

**АВТНОМНАЯ НЕКОММЕРЧЕСКАЯ ОБРАЗОВАТЕЛЬНАЯ
ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАУЧНО-ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «СИРИУС»
(АНОО ВО «УНИВЕРСИТЕТ «СИРИУС»)**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Основные алгоритмы машинного обучения

Уровень образования: высшее образование – программа магистратуры

Направление подготовки: 06.04.01 Биология
09.04.03 Прикладная информатика

Направленность (профиль): Биоинформатика

1. Общая характеристика дисциплины (модуля)

1.1. Цель дисциплины: формирование у студентов теоретических знаний и практических навыков численных методов и методов оптимизации для решения различных прикладных задач в математической робототехнике.

1.2. Задачи дисциплины: развитие у студентов математического мышления, умения ставить, исследовать и решать сложные задачи прикладной математики и информатики.

1.3. Общая трудоемкость дисциплины (модуля): 3 з.е.

1.4. Планируемые результаты обучения по дисциплине (модулю):

Формируемые компетенции (код компетенции, формулировка)	Планируемые результаты обучения по дисциплине (модулю) (индикаторы достижения компетенций)
ПК-2. Способен использовать современный математический аппарат при проведении научных исследований	ИПК-2.1. Знает современные алгоритмы, средства разработки и программные средства, а также принципы написания программ на различных языках программирования
	ИПК-2.2. Осуществляет анализ и выбор методов решения профессиональных задач на основе теоретических знаний в области информационных технологий
	ИПК-2.3. Разрабатывает оригинальные алгоритмы и программные средства для решения профессиональных задач в области биоинформатики, биоинженерии, биотехнологии и фарминдустрии.
ПК-5. Способен определять необходимые системные и программные средства для разработки и отладки прикладного программного обеспечения в современных специализированных программных комплексах, а также реализовывать в них новые алгоритмы	ИПК-5.1. Знает математические алгоритмы и принципы определения необходимых системных и программных средств для решения профессиональных задач
	ИПК-5.2. Определяет необходимые системные и программные средства для разработки и отладки прикладного программного обеспечения в современных специализированных программных комплексах
	ИПК-5.3. Реализует новые алгоритмы в современных специализированных программных комплексах

2. Структура и содержание дисциплины (модуля)

2.1. Объем дисциплины (модуля) и виды учебной деятельности:

Виды учебной деятельности	1 семестр	Всего
Контактная работа обучающихся с преподавателем, всего ч.	72	72
Лекционные занятия, ч.	36	36
Практические (семинарские) занятия, ч.	34	34
Лабораторные занятия, ч.	х	х
Промежуточная аттестация – экзамен, ч	х	х
Промежуточная аттестация – зачет с оценкой, ч	х	х

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Основные алгоритмы машинного обучения»	Лист 3 Листов 11
----------------------------------	--	---------------------

Промежуточная аттестация – зачет, ч	2	2
Самостоятельная работа обучающихся, всего ч.	36	36
Общая трудоемкость, ч.	108	108
Общая трудоемкость, з.е.	3	3

2.2. Структура дисциплины (модуля) по разделам (темам) и видам учебной деятельности:

Наименования разделов (тем) дисциплины (модуля)	Лекционные занятия, ч	Практические (семинарские) занятия, ч	Лабораторные занятия, ч	Промежуточная аттестация, ч	Самостоятельная работа, ч	Всего, ч	Форма текущего контроля / промежуточной аттестации
Введение.	2	2			2	6	контрольная работа
Раздел 1. Предмет и задачи машинного обучения и анализа данных.	2	2			2	6	контрольная работа
Раздел 2. Линейная регрессия. Градиентный спуск.	2	2			2	6	контрольная работа
Раздел 3. Работа с признаками. One-hot-encoding, ordinal encoding, label encoding, hashing trick. Работа с текстами. TF-IDF	2	2			2	6	контрольная работа
Раздел 4. Общий вид метрического классификатора.	2	2			2	6	контрольная работа
Раздел 5. Метрики качества. Точность, полнота, F-мера. Анализ с помощью ROC кривой.	2	2			2	6	контрольная работа
Раздел 6. Алгоритмы кластеризации с фиксированным количеством кластеров.	2	2			2	6	контрольная работа
Раздел 7. Перцептрон и разделяющая гиперплоскость.	2	2			2	6	контрольная работа
Раздел 8. Голосование. Бутстраппинг. Случайный лес.	2	2			2	6	контрольная работа
Раздел 9. Нейронные сети и алгоритм обратного распространения градиента. Глубокое обучение.	2	2			2	6	контрольная работа
Раздел 10. Задачи компьютерного зрения.	2	2			2	6	контрольная работа

Сверточные нейронные сети («CNN»).							
Раздел 11. Архитектуры нейронных сетей для классификации изображений и для задачи детектирования объектов.	2	2			2	6	контрольная работа
Раздел 12. Рекуррентные нейронные сети. Виды связей.	2	2			2	6	контрольная работа
Раздел 13. Механизмы Attention и Self-attention.	2	2			2	6	контрольная работа
Раздел 14. Автоэнкодеры. Шумоподавляющие автоэнкодеры.	2	2			2	6	контрольная работа
Раздел 15. Обучение с подкреплением. Глубокое Q-обучение.	2	2			2	6	контрольная работа
Раздел 16. Интерпретация моделей машинного обучения.	2	2			2	6	контрольная работа
Раздел 17. Огромные языковые модели. ChatGPT	2				2	4	контрольная работа
Промежуточная аттестация				2		2	Зачёт
Итого	36	34	0	2	36	108	

2.3. Содержание разделов (тем) дисциплины (модуля):

Наименования разделов (тем) дисциплины	Содержание разделов (тем) дисциплины
Введение.	Краткая история дисциплины. Общая постановка задачи оптимизации. Примеры
Раздел 1. Предмет и задачи машинного обучения и анализа данных.	Основные принципы, задачи и подходы, использование в различных областях науки и индустрии. Основные этапы эволюции алгоритмов машинного обучения.
Раздел 2. Линейная регрессия.	Градиентный спуск. Полиномиальная регрессия. Смещение и дисперсия. Проблема переобучения. Регуляризация. Гребневая регрессия и лассо-регрессия.
Раздел 3. Работа с признаками. One-hot-encoding, ordinal encoding, label encoding, hashing trick. Работа с текстами. TF-IDF	One-hot-encoding, ordinal encoding, label encoding, hashing trick. Работа с текстами. TF-IDF
Раздел 4. Общий вид метрического классификатора.	Алгоритм К ближайших соседей. Метод локтя для выбора числа соседей. Проблема выбора начальных точек, алгоритм K++.
Раздел 5. Метрики качества. Точность, полнота, F-мера. Анализ с помощью ROC кривой.	Алгоритм построения деревьев решений. Критерий информационного выигрыша и критерий Джини. Леса решающих деревьев.
Раздел 6. Алгоритмы кластеризации с фиксированным количеством кластеров.	Алгоритмы кластеризации по плотности. Иерархическая кластеризация.

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Основные алгоритмы машинного обучения»	Лист 5 Листов 11
-------------------------------	--	---------------------

Раздел 7. Перцептрон и разделяющая гиперплоскость.	Перцептрон и разделяющая гиперплоскость.
Раздел 8. Голосование. Бутстраппинг.	Случайный лес. Бустинг, адаптивный бустинг, градиентный бустинг.
Раздел 9. Нейронные сети и алгоритм обратного распространения градиента. Глубокое обучение.	Функции активации. Батчи и эпохи. Затухающие и разрывающиеся градиенты. Целевая функция softmax, расстояние Кульбака – Лейблера, относительная энтропия.
Раздел 10. Задачи компьютерного зрения.	Сверточные нейронные сети («CNN»). Свертка и пулинг. Batch Normalization.
Раздел 11. Архитектуры нейронных сетей для классификации изображений и для задачи детектирования объектов.	Задача семантической сегментации. Трекинг движущихся объектов.
Раздел 12. Рекуррентные нейронные сети. Виды связей.	Обратное распространение ошибки градиентов через время. Архитектуры LSTM и GRU
Раздел 13. Механизмы Attention и Self-attention.	Архитектура трансформер. Декодер и энкодер. BERT. GPT3
Раздел 14. Автоэнкодеры. Шумоподавляющие автоэнкодеры.	Вариационные автоэнкодеры. Генеративно-состязательные сети.
Раздел 15. Обучение с подкреплением.	Глубокое Q-обучение. Приложения глубокого обучения к робототехнике.
Раздел 16. Интерпретация моделей машинного обучения.	Моделенезависимые методы. Shar. Lime. Интерпретация нейронных сетей.
Раздел 17. Огромные языковые модели. ChatGPT	ChatGPT

2.4. Самостоятельная работа по дисциплине (модулю):

Самостоятельная работа предусматривает:

- самостоятельное изучение теоретического материала;
- подготовку к контрольной работе.

3. Текущий контроль успеваемости и промежуточная аттестация по дисциплине (модулю). Оценочные материалы

3.1. Текущий контроль успеваемости проводится в течение семестра в следующих формах:

Наименования разделов (тем) дисциплины (модуля)	Форма текущего контроля	Оценочные материалы
Введение.	контрольная работа	перечень заданий к контрольной работе
Раздел 1. Предмет и задачи машинного обучения и анализа данных.	контрольная работа	перечень заданий к контрольной работе
Раздел 2. Линейная регрессия.	контрольная работа	перечень заданий к контрольной работе

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Основные алгоритмы машинного обучения»	Лист 6 Листов 11
-------------------------------	--	---------------------

Раздел 3. Работа с признаками. One-hot-encoding, ordinal encoding, label encoding, hashing trick. Работа с текстами. TF-IDF	контрольная работа	перечень заданий к контрольной работе
Раздел 4. Общий вид метрического классификатора.	контрольная работа	перечень заданий к контрольной работе
Раздел 5. Метрики качества. Точность, полнота, F-мера. Анализ с помощью ROC кривой.	контрольная работа	перечень заданий к контрольной работе
Раздел 6. Алгоритмы кластеризации с фиксированным количеством кластеров.	контрольная работа	перечень заданий к контрольной работе
Раздел 7. Перцептрон и разделяющая гиперплоскость.	контрольная работа	перечень заданий к контрольной работе
Раздел 8. Голосование. Бутстраппинг.	контрольная работа	перечень заданий к контрольной работе
Раздел 9. Нейронные сети и алгоритм обратного распространения градиента. Глубокое обучение.	контрольная работа	перечень заданий к контрольной работе
Раздел 10. Задачи компьютерного зрения.	контрольная работа	перечень заданий к контрольной работе
Раздел 11. Архитектуры нейронных сетей для классификации изображений и для задачи детектирования объектов.	контрольная работа	перечень заданий к контрольной работе
Раздел 12. Рекуррентные нейронные сети. Виды связей.	контрольная работа	перечень заданий к контрольной работе
Раздел 13. Механизмы Attention и Self-attention.	контрольная работа	перечень заданий к контрольной работе
Раздел 14. Автоэнкодеры. Шумоподавляющие автоэнкодеры.	контрольная работа	перечень заданий к контрольной работе
Раздел 15. Обучение с подкреплением.	контрольная работа	перечень заданий к контрольной работе
Раздел 16. Интерпретация моделей машинного обучения.	контрольная работа	перечень заданий к контрольной работе
Раздел 17. Огромные языковые модели. ChatGPT	контрольная работа	перечень заданий к контрольной работе

3.2. Оценочные материалы для текущего контроля.

Примерный перечень заданий для контрольных работ:

Задание №1 выдается студентам в одном варианте и состоит из 3 задач. Каждой задаче присвоен свой балл. Срок выполнения домашнего задания - 2 недели. Форма представления обучающимися домашнего задания - представленные в письменном виде решения задач.

Задача 1 [3 балла]. Реализуйте алгоритм kNN классификации по k

ближайшим соседям, используя простое евклидовое расстояние.

Задача 2 [3 балла]. Реализуйте алгоритм k-means для кластеризации на 2-4 кластера.

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

Задание №2 выдается студентам в одном варианте и состоит из 4 задач. Каждой задаче присвоен свой балл. Срок выполнения домашнего задания - 2 недели. Форма представления обучающимися домашнего задания - представленные в письменном виде решения задач.

Задача 1 [2,5 балла]. 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 [2,5 балла]. 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 [2,5 балла]. 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 [2,5 балла]. 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.

Задание №3 выдается студентам в одном варианте и состоит из 4 задач. Каждой задаче присвоен свой балл. Срок выполнения домашнего задания - 2 недели. Форма представления обучающимися домашнего задания - представленные в письменном виде решения задач.

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

Задача 2 [2,5 балла]. Реализуйте алгоритм гребневой регрессии и найдите оптимальный параметр регуляризации для случаев из задачи 1.

Задача 3 [2,5 балла]. Найдите максимум функции с помощью алгоритма кросс-энтропийного поиска, изображая распределение на каждом шаге. Задача 4 [2,5 балла]. Найдите лучший путь в задаче коммивояжера с помощью алгоритма отжига.

Критерии оценки контрольных работ.

Критерий	Зачтено	Не зачтено
Процент правильных ответов	Успешно/в целом успешно применяет инновационные инструменты и методы при определении путей решения профессиональных задач.	Не применяет/не в полной мере применяет инновационные инструменты и методы при определении путей решения профессиональных задач.

3.3. Формой промежуточной аттестации по дисциплине является зачет.

Результаты промежуточной аттестации оцениваются как «зачтено» и «не зачтено».

Оценка «зачтено» означает успешное прохождение промежуточной аттестации по дисциплине.

3.4. Оценочные материалы для промежуточной аттестации:

Перечень вопросов для подготовки к зачету:

1. Препроцессинг. Масштабирование. Нормировка. Полиномиальные признаки. One-hot encoding. Hashing trick. TF-IDF.
2. Кластеризация. kMeans, MeanShift, DBSCAN, Affinity Propagation.
3. Смещение и дисперсия (bias and variance). Вывод разложения смещения-дисперсии для квадратичной ошибки.
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. Геометрическая интерпретация.
7. Ошибка внутри и вне выборки. Ошибка обобщения. Неравенство Хёфдинга. Валидация и кросс-валидация.
8. Линейная регрессия. Регуляризация. Полиномиальная регрессия. Гребневая регрессия.
9. Размерность Вапника-Червоненкиса. Размерность Вапника-Червоненкиса для перцептрона.
10. Логистическая регрессия. Градиентный спуск. Метрики классификации.
11. Пороговые условия. Эффективность по Парето. Precision-Recall и ROC кривые. AUC.
12. Метод опорных векторов. Постановка задачи. Формулировка и решение двойственной задачи. Типы опорных векторов. Ядра.
13. Деревья решений. Информационный выигрыш, критерий Джини. Регуляризация деревьев. Небрежные решающие деревья.
14. Наивный байесовский классификатор. Типы оценки распределений признаков (Gaussian, Bernoulli, Multinomial). EM алгоритм.

15. Нейронные сети. Перцептрон Розенблатта. Функции активации. Обратное распространение градиента. Softmax.
16. Стохастическая оптимизация. Hill Climb. Отжиг. Генетический алгоритм.
17. Метрические классификаторы. kNN. WkNN. Отбор эталонов. DROP5. Kdtree.
18. Задача понижения размерности. PCA. Построение базиса
19. Работа с текстами. Эмбединги. word2vec. doc2vec
20. Сверточные нейронные сети. Свертки, операция пулинга. Dropout. Batch Normalization.
21. Архитектуры нейронных сетей для классификации изображений. Основные идеи от Alexnet, VGG, Inception, ResNet.
22. Архитектуры нейронных сетей для детектирования объектов. Основные идеи R-CNN, Fast R-CNN, Faster R-CNN.
23. Архитектура YOLO для детектирования объектов. Задача семантической сегментации. Основные идеи SegNet и UNET.
24. Рекуррентные нейронные сети. Виды связей. Обратное распространение градиентов через время.
25. Архитектуры LSTM и GRU. Механизм Attention.
26. Механизм self-attention.
27. Автоэнкодеры. Виды. Механизм обучения.
28. Вариационные автоэнкодеры. Reparameterization trick.
29. Архитектура трансформеров. Энкодер. Декодер.
30. BERT. Основные идеи, подход к обучению. GPT3.
31. Интерпретация моделей. Permutation importance. LIME. SHAP.

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

4.1. Перечень основной литературы:

1. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2026. — 89 с.

4.2. Перечень дополнительной литературы:

1. Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning Series) / Murphy, Kevin P. MIT Press. 2014
2. Encyclopedia of Machine Learning and Data Mining / Sammut. Springer. 2016

3. П. Флах. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных Machine Learning and Knowledge Discovery in Databases / Peter A. Flach; Tijn Bie; Nello Cristianini. Springer Berlin Heidelberg. 2012

4. An introduction to statistical learning: with applications in R / G. James, D. Witten, T. Hastie, R. Tibshirani. – New York: Springer, 2013. – 426 с.

5. Data Analysis, Machine Learning and Knowledge Discovery / Spiliopoulou, Myra; Janning, Ruth; Schmidt-Thieme, Lars; Gesellschaft für Klassifikation. Springer International Publishing. 2014

4.3. Перечень современных профессиональных баз данных и ресурсов информационно-телекоммуникационной сети «Интернет»:

1. <https://stepik.org/course/4852/promo>

4.4. Учебно-методическое обеспечение дисциплины (модуля) представлено в Приложении к рабочей программе дисциплины (модуля).

5. Материально-техническое и программное обеспечение дисциплины

5.1. Материально-техническое обеспечение:

Вид аудитории	Технические средства и оборудование
<i>Учебная аудитория для проведения лекционных занятий</i>	Альфа 5.1 - учебная аудитория для проведения учебных занятий, предусмотренных программой магистратуры. Доска магнитно-маркерная поворотная BoardSYS Twist 100x160 ПО-15Ф 1 шт. Флипчарт 70*100 на роликах 1 шт. Стол-кафедра 1 шт. Стол аудиторный 1 шт. Столы-трансформеры Summa GA ученические 25 шт. Стулья на колесах ученические 25 шт. Ноутбук HP 1 шт. Интерактивная панель NexTouch Nextpanel 86” 1 шт. Радиосистема Arthur Forty U-9700C PSC (UHF) в комплекте. Акустическая система Behringer B215D 2 шт. Веб-камера 4К с технологией искусственного интеллекта JazzTel JT-Vintage-4K 1 шт. Комплект электронных презентаций.
<i>Учебная аудитория для проведения практических занятий – Компьютерный класс</i>	Бета 4.1 – учебная аудитория для проведения практических занятий (компьютерный класс). Доска магнитно-маркерная поворотная BoardSYS Twist 100x160 ПО-15Ф 1 шт. Флипчарт 70*100 на роликах 1 шт. Стол преподавателя аудиторный 1 шт. Столы и стулья ученические 42 шт. Компьютеры Lenovo ThinkCentre M920s SFF в комплекте с мониторами ПУАМА 27” и периферией – 42 шт. Интерактивная панель NexTouch Nextpanel 86” 1 шт. Радиосистема Arthur Forty U-9700C PSC (UHF) в комплекте. Акустическая система Behringer B215D 2 шт. Веб-камера 4К с технологией искусственного интеллекта JazzTel JT-Vintage-4K 1 шт. Комплект электронных презентаций.

АНОО ВО «Университет «Сириус»	Рабочая программа дисциплины (модуля) «Основные алгоритмы машинного обучения»	Лист 11 Листов 11
----------------------------------	--	----------------------

5.2. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе российского производства: не предусмотрено