. Наивный байесовский классификатор. Метод k ближайших соседей. Оценка качества обучения и выбор модели. Метод перекрестного контроля. Переобучение и недообучение.	13	3	7	10	3
<i>Тема 7.</i> Методы предобработки данных. Методы снижения размерности. Метод главных компонент.	9	3	3	6	3
<i>Тема 8.</i> Понятие регрессии. Линейная регрессия. Метод наименьших квадратов. Борьба с переобучением. Отбор признаков. Линейная регрессия для задачи классификации. Линейный и квадратичный дискриминантный анализ. Логистическая регрессия.	10	3	4	7	3
<i>Тема 9.</i> Машина опорных векторов. Ядра и спрямляющие пространства.	9	3	3	6	3
<i>Тема 10.</i> Деревья решений. Алгоритм CART. Ансамбли решающих правил. Бэггинг и случайный лес. Бустинг.	9	3	3	6	3
<i>Тема 11.</i> Нейронные сети. Персептрон Розенблатта. Алгоритм обратного распространения ошибки. Глубокое обучение.	9	3	3	6	3

Тема 12. Обучение без учителя. Задача кластеризации. Метод центров тяжести. Метод медоидов. Другие методы кластеризации: EM-алгоритм, DBSCAN, иерархическая кластеризация.	9	3	3	6	3
В т.ч. текущий контроль	2				
<b>Промежуточная аттестация – экзамен (36 часов)</b>					

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

#### 4. Образовательные технологии

В соответствии с рабочей программой и тематическим планом изучение дисциплины проходит в виде аудиторной и самостоятельной работы студентов. Учебный процесс в аудитории осуществляется в форме лекционных и практических занятий.

Для достижения планируемых результатов обучения в дисциплине «Научное программирование и машинное обучение» используются следующие образовательные технологии:

1. *Традиционные технологии:* информационные лекции и практические занятия в компьютерном классе.
2. *Информационно-коммуникационные технологии:* лекции-визуализации с презентацией изучаемого материала, беседы, различные формы самостоятельной работы студентов (самостоятельное изучение литературы, составление опорных конспектов, решение задач в области научного программирования и машинного обучения).
3. *Интерактивные технологии:* демонстрация процесса написания программ на примере решения конкретных задач и вовлечение обучающихся в процесс программирования.

#### 5. Учебно-методическое обеспечение самостоятельной работы обучающихся

Основные виды самостоятельной работы студентов в рамках освоения дисциплины:

- изучение понятийного аппарата дисциплины;
- проработка тем дисциплины;
- подготовка к практическим занятиям;
- подготовка к устным опросам;
- выполнение домашних заданий (задач по научному программированию и машинному обучению);
- работа с основной и дополнительной литературой;
- изучение материала по темам дисциплины в сети Интернет;
- подготовка к экзамену.

#### Распределение форм текущего контроля по темам дисциплины

Раздел / тема учебной дисциплины	Проверяемые знания, умения и владения	Форма текущего контроля и оценивания самостоятельной работы студентов
Тема 1. Различные способы визуализации научных данных с использованием Matplotlib.	ПК-1	Устный опрос Практические задания

<i>Тема 2. Основы статистики в Python. Описательная статистика. Построение диаграмм размаха, гистограмм, функций плотности вероятности, планок погрешностей.</i>	ПК-1	Устный опрос Практические задания
<i>Тема 3. Основы статистики в Python. Корреляционный анализ. Проверка статистических гипотез.</i>	ПК-1	Устный опрос Практические задания
<i>Тема 4. Численное решение дифференциальных уравнений, решение систем линейных уравнений с помощью библиотек SciPy и NumPy. Методы Монте-Карло. Численное моделирование динамики процесса Морана в модели нейтрального дрейфа.</i>	ПК-1	Устный опрос Практические задания
<i>Тема 5. Введение в машинное обучение. Постановка основных классов задач в машинном обучении: обучение с учителем, обучение без учителя. Примеры практических задач.</i>	ПК-1, ПК-3	Устный опрос Практические задания
<i>Тема 6. Вероятностная постановка задачи обучения с учителем. Наивный байесовский классификатор. Метод k ближайших соседей. Оценка качества обучения и выбор модели. Метод перекрестного контроля. Переобучение и недообучение.</i>	ПК-1, ПК-3	Устный опрос Практические задания
<i>Тема 7. Методы предобработки данных. Методы снижения размерности. Метод главных компонент.</i>	ПК-1, ПК-3	Устный опрос Практические задания
<i>Тема 8. Понятие регрессии. Линейная регрессия. Метод наименьших квадратов. Борьба с переобучением. Отбор признаков. Линейная регрессия для задачи классификации. Линейный и квадратичный дискриминантный анализ. Логистическая регрессия.</i>	ПК-1, ПК-3	Устный опрос Практические задания
<i>Тема 9. Машина опорных векторов. Ядра и спрямляющие пространства.</i>	ПК-3	Устный опрос Практические задания
<i>Тема 10. Деревья решений. Алгоритм CART. Ансамбли решающих правил. Бэггинг и случайный лес. Бустинг.</i>	ПК-3	Устный опрос Практические задания
<i>Тема 11. Нейронные сети. Персептрон Розенблатта. Алгоритм обратного распространения ошибки. Глубокое обучение.</i>	ПК-3	Устный опрос Практические задания

Тема 12. Обучение без учителя. Задача кластеризации. Метод центров тяжести. Метод медоидов. Другие методы кластеризации: EM-алгоритм, DBSCAN, иерархическая кластеризация.	ПК-3	Устный опрос Практические задания
--	------	--------------------------------------

## 5.1 Методические указания по подготовке студентов к текущему и промежуточному контролю по дисциплине «Научное программирование и машинное обучение»

### Подготовка к устному опросу

Устный опрос представляет собой систему вопросов, позволяющих оценить уровень знаний по основным разделам, темам, проблемам дисциплины, а также умений обучающегося синтезировать материал предшествующих дисциплин.

При подготовке к устному опросу необходимо:

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

Вопросы к устным опросам приведены в п. 6.4. данной рабочей программы.

### Подготовка к экзамену

Промежуточная аттестация по итогам освоения дисциплины проходит в форме **экзамена**.

Подготовка к экзамену является концентрированной систематизацией всех полученных знаний по дисциплине «Машинное обучение».

В начале семестра рекомендуется внимательно изучить перечень вопросов к 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 существа того или иного вопроса за счет:

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

После изучения соответствующей тематики рекомендуется проверить наличие и формулировки вопроса по этой теме в перечне вопросов к экзамену, а также попытаться изложить ответ на этот вопрос. Если возникают сложности при раскрытии материала, следует вновь обратиться к лекционному материалу, материалам практических занятий, уточнить терминологический аппарат темы, а также проконсультироваться с преподавателем.

Вопросы к экзамену приведены в п. 6.4. данной рабочей программы.

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

**6.1 Перечень компетенций выпускников образовательной программы с указанием результатов обучения (знаний, умений, владений), характеризующих этапы их формирования, описание показателей и критериев оценивания компетенций на различных этапах их формирования.**

(ПК-1): способность творчески использовать в научной и производственно-технологической деятельности знания фундаментальных и прикладных разделов дисциплин (модулей), определяющих направленность (профиль) программы магистратуры. Базовый этап формирования.

Индикаторы компетенции	Критерии оценивания						
	«плохо»	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«очень хорошо»	«отлично»	«превосходно»
<b>Знать:</b> основные способы визуализации научных данных; основы математической статистики; специфику машинного обучения; основные типы задач в машинном обучении; некоторые методы решения задач классификации и восстановления регрессии (метод k ближайших соседей, линейная регрессия)	Отсутствие знаний материала	Наличие грубых ошибок в основном материале	Знание основного материала при наличии ошибок	Знание основного материала с заметными погрешностями	Знание основного материала с незначительными и погрешностями	Знание основного материала без ошибок	Знание основного и дополнительного материала без ошибок
<b>Уметь:</b> обрабатывать, анализировать и визуализировать научные данные; применять на практике технологии Numpy, Pandas, Matplotlib, Scipy, scikit-learn для разработки программ для решения задач научной деятельности	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

<b>Владеть:</b> навыками решения научных задач с помощью языка программирования Python	Отсутст вие минимал ьных владени й. Невозмо жность оценить владени е вследств ие отказа обучаю щегося от ответа	При решении стандартн ых задач не продемон стрирован о владение материало м. Имели место грубые ошибки	Продемонст рированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продемонст рированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстр ированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонст рированы все основные умения, решены все основные задачи с отдельными несушестве нными недочетами, выполнены все задания в полном объеме	Продемон стрирован ы все основные умения. Решены все основные задачи. Выполнен ы все задания, в полном объеме без недочетов
Шкала оценок по проценту правильно выполненных заданий	0-20%	21-50%	51-70%	71-80%	81-90%	91-99%	100%

(ПК-3): способность применять методические основы проектирования, выполнения полевых и лабораторных биологических, экологических исследований, использовать современную аппаратуру и вычислительные комплексы (в соответствии с направленностью (профилем) программы магистратуры). Начальный этап формирования.

Индикаторы компетенции	Критерии оценивания						
	«плохо»	«неудовлетв орительно»	«удовлетворит ельно»	«хорошо»	«очень хорошо»	«отлично»	«превосходн о»
<b>Знать:</b> методические основы проектирования моделей машинного обучения; основные алгоритмы решения задачи восстановления регрессии; основные алгоритмы решения задачи классификации; алгоритмы кластеризации; экспериментальны е методы оценки качества обучения	Отсутствие знаний материала	Наличие грубых ошибок в основном материале	Знание основного материала при наличии ошибок	Знание основного материала с заметными погрешностям и	Знание основного материала с незначительным и погрешностями	Знание основного материала без ошибок	Знание основного и дополнител ьного материала без ошибок
<b>Уметь:</b> использовать имеющиеся знания для решения практических задач машинного обучения с	Отсутст вие минимал ьных умений. Невозмо жность оценить	При решении стандартн ых задач не продемон стрирован	Продемонст рированы основные умения. Решены типовые задачи с негрубыми	Продемонст рированы все основные умения. Решены все основные задачи с негрубыми	Продемонстри рованы все основные умения. Решены все основные задачи.	Продемонст рированы все основные умения, решены все основные задачи с отдельными	Продемон стрирован ы все основные умения. Решены все

помощью Python; оценивать на практике качество обучения используемых моделей	наличие умений вследствие отказа обучающегося от ответа	ы основные умения. Имели место грубые ошибки	ошибками. Выполнены все задания, но не в полном объеме	ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами	Выполнены все задания, в полном объеме, но некоторые с недочетами	несущественными недочетами, выполнены все задания в полном объеме	основные задачи. Выполнены все задания, в полном объеме без недочетов
<b>Владеть:</b> навыками решения научных задач с использованием методов машинного обучения; навыками использования библиотеки scikit-learn для решения задач машинного обучения	Отсутствие минимальных владений. Невозможность оценить владение вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрировано владение материалом. Имели место грубые ошибки	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме, но некоторые с недочетами	Продемонстрированы все основные умения, решены все основные задачи с отдельными несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
Шкала оценок по проценту правильно выполненных заданий	0-20%	21-50%	51-70%	71-80%	81-90%	91-99%	100%

## 6.2 Описание шкал оценивания результатов обучения по дисциплине

Промежуточный контроль качества усвоения студентами содержания дисциплины проводится в виде **экзамена**, на котором определяется:

- уровень усвоения студентами основного учебного материала по дисциплине;
- уровень понимания студентами изученного материала;
- способности студентов использовать полученные знания для решения конкретных задач.

Экзамен проводится в устной форме в виде ответа обучающегося на теоретические вопросы курса (с предварительной подготовкой), а также представления экзаменационного практического задания (программы на языке Python). Практическое задание состоит из нескольких задач, подразумевающих определенную последовательность действий. Собеседование проводится в форме вопросов, на которые студент должен дать развернутый ответ. Практическая часть экзамена предусматривает беседу по подготовленному экзаменационному практическому заданию. Итоговая оценка за экзамен получается как средняя арифметическая из теоретического ответа по билету и оценки за практическое задание.

### Шкала оценивания устного ответа на экзамене

Оценка	Уровень подготовки
Превосходно	Высокий уровень подготовки, безупречное владение теоретическим материалом, студент демонстрирует творческий подход к решению нестандартных ситуаций. Студент дал полный и развернутый ответ на все теоретические вопросы билета. Студент активно работал на практических занятиях*, чему подтверждением является высокий средний балл за текущую успеваемость.
Отлично	Высокий уровень подготовки с незначительными ошибками. Студент дал полный и развернутый ответ на все теоретические вопросы билета. Студент активно работал на практических занятиях, чему подтверждением является высокий средний балл за текущую успеваемость.
Очень хорошо	Хорошая подготовка. Студент дал полный ответ на все теоретические вопросы билета, но допустил небольшие неточности в определениях понятий, процессов и т.п. Студент активно работал на практических занятиях, имеет высокий средний балл за текущую успеваемость.
Хорошо	В целом хорошая подготовка с заметными ошибками или недочетами. Студент дал ответ на все теоретические вопросы билета, но допустил неточности в определениях понятий, процессов и т.п. Имеются ошибки при ответах на дополнительные и уточняющие вопросы экзаменатора. Студент работал на практических занятиях, имеет хороший средний балл за текущую успеваемость.
Удовлетворительно	Минимально достаточный уровень подготовки. Студент показал минимальный уровень теоретических знаний, сделал существенные ошибки при ответе на экзаменационный вопрос, но при ответах на наводящие вопросы, смог правильно сориентироваться и в общих чертах дать правильный ответ. Студент посещал практические занятия, но имеет низкие средний балл за текущую успеваемость.
Неудовлетворительно	Подготовка недостаточная и требует дополнительного изучения материала. Студент дал ошибочные ответы, как на теоретические вопросы билета, так и на наводящие и дополнительные вопросы экзаменатора. Студент посещал практические занятия, но имеет очень низкий средний балл за текущую успеваемость.
Плохо	Студент отказался отвечать на экзаменационный билет.

\*информация предоставляется преподавателем, ведущим практические занятия.

### Критерии оценивания практического задания

Предел длительности контроля	10-15 минут
Предлагаемое количество заданий	1
Критерии оценки:	
«отлично»	Практическое задание выполнено в полном объеме (100% правильно выполненных задач) и с соблюдением всех требований, указанных в условии задания. Написанная программа является полностью самостоятельной разработкой. Ответ на вопросы по заданию изложен четко и логично.

«хорошо»	Практическое задание выполнено не в полном объеме (80-99% правильно выполненных задач). Написанная программа является полностью самостоятельной разработкой. Ответ на вопросы по заданию, в целом, изложен четко и логично с небольшими недочётами.
«удовлетворительно»	Практическое задание выполнено не в полном объеме (50-79% правильно выполненных задач) или имеются частичные заимствования программного кода. Ответ на вопросы по заданию неполный и поверхностный.
«неудовлетворительно»	Практическое задание выполнено не в полном объеме (менее 50% правильно выполненных задач) или не выполнено совсем или программный код заимствован или программа не работоспособна. Ответ на вопросы по заданию неполный и поверхностный.

### 6.3 Критерии и процедуры оценивания результатов обучения по дисциплине (модулю), характеризующих сформированность компетенций

Для оценивания результатов обучения в виде знаний используются следующие процедуры и технологии:

- устный опрос
- собеседование по вопросам на экзамене

Для оценивания результатов обучения в виде умений и владений используются следующие процедуры и технологии:

- практические задания (задания по научному программированию и машинному обучению)

### 6.4 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения, характеризующих сформированность компетенций и (или) для итогового контроля сформированности компетенции.

6.4.1. Для оценки сформированности знаний компетенций ПК-1 и ПК-3 используются:

#### 1. Устный опрос по вопросам, выносимым на практические занятия.

Устный опрос используется для контроля знаний студентов в качестве проверки результатов освоения основных понятий и принципов по дисциплине «Научное программирование и машинное обучение».

#### **Примеры вопросов к устным опросам для оценки знаний компетенции ПК-1**

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?

### ***Примеры вопросов к устным опросам для оценки знаний компетенции ПК-3***

1. Для чего используются обучающие и тестовые выборки?
2. В чём заключается суть метода перекрёстного контроля?
3. В чём заключается суть метода главных компонент?
4. Какие существуют эвристические подходы для выбора количества главных компонент?
5. Для чего используется метод наименьших квадратов?
6. Что такое коэффициент детерминации?
7. Какие существуют причины переобучения в задаче восстановления регрессии?
8. Какие предположения задаются в линейном дискриминантном анализе?
9. Что такое логистическая регрессия?
10. Что такое оптимальная разделяющая гиперплоскость?
11. В чём заключается суть использования дерева решений?
12. Какие существуют алгоритмы построения деревьев решений и в чём их отличие?
13. Что такое энтропия?
14. В чём заключается суть баггинга?
15. Чем отличаются экстремально случайные деревья от случайного леса?
16. В чём заключается суть бустинга?
17. Что такое нейронная сеть?
18. Что такое метод обратного распространения ошибки?
19. Что такое глубокое обучение?
20. Какая задача является задачей обучения без учителя?
21. В чём заключается суть метода центров тяжести?
22. Чем метод медоидов отличается от метода центров тяжести?
23. Что такое иерархическая кластеризация?

## ***2. Собеседование по вопросам, выносимым на экзамен.***

### ***Вопросы для подготовки 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.

6.4.2. Для оценки сформированности умений и владений компетенций ПК-1 и ПК-3 используются:

### **Практические задания**

Практические задания предполагают решение задач в области научного программирования и машинного обучения (написание соответствующих программ с использованием необходимых библиотек языка Python в приложении Jupyter Notebook).

#### **Примеры практических заданий для оценки умений и владений компетенции ПК-1**

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, среднеквадратичное 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.

### ***Примеры практических заданий для оценки умений и владений компетенции ПК-3***

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  $k$  ближайших соседей. Постройте графики зависимости ошибки этого метода на обучающей и тестовой выборках от  $k$ . 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  $k$  ближайших соседей. Выберите наилучшую модель и проверьте ее качество на тестовой выборке.
- 6) Примените к рассматриваемым данным
  - Линейный дискриминантный анализ
  - Квадратичный дискриминантный анализ
  - Логистическую регрессию

Для каждого метода вычислите ошибки на обучающей и тестовой выборках.

- 7) На рассматриваемых данных обучите следующие классификаторы:
  - Random Forest
  - Extremely Random Trees

Постройте 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).

## **6.5 Методические материалы, определяющие процедуры оценивания.**

Положение «О проведении текущего контроля успеваемости и промежуточной аттестации обучающихся в ННГУ», утверждённое 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.

## **7. 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. Программируем 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](http://scikit-learn.org/stable/user_guide.html).

г) программное обеспечение

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

## 8. 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 ФГОС ВО с учетом рекомендаций ОПОП ВО по направлению 06.04.01. Биология, магистерская программа «Биоинформатика».

Автор(ы) \_\_\_\_\_ м.н.с., О.С. Вершинина

Рецензент \_\_\_\_\_

Директор ИББМ \_\_\_\_\_ д.б.н., доц. М.В. Ведунова

Программа одобрена на заседании методической комиссии ИББМ от 24 февраля 2021 г., протокол № 4. \_\_\_\_\_