

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 президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

Рабочая программа дисциплины

Анализ данных

Уровень высшего образования

Специалитет

Направление подготовки / специальность

30.05.03 - Медицинская кибернетика

Направленность образовательной программы

Медицинская кибернетика

Форма обучения

очная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.О.55 Анализ данных относится к обязательной части 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
ОПК-4: Способен определять стратегию и проблематику исследований, выбирать оптимальные способы их решения, проводить системный анализ объектов исследования, отвечать за правильность и обоснованность выводов, внедрение полученных результатов в практическое здравоохранение	ОПК-4.1: Анализирует проблему исследования и находит необходимую информацию для выбора оптимальной стратегии ОПК-4.2: Критически рассматривает возможные варианты решения задачи, определяет оптимальный вариант для формирования стратегии исследования ОПК-4.3: Проводит системный анализ объектов исследования и формирует правильные и корректные выводы ОПК-4.4: Обосновывает эффективность внедрения полученных результатов в практическое здравоохранение	ОПК-4.1: Анализировать основные концептуальные и теоретические модели искусственного интеллекта ОПК-4.2: Рассматривать возможные варианты решения задачи с использованием библиотеки Scikit-Learn или среды для статистических вычислений R ОПК-4.3: Проводить системный анализ и разрабатывать концептуальные и теоретические модели ОПК-4.4: Обосновывать эффективность внедрения концептуальных и теоретических моделей искусственного интеллекта	Задания	Зачёт: Контрольные вопросы
ОПК-6: Способен обеспечивать информационно-технологическую поддержку в области здравоохранения; применять средства информационно-коммуникационных	ОПК-6.1: Знает биоинформационные технологии ОПК-6.2: Умеет применять информационные технологии в области здравоохранения, в профессиональной деятельности ОПК-6.3: Владеет	ОПК-6.1: Знать основы языка Python или среды вычислений R. ОПК-6.2: Уметь оценивать качество методов ОПК-6.3:	Тест	Зачёт: Контрольные вопросы

технологий и ресурсы биоинформатики в профессиональной деятельности; выполнять требования информационной безопасности	требованиями информационной безопасности	Владеть технологиями проектирования и анализа данных, современными инструментальными средствами разработки и внедрения данных.		
---	--	--	--	--

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	3
Часов по учебному плану	108
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	32
- занятия семинарского типа (практические занятия / лабораторные работы)	32
- КСР	1
самостоятельная работа	43
Промежуточная аттестация	0 Зачёт

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	
	О Ф О	О Ф О	О Ф О	О Ф О	О Ф О
Введение в проблематику искусственного интеллекта.	6	2	2	4	2
Вероятностная постановка задачи обучения по прецедентам	6	2	2	4	2
Экспериментальные методы оценки качества обучения	7	2	2	4	3
Метод наименьших квадратов для решения задачи восстановления регрессии	7	2	2	4	3
Проблема переобучения при решении задачи восстановления регрессии	7	2	2	4	3
Метод ближайших соседей для решения задачи классификации	7	2	2	4	3
Наивный байесовский классификатор	7	2	2	4	3

Линейный дискриминантный анализ	7	2	2	4	3
Нейронные сети. Персептрон Розенблатта	6	2	2	4	2
Машина опорных векторов	8	2	2	4	4
Деревья решений	7	2	2	4	3
Ансамбли решающих правил	7	2	2	4	3
Обучение без учителя	7	2	2	4	3
Иерархическая кластеризация	7	2	2	4	3
Основы теории Вапника–Червоненкиса	11	4	4	8	3
Аттестация	0				
КСР	1			1	
Итого	108	32	32	65	43

Содержание разделов и тем дисциплины

Введение в проблематику искусственного интеллекта.
 Вероятностная постановка задачи обучения по прецедентам
 Экспериментальные методы оценки качества обучения
 Метод наименьших квадратов для решения задачи восстановления регрессии
 Проблема переобучения при решении задачи восстановления регрессии
 Метод ближайших соседей для решения задачи классификации
 Линейный дискриминантный анализ и байесовский классификатор
 Нейронные сети. Персептрон Розенблатта
 Машина опорных векторов
 Деревья решений
 Ансамбли решающих правил
 Обучение без учителя
 Иерархическая кластеризация
 Основы теории Вапника–Червоненкиса

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.

Виды самостоятельной работы студентов:

- проработка теоретического материала лекционных занятий;
- подготовка домашних заданий к научно-практическим занятиям;
- подготовка к выполнению письменных контрольных работ;
- подготовка к промежуточной аттестации в форме зачета или экзамена

4.1. Проработка теоретического материала лекционных занятий

Выполняется самостоятельно с использованием лекционных материалов и материалов, разобранных в литературе (список обязательной и дополнительной литературы приводится). Контроль выполняется в форме проведения ежемесячного письменного экспресс - опроса по понятиям, фактам, формулировкам, выполняемого в течение 15 минут на практических занятиях. Экспресс – опрос оценивается оценками «Зачтено» – «Не зачтено».

4.2. Подготовка домашних заданий к научно-практическим занятиям

Домашние задания выдаются на практических занятиях.

Проверка выполнения домашних заданий проводится в начале каждого практического занятия.

Используется две формы контроля: – выборочная проверка выполнения заданий у двух-трех человек из группы; – проверка в форме коллективного обсуждения у доски результатов выполнения отдельных заданий одним или двумя студентами.

4.3. Подготовка к выполнению письменных контрольных работ

В течение учебного семестра проводится две домашние контрольные работы по материалам всех разделов лекционного курса.

Для подготовки к контрольным работам рекомендуется повторно прочитать теоретические разделы в задачнике, просмотреть полезные разделы в соответствующих источниках из списка рекомендованной литературы, а также самостоятельно решать несколько задач по теме контрольной работы из указанного задачника.

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 - Задания) для оценки сформированности компетенции ОПК-4:

1. Методом линейного дискриминантного анализа для каждого класса построить дискриминантную функцию и записать уравнение разделяющей поверхности.
2. Методом квадратичного дискриминантного анализа построить дискриминантные функции.
3. С помощью наивного байесова классификатора оценить вероятности $P(Y = 1 | x_1 = 1, x_1 = 2)$

Критерии оценивания (оценочное средство - Задания)

Оценка	Критерии оценивания
зачтено	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	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно»

5.1.2 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ОПК-6:

1. Какая из следующих задач является задачей обучения без учителя?

а. восстановление регрессии

б. классификация

с. кластеризация

2. Какая из следующих задач является задачей обучения с учителем?

а. классификация

б. кластеризация

с. восстановление функции распределения

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	51% и выше верно выполненных заданий
не зачтено	0-50% верно выполненных заданий

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

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатор 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	При решении стандартных задач не продемонстрированы основные умения. Имели	Продемонстрированы основные умения. Решены типовые задачи с	Продемонстрированы все основные умения. Решены все основные	Продемонстрированы все основные умения. Решены все основные	Продемонстрированы все основные умения. Решены все	Продемонстрированы все основные умения. Решены все основные задачи.

	отказа обучающегося от ответа	место грубые ошибки	негрубыми ошибками. Выполнены все задания, но не в полном объеме	задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	задачи. Выполнены все задания в полном объеме, но некоторые с недочетами	основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	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 некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач с некоторым и недочетами	Продемонстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрирован 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 ОПК-4

1. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия.
2. Нейронные сети. Персептрон Розенблатта. Алгоритм обучения персептрона как метод стохастического градиентного спуска. Нейронные сети для решения задач классификации и восстановления регрессии. Обучение сети. Регуляризация как метод борьбы с переобучением. Понятие о глубоких нейронных сетях.
3. Машина опорных векторов. Ядра и спрямляющие пространства.
4. Иерархическая кластеризация. Агломеративные и разделяющие методы.
5. Основы теории Вапника–Червоненкиса. Лемма Бернштейна. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечного класса решающих правил. Обоснование принципа минимизации эмпирического риска.
6. Размерность Вапника–Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска.

5.3.2 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ОПК-6

1. Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов классификатор.
2. Экспериментальные методы оценки качества обучения.
3. Метод наименьших квадратов для решения задачи восстановления регрессии. Его вывод на основе метода максимального правдоподобия. Линейная регрессионная модель. Система нормальных уравнений. Основы регрессионного анализа (проверка значимости коэффициентов, коэффициент детерминации Пирсона, доверительные интервалы, анализ остатков).
4. Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов.
5. Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа.
6. Наивный байесовский классификатор.
7. Деревья решений. Метод CART (classification and regression trees) для решения задач классификации и восстановления регрессии. Отсечения ветвей и выбор финального дерева. Методы обработки пропущенных значений.
8. Ансамбли решающих правил (классификаторов). Простое и взвешенное голосование. Бустинг. Алгоритм AdaBoost. Оценка ошибки предсказания. Бустинг и аддитивные модели. Градиентный бустинг. Алгоритм градиентного бустинга деревьев решений (MART). Баггинг. Алгоритм случайных деревьев (случайный лес.).

9. Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких множеств. ЕМ-алгоритм.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Все компетенции (части компетенции), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно». Задание выполнено.
не зачтено	Компетенция сформирована на уровне «неудовлетворительно», или хотя бы одна компетенция сформирована на уровне «плохо». Отказ от выполнения задания.

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

Основная литература:

1. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. - Москва : Юрайт, 2024. - 448 с. - (Высшее образование). - URL: <https://urait.ru/bcode/557384> (дата обращения: 15.08.2024). - ISBN 978-5-534-19964-2 : 1869.00. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=901624&idb=0>.
2. Сандуляк Д. А. Анализ и обработка экспериментальных данных : учебно-методическое пособие / Сандуляк Д. А., Полисмакова М. Н. - Москва : РТУ МИРЭА, 2023. - 64 с. - Книга из коллекции РТУ МИРЭА - Инженерно-технические науки. - ISBN 978-5-7339-1902-7., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=888460&idb=0>.

Дополнительная литература:

1. Борисова Людмила Робертовна (Финансовый университет при Правительстве Российской Федерации). Математика и анализ данных с поддержкой MS Excel и языка R : Учебное пособие / Финансовый университет при Правительстве Российской Федерации. - Москва : Прометей, 2023. - 728 с. - ВО - Бакалавриат. - ISBN 978-5-00172-445-2., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=890101&idb=0>.

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

1. The R Project for Statistical Computing <https://www.r-project>
2. Welcome to Python.org <https://www.python.org/>
3. scikit-learn: machine learning in Python scikit-learn.org/

7. Материально-техническое обеспечение дисциплины (модуля)

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

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/специальности 30.05.03 - Медицинская кибернетика.

Автор(ы): Конаков Антон Алексеевич, кандидат физико-математических наук.

Заведующий кафедрой: Конаков Антон Алексеевич, кандидат физико-математических наук.

Программа одобрена на заседании методической комиссии от 05.12.2023г., протокол № 2.