

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

Машинное обучение и искусственный интеллект

Уровень высшего образования

Бакалавриат

Направление подготовки / специальность

05.03.06 - Экология и природопользование

Направленность образовательной программы

Информационные технологии в экологии

Форма обучения

очная

г. Нижний Новгород

2024 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.18 Машинное обучение и искусственный интеллект относится к части, формируемой участниками образовательных отношений образовательной программы.

2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ПК-15-э: Владеет знаниями о теоретических основах биогеографии, экологии животных, растений и микроорганизмов	ПК-15-э.1: Знает: - теоретические основы биогеографии ПК-15-э.2: Умеет: - использовать теоретические знания основ экологии животных и растений ПК-15-э.3: Владеет: - основными положениями экологии микроорганизмов	ПК-15-э.1: Знает принципы применения машинного обучения в задачах экологии ПК-15-э.2: Умеет разрабатывать приложения машинного обучения искусственного интеллекта к задачам экологии ПК-15-э.3: Владеет навыками адаптации машинного обучения к задачам экологии	Собеседование	Зачёт: Контрольные вопросы
ПК-2-ит: Способен осуществлять проектирование программного обеспечения ИС и разрабатывать техническую документацию на его компоненты	ПК-2-ит.1: Демонстрирует знание современных языков и систем программирования, технологий проектирования программного обеспечения ПК-2-ит.2: Демонстрирует умение сформулировать требования к разрабатываемому программному обеспечению, выполнить его реализацию и оформить техническую документацию на его компоненты ПК-2-ит.3: Имеет практический опыт проектирования программного обеспечения	ПК-2-ит.1: Знает устройство и основные архитектуры нейронных сетей, применяющиеся в задачах экологии. ПК-2-ит.2: Умеет комбинировать различные архитектурные решения, функции потерь и приемы подготовки данных для оптимального решения поставленных задач машинного обучения. ПК-2-ит.3:	Тест	Зачёт: Контрольные вопросы

	конкретной ИС и разработки технической документации на ее компоненты	Владеет навыками разработки технической документации для прикладных ИС на основе машинного обучения и искусственного интеллекта.		
ПК-4-ит: Способен проводить исследование и описание процессов принятия решений в конкретной предметной (проблемной) области с применением современных информационных технологий, в том числе основанных на моделях и методах искусственного интеллекта	<p>ПК-4-ит.1: Демонстрирует знание современных моделей и методов интеллектуальной поддержки процессов принятия решений</p> <p>ПК-4-ит.2: Демонстрирует умение применять системный подход к исследованию и описанию предметной (проблемной) области, формированию требований к ИС (ИИС) с учетом возможностей интеллектуальных технологий</p> <p>ПК-4-ит.3: Имеет практический опыт исследования и описания конкретной предметной области, разработки технического задания, эскизного и технического проектов ИС (ИИС)</p>	<p>ПК-4-ит.1: Знает теоретические основы принятия решений на основе моделей машинного обучения и искусственного интеллекта.</p> <p>ПК-4-ит.2: Умеет определять требования к разработке ИС на основе машинного обучения и искусственного интеллекта.</p> <p>ПК-4-ит.3: Владеет навыками разработки ИС на основе машинного обучения и искусственного интеллекта.</p>	Расчетно-графическое задание	Зачёт: Контрольные вопросы

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очная
Общая трудоемкость, з.е.	2
Часов по учебному плану	72
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	24
- занятия семинарского типа (практические занятия / лабораторные работы)	24
- КСР	1
самостоятельная работа	23
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
Тема 1. Типы задач машинного обучения	11	4	4	8	3
Тема 2. Алгоритмы кластеризации	12	4	4	8	4
Тема 3. Деревья решений	12	4	4	8	4
Тема 4. Нейронные сети	12	4	4	8	4
Тема 5. Глубокое обучение	12	4	4	8	4
Тема 6. Ансамблевые методы	12	4	4	8	4
Аттестация	0				
КСР	1				1
Итого	72	24	24	49	23

Содержание разделов и тем дисциплины

Тема 1. Типы задач машинного обучения

Тема 2. Алгоритмы кластеризации

Тема 3. Деревья решений

Тема 4. Нейронные сети

Тема 5. Глубокое обучение

Тема 6. Ансамблевые методы

Практические занятия /лабораторные работы организуются, в том числе, в форме практической подготовки, которая предусматривает участие обучающихся в выполнении отдельных элементов работ, связанных с будущей профессиональной деятельностью.

На проведение практических занятий / лабораторных работ в форме практической подготовки отводится: очная форма обучения - 24 ч.

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.

Для обеспечения самостоятельной работы обучающихся используются:
- электронный курс "-".

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 - Собеседование) для оценки сформированности компетенции ПК-15-э:

1. Основные типы данных в задачах биогеографии.
2. Основные типы данных в задачах экологии животных и растений.
3. Основные типы данных в задачах экологии микроорганизмов.
4. Приложения алгоритмов кластеризации в задачах экологии.
5. Приложения деревьев решений в задачах экологии.
6. Приложения нейронных сетей в задачах экологии.
7. Приложения глубокого обучения в задачах экологии.
8. Приложения бустинга в задачах экологии.
9. Приложения методов снижения размерности в задачах экологии.
10. Перспективы применения методов искусственного интеллекта в задачах экологии.

Критерии оценивания (оценочное средство - Собеседование)

Оценка	Критерии оценивания
зачтено	Достаточный уровень подготовки. Студент показывает допустимый уровень теоретических знаний, может делать существенные ошибки, но при ответах на наводящие вопросы, может правильно сориентироваться и в общих чертах дать правильный ответ.
не зачтено	Подготовка недостаточная и требует дополнительного изучения материала. Студент дает ошибочные ответы, как на теоретические вопросы, так и на наводящие и дополнительные вопросы преподавателя.

5.1.2 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ПК-2-ит:

1. К какому типу задач машинного обучения относится задача ранжирования?

- а) обучение с учителем;
- б) обучение без учителя.

2. Чем шире пространство возможных моделей, тем больше вероятность переобучения?

- а) да;
- б) нет.

3. Чем больше обучающая выборка, тем больше вероятность переобучения?

- а) да;
- б) нет.

4. ROC-кривая показывает:

- а) зависимость вероятности детектирования объекта от вероятности «ложной тревоги»;
- б) зависимость вероятности «пропуска цели» объекта от вероятности «ложной тревоги».

5. Какой из методов классификации основан на идее максимизации зазора между классами?

- а) логистическая регрессия;
- б) метод опорных векторов.

6. Если удалить только один опорный вектор из обучающей выборки решающее правило, полученное методом опорных векторов, не изменится:

- а) да;
- б) нет.

7. Логистическая регрессия реализует нелинейное решающее правило, поскольку логистическая функция не линейна:

- а) да;
- б) нет.

8. Какие из вероятностных моделей задаются условным распределением предсказаний (меток класса) при заданном описании объекта?

- а) порождающие;
- б) дискриминативные.

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	50–100% правильных ответов

Оценка	Критерии оценивания
не зачтено	менее 50% правильных ответов

5.1.3 Типовые задания (оценочное средство - Расчетно-графическое задание) для оценки сформированности компетенции ПК-4-ит:

Расчетно-графическое задание 1.

Задача 1. Реализуйте алгоритм kNN классификации по k ближайшим соседям, используя простое евклидовое расстояние.

Задача 2. Реализуйте алгоритм k-means для кластеризации на 2-4 кластера.

Задача 3. Реализуйте алгоритм DBSCAN, найдите параметры для кластеризации на 4 кластера.

Расчетно-графическое задание 2.

Задача 1. Реализуйте алгоритмы построения дерева с критерием информационного выигрыша и критерием Джини и определению класса по мажоритарному классу в листе. Найдите оптимальную глубину дерева в обоих случаях (в отрезке 2-10).

Задача 2. Примените метод SVM (например, из библиотеки sklearn) для датасета blobs2. Визуализируйте результат (разбиение плоскости и опорные вектора) при разных вариантах ядер (линейное; полиномиальное степеней 2,3,5; RBF).

Задача 3. Реализуйте алгоритм логистической регрессии со стохастическим градиентным спуском, обучите его на датасете spambase_old (train) и проверьте на датасете spambase_new (val). Получите ROC кривые для вариантов без нормировки и с нормировкой признаков.

Задача 4. Модифицируйте модель из задачи 3, заменив последний нейрон на 10 нейронов, и реализовав мультиклассовую классификацию с softmax в качестве решающей функции и кросс-энтропией в качестве функции потерь и обучите на подготовленном датасете mnist.

Расчетно-графическое задание 3.

Задача 1. Реализуйте алгоритм линейной регрессии, и полиномиальной регрессии (для датасета noisysine – степеней от 2 до 5, для датасета hydrodynamics – степени 2) без регуляризации.

Задача 2. Реализуйте алгоритм гребневой регрессии и найдите оптимальный параметр регуляризации для случаев из задачи 1.

Задача 3. Найдите максимум функции с помощью алгоритма кросс-энтропийного поиска, изображая распределение на каждом шаге.

Задача 4. Найдите лучший путь в задаче коммивояжёра с помощью алгоритма отжига.

Критерии оценивания (оценочное средство - Расчетно-графическое задание)

Оценка	Критерии оценивания
зачтено	50–100% правильно выполненных задач
не зачтено	менее 50% правильно выполненных задач

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 отдельным и 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 некоторым и недочетами	Продemonстрированы базовые навыки при решении стандартных задач с некоторым и недочетами	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продemonстрирован творческий подход к решению нестандартных задач

Шкала оценивания при промежуточной аттестации

Оценка	Уровень подготовки
--------	--------------------

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

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

5.3.1 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-15-э

1. Основные типы данных в задачах экологии.
2. Приложения алгоритмов кластеризации и деревьев решений в задачах экологии.
3. Приложения нейронных сетей и глубокого обучения в задачах экологии.
4. Приложения бустинга в задачах экологии.
5. Приложения методов снижения размерности в задачах экологии.

5.3.2 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-2-ит

1. Дать определение машинному обучению.
2. Дать определение Data Mining.
3. Какие три компонента включает в себя машинное обучение?
4. Какая цель машинного обучения?
5. Что такое пространство признаков?
6. Перечислить этапы машинного обучения.
7. Перечислить задачи машинного обучения.
8. Что включает в себя процесс решения задачи машинного обучения?
9. Какие категории включает в себя классическое обучение?
10. Перечислить популярные алгоритмы классификации.
11. Для чего используется регрессия?
12. При решении каких задач используется обучение без учителя?

13. Что означает задача кластеризации?
14. Какие цели преследует кластеризация?
15. В чем состоит задача сокращения размерности данных?
16. Перечислить популярные алгоритмы сокращения размерности данных.
17. В чем заключается метод ассоциаций?
18. Дать определение метода обучения с подкреплением.
19. Что такое ансамбль?
20. В чем состоит процедура стекинга?
21. В чем состоит процедура беггинга??
22. В чем состоит процедура бустинга?
23. Дать определение нейронной сети.

5.3.3 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-4-ит

1. Препроцессинг. Масштабирование. Нормировка. Полиномиальные признаки. One-hot encoding.
2. Кластеризация. kMeans, MeanShift, DBSCAN, Affinity Propagation.
3. Смещение и дисперсия (bias and variance). Понятие средней гипотезы.
4. Ансамблевые методы. Soft and Hard Voting. Bagging. Случайные леса. AdaBoost.
5. Типы обучения: с учителем, без учителя, с подкреплением, с частичным участием учителя, активное обучение.
6. Бустинг деревьев решений.
7. Ошибка внутри и вне выборки. Ошибка обобщения. Неравенство Хёфдинга. Валидация и кросс-валидация.
8. Линейная регрессия. Полиномиальная регрессия. Гребневая регрессия.
9. Размерность Вапника-Червоненкиса. Размерность Вапника-Червоненкиса для перцептрона.
10. Логистическая регрессия. Градиентный спуск.
11. Пороговые условия. Эффективность по Парето. Precision-Recall и ROC кривые. AUC.
12. Ансамблевые методы регрессии. RANSAC. Theil-Sen. Huber.
13. Перцептрон. Перцептрон с карманом.
14. Метод опорных векторов. Постановка задачи. Формулировка и решение двойственной задачи. Типы опорных векторов. Ядра.
15. Гипотезы и дихотомии. Функция роста. Точка поломки. Доказательство полиномиальности функции роста в присутствии точки поломки.
16. Деревья решений. Информационный выигрыш, критерий Джини. Регуляризация деревьев. Небрежные решающие деревья.
17. Байесовский классификатор. Типы оценки распределений признаков (Gaussian, Bernoulli, Multinomial). ЕМ алгоритм.
18. Нейронные сети. Перцептрон Розенблатта. Функции активации. Обратное распространение градиента. Softmax.
19. Стохастическая оптимизация. Hill Climb. Отжиг. Генетический алгоритм.
20. Метрические классификаторы. kNN. WkNN. Отбор эталонов. DROP5. Kdtree.

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
зачтено	50–100% правильно выполненных задач
не зачтено	менее 50% правильно выполненных задач

Оценка	Критерии оценивания

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

Основная литература:

1. Платонов А. В. Машинное обучение : учебное пособие / А. В. Платонов. - Москва : Юрайт, 2022. - 85 с. - (Высшее образование). - URL: <https://urait.ru/bcode/508804> (дата обращения: 14.08.2022). - ISBN 978-5-534-15561-7 : 319.00. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=821948&idb=0>.

Дополнительная литература:

1. Барский А. Б. Искусственный интеллект и логические нейронные сети : курс лекций / Барский А. Б. - Санкт-Петербург : Интермедия, 2019. - 360 с. - Книга из коллекции Интермедия - Информатика. - ISBN 978-5-4383-0155-4., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=880301&idb=0>.

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

1. <http://www.r-project.org/> – R Project: The R Project for Statistical Computing – язык программирования и среда разработки для статистических вычислений.
2. <http://cran.r-project.org/> – CRAN: The Comprehensive R Archive – архив пакетов для расширения языка программирования R.
3. <http://www.rstudio.com/> – R Studio – среда разработки для языка программирования R.
4. <http://www.statsoft.ru/home/textbook/> – электронный учебник по статистике и планированию эксперимента.
5. <http://r-analytics.blogspot.ru/> – R: Анализ и визуализация данных.

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 05.03.06 - Экология и природопользование.

Автор(ы): Якимов Василий Николаевич, доктор биологических наук, доцент.

Заведующий кафедрой: Якимов Василий Николаевич, доктор биологических наук.

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