

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 № 6 от 31.05.2023 г.

Рабочая программа дисциплины

Машинное обучение

Уровень высшего образования

Специалитет

Направление подготовки / специальность

30.05.02 - Медицинская биофизика

Направленность образовательной программы

Медицинская биофизика

Форма обучения

очная

г. Нижний Новгород

2023 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.О.68 Машинное обучение 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-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	ОПК-6.1: Знает основные алгоритмы решения задачи восстановления регрессии и классификации, алгоритмы кластеризации, основные концептуальные и теоретические модели машинного обучения ОПК-6.2: Умеет использовать имеющиеся знания для решения практических задач машинного обучения ОПК-6.3: Владеет навыками использования среды статистических вычислений R или библиотеки ScikitLearn для решения задач машинного обучения, разработки концептуальных и теоретических моделей машинного обучения	Контрольная работа Тест	Зачёт: Контрольные вопросы

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)	0
- КСР	1
самостоятельная работа	75
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/ лабора торные работы), часы	Всего	
	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0	0 Ф 0
Тема 1. Содержательная постановка задачи машинного обучения	7	2		2	5
Тема 2. Вероятностная постановка задачи обучения по прецедентам	9	4		4	5
Тема 3. Экспериментальные методы оценки качества обучения	7	2		2	5
Тема 4. Метод наименьших квадратов для решения задачи восстановления регрессии	7	2		2	5
Тема 5. Проблема переобучения при решении задачи восстановления регрессии	7	2		2	5
Тема 6. Метод ближайших соседей для решения задачи классификации	7	2		2	5
Тема 7. Наивный байесовский классификатор	7	2		2	5
Тема 8. Линейный дискриминантный анализ	7	2		2	5
Тема 9. Нейронные сети. Персептрон Розенблатта	7	2		2	5
Тема 10. Машина опорных векторов	7	2		2	5
Тема 11. Деревья решений	7	2		2	5
Тема 12. Ансамбли решающих правил	7	2		2	5
Тема 13. Обучение без учителя	7	2		2	5
Тема 14. Иерархическая кластеризация	7	2		2	5
Тема 15. Основы теории Вапника–Червоненкиса	7	2		2	5
Аттестация	0				

КСР	1			1	
Итого	108	32	0	33	75

Содержание разделов и тем дисциплины

Содержательная постановка задачи машинного обучения
 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
 Наивный байесовский классификатор
 Линейный дискриминантный анализ
 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 занятиям;
- подготовка к выполнению письменных контрольных работ;
- подготовка к промежуточной аттестации в форме зачета

При подготовке может быть использовано методическое пособие: Золотых Н.Ю. Машинное обучение. Курс лекций. Нижний Новгород: ННГУ, 2007. <http://www.uic.nnov.ru/~zny/ml>

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 - Контрольная работа) для оценки сформированности компетенции ОПК-6:

1. Дана обучающая выборка

x_1	0	1	1	0	0	1	1	2	6
x_2	3	3	1	0	1	1	2	3	1
y	0	0	0	0	1	1	1	1	1

Методом линейного дискриминантного анализа для каждого класса построить дискриминантную функцию и записать уравнение разделяющей поверхности.

2. Дана обучающая выборка (см. таблицу выше). Методом квадратичного дискриминантного анализа построить дискриминантные функции.

3. Дана обучающая выборка (см. таблицу выше). С помощью наивного байесова классификатора оценить вероятности $P(Y=1 | x_1=1, x_2=2)$

4. Дана обучающая выборка:

x	-1	0	0	1	2
y	1	-2	1	7	8

Методом наименьших квадратов построить полиномиальную модель вида $f(x) = \beta_0 + \beta_1x + \beta_2x^2$.

5. Дана обучающая выборка (см. выше). Методом ридж-регрессии построить полиномиальную модель вида $f(x) = \beta_0 + \beta_1x + \beta_2x^2$, если параметр регуляризации $\lambda = 2$.

Критерии оценивания (оценочное средство - Контрольная работа)

Оценка	Критерии оценивания
зачтено	Выполнение более 60% заданий
не зачтено	Выполнение менее 60% заданий

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

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

- a. восстановление регрессии
- b. классификация
- c. кластеризация (+)

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

- a. классификация (+)
- b. кластеризация
- c. восстановление функции распределения

3. Переобучением называется

- a. явление, возникающее при решении задач обучения с учителем, при котором ошибка решающего правила на тестовой выборке много больше ошибки на обучающей выборке (+)
- b. явление, возникающее при решении задач обучения с учителем, при котором ошибка решающего правила на тестовой выборке много меньше ошибки на обучающей выборке
- c. явление, возникающее при решении задач обучения с учителем, при котором решающее правило допускает большую ошибку на тестовой выборке

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	Выполнение более 60% заданий
не зачтено	Выполнение менее 60% заданий

Оценка	Критерии оценивания

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 ОПК-6

1. Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов лассификатор.
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. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия.

8. 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.

9. Машина опорных векторов. Ядра и спрямляющие пространства.

10. 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

11. 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.).

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

13. Иерархическая кластеризация. Агломеративные и разделяющие методы.

14. 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.

15. Размерность Вапника–Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	Уровень подготовки в объеме, соответствующем программе дисциплины. Допущено несколько несущественных ошибок
не зачтено	Уровень подготовки ниже минимальных требований. Имели место грубые ошибки

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

Основная литература:

1. Платонов Алексей Владимирович. Машинное обучение : учебное пособие для вузов / А. В. Платонов. - Москва : Юрайт, 2024. - 85 с. - (Высшее образование). - URL: <https://urait.ru/bcode/544780> (дата обращения: 15.08.2024). - ISBN 978-5-534-15561-7 : 349.00. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?>

Action=FindDocs&ids=904561&idb=0.

Дополнительная литература:

1. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / Рашка С. - Москва : ДМК-пресс, 2017., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=659331&idb=0>.
2. Баланов А. Н. Машинное обучение и искусственный интеллект : учебное пособие для спо / Баланов А. Н. - Санкт-Петербург : Лань, 2024. - 80 с. - Книга из коллекции Лань - Информатика. - ISBN 978-5-507-49195-7., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=898834&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, компьютерами.

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 30.05.02 - Медицинская биофизика.

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 06.09.2022, протокол № 1.