

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы физики
и исследований им. Ландау**

А.В. Рогачев

	Рабочая программа дисциплины (модуля)
по дисциплине:	Машинное обучение для физиков
по направлению:	Прикладные математика и физика
профиль подготовки:	Физика и педагогика
	Физтех-школа физики и исследований им. Ландау
	Физтех-кластер академической и научной карьеры
курс:	4
квалификация:	бакалавр

Семестры, формы промежуточной аттестации:

7 (осенний) - Дифференцированный зачет

8 (весенний) - Дифференцированный зачет

Аудиторных часов: 120 всего, в том числе:

лекции: 60 час.

семинары: 60 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 60 час.

Всего часов: 180, всего зач. ед.: 4

Программу составили:

А.Е. Устюжанин, канд. физ.-мат. наук

М.И. Гущин, канд. физ.-мат. наук

Программа обсуждена на заседании Физтех-кластера академической и научной карьеры 16.04.2021

Аннотация

Технологии искусственного интеллекта (ИИ) - отличный пример междисциплинарной области исследований. Мы созерцаем его успех в распознавании изображений, голоса и управлении роботами. Однако они далеки от волшебства и в значительной степени основаны на подходах машинного обучения (ML), которое в значительной степени опирается на подходы к анализу данных, статистику, компьютерное моделирование, методы оптимизации и особый образ мышления, включающий эвристику обучения и здравый научный смысл. Чтобы оценить междисциплинарность ML, необходимо познакомиться с этими школами. Не будет большим преувеличением сказать, что искусственный интеллект - это современный подход к построению лучшей картины мира.

В этом курсе будут освещены идеи, подходы, разработки и методы, которые связывают современный ИИ с его предшественниками. С особым вниманием мы сосредоточимся на методах оптимизации, лежащих в основе подавляющего большинства методов искусственного интеллекта, рассмотрим классический пример оценки состояния квантовой системы, представленной кубитом.

Данный курс познакомит вас с основными направлениями и алгоритмами машинного обучения, применяющимися в современных исследованиях в различных областях науки и бизнеса: глубокое обучение, сверточные нейронные сети, компьютерное зрение, временные ряды, генеративные модели, автокодировщики, нейродиффузоры и методы оптимизации.

Лекционные занятия проводятся в интерактивном режиме видео-трансляций.

1. Цели и задачи

Цель дисциплины

Целью освоения дисциплины являются формирование у студентов теоретических знаний и практических навыков по основам машинного обучения, овладение студентами инструментарием, моделями и методами машинного обучения, а также приобретение навыков исследователя данных. Данный курс стремится охватить как базовые так и более продвинутые концепции, а также помочь понять здравый смысл, лежащий в основе столь популярных в наши дни технологий искусственного интеллекта. Ориентируясь на междисциплинарный подход к этим методам продемонстрировать его, начиная с моделей мышления, которые ученые расширили на протяжении веков, до подходов к оптимизации и к практическому примеру анализа квантовых систем. Познакомить с основными направлениями и алгоритмами машинного обучения: глубокое обучение, сверточные нейронные сети, компьютерное зрение, временные ряды, генеративные модели, автокодировщики, нейродифференциальные уравнения и методы оптимизации.

Задачи дисциплины

Познакомить студентов с методами и задачами машинного обучения и научить применять полученные знания в будущих исследованиях.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи

УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	УК-6.2 Способен планировать самостоятельную деятельность в решении профессиональных задач; подвергать критическому анализу проделанную работу; находить и творчески использовать имеющийся опыт в соответствии с задачами саморазвития
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
	ОПК-4.4 Владеет навыками работы с компьютером и компьютерными сетями с целью получения, хранения и обработки научной (технической, технологической) информации
ПК-2 Способен анализировать полученные в ходе научно-исследовательской работы данные и делать научные выводы (заключения)	ПК-2.1 Владеет методами статистической обработки и анализа научных данных
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

ключевые понятия, цели и задачи использования машинного обучения; методологические основы применения алгоритмов машинного обучения.

уметь:

самостоятельно выбирать метод машинного обучения, соответствующий исследовательской задаче, интерпретировать полученные результаты.

владеть:

навыками самостоятельного обучения, чтения и анализа академической литературы по применению методов машинного обучения, построения и оценки качества моделей.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Введение в машинное обучение	3	3		3
2	Оптимизация (методом градиента)	2	2		2
3	Глубокое обучение: Основы	3	3		3
4	Метрики, дилемма смещения-дисперсии: тестирование гипотез	2	2		2
5	Глубокое обучение: нейросети для изображений	2	2		2

6	Глубокое обучение: нейросети для последовательностей	3	3		3
7	Глубокое обучение: Введение в задачи обучения без учителя	3	3		3
8	Архитектура компьютерного зрения	4	4		4
9	Временные ряды / последовательности	5	5		5
10	Структуры графов	3	3		3
11	Кластеризация	2	2		2
12	Автокодировщики	2	2		2
13	Генеративно-сопоставительные сети (GAN)	4	4		4
14	Улучшенные GAN	2	2		2
15	Потоки (flows)	2	2		2
16	Детекция аномалий. Глубокая детекция аномалий с учителем	2	2		2
17	Методы оптимизации	2	2		2
18	Введение в обучение с подкреплением	3	3		3
19	Введение в суррогатные модели	3	3		3
20	Обучение признакам	3	3		3
21	Нейральные системы обыкновенных дифференциальных уравнений	2	2		2
22	Выводы на основе моделирования	3	3		3
Итого часов		60	60		60
Подготовка к экзамену		0 час.			
Общая трудоёмкость		180 час., 4 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 7 (Осенний)

1. Введение в машинное обучение

Данная тема является вводной. В ней студенты познакомятся с предпосылками зарождения искусственного интеллекта, а также с примерами задач машинного обучения и возможными направлениями дальнейшего развития.

2. Оптимизация (методом градиента)

В данной теме студенты узнают о базовых задачах оптимизации, о градиентном спуске и методе Ньютона и применят полученные знания в практической задаче.

3. Глубокое обучение: Основы

В данной теме студенты узнают, что из себя представляет нейронная сеть, познакомятся с ее математическим аппаратом. Узнают какие задачи можно решать с помощью глубоких нейронных сетей и как их тренировать.

4. Метрики, дилемма смещения-дисперсии: тестирование гипотез

В данной теме студенты познакомятся с метриками оценки качества алгоритмов регрессии и классификации, узнают о проблеме смещения и дисперсии, поймут, как можно бороться переобучением, а также научатся тестировать гипотезы.

5. Глубокое обучение: нейросети для изображений

В данной теме студенты познакомятся со сверточными сетями, узнают, в чем их преимущество при работе с изображениями по сравнению с обычными нейронными сетями и попробуют применить полученные знания на практике.

6. Глубокое обучение: нейросети для последовательностей

В данной теме студенты узнают, что такое рекуррентные нейронные сети, какие бывают архитектуры, какие задачи можно решить с их помощью.

7. Глубокое обучение: Введение в задачи обучения без учителя

В данной теме студенты узнают о том, какие задачи решаются с помощью обучения без учителя и какие алгоритмы, основанные на нейронных сетях применимы для решения таких задач.

8. Архитектура компьютерного зрения

В данной теме студенты познакомятся с одним из направлений задач глубокого обучения без участия учителя - компьютерным зрением. Узнают какие задачи могут быть решены с его помощью: распознавание объектов, сегментация, генерация, анализ видео.

9. Временные ряды / последовательности

В данной теме студенты узнают, что из себя представляют временные ряды, познакомятся с методами анализа и прогнозирования.

10. Структуры графов

В данной теме студенты познакомятся с графовыми нейронными сетями, узнают об областях их применимости и применят полученные знания на практике.

Семестр: 8 (Весенний)

11. Кластеризация

В данной теме студенты познакомятся с еще одним видом машинного обучения без учителя - кластеризацией, изучат алгоритмы и метрики оценки качества алгоритмов кластеризации.

12. Автокодировщики

В данной теме студенты познакомятся со специальной структурой нейронных сетей, позволяющей применять обучение без учителя - автокодировщиком. Узнают о его архитектуре, принципах работы и области применения.

13. Генеративно-сопоставительные сети (GAN)

В данной теме студенты познакомятся с еще одним алгоритмом машинного обучения без учителя, построенного на комбинации двух нейронных сетей - GAN. Студенты разберутся в архитектуре GAN, используемых методах и решаемых с их помощью задачах.

14. Улучшенные GAN

В данной теме студенты познакомятся с архитектурой улучшенных GAN, узнают об их преимуществах.

15. Потoki (flows)

В данной теме студенты узнают, что такое нормализующие потоки, для чего они нужны и где используются.

16. Детекция аномалий. Глубокая детекция аномалий с учителем

В данной теме студенты областью машинного обучения, которая называется детекцией аномалий, узнают сферы применения и познакомятся популярными подходами для изучения аномалий.

17. Методы оптимизации

В данной теме студенты познакомятся методами не градиентной оптимизации, узнают, какие задачи можно решить с ее помощью.

18. Введение в обучение с подкреплением

В данной теме студенты познакомятся с еще одной ветвью машинного обучения наравне с методами МО с учителем и без учителя - обучением с подкреплением. Узнают, в чем их принципиальное отличие. Познакомятся с основной концепцией и задачами, которые здесь решаются.

19. Введение в суррогатные модели

В данной теме студенты познакомятся с суррогатными моделями, узнают об их предназначении и о том, какие задачи они помогают решать.

20. Обучение признакам

В данной теме студенты познакомятся с набором техник, позволяющих автоматически обнаружить представления, необходимые для выявления признаков или классификации сырых данных - обучение признакам, а также узнают о задачах, решаемых с помощью этих техник.

21. Нейральные системы обыкновенных дифференциальных уравнений

В данной теме студенты познакомятся с тем, что называется нейральными системами ОДУ, узнают о задачах и области их применимости.

22. Выводы на основе моделирования

Данная тема является заключительной для данного курса и направлена на выработку у учащихся навыков проведения анализа и составления выводов на основе проведенного моделирования.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Данный курс является онлайн-курсом, для его прохождения обучающимися понадобится доступ к компьютеру и интернету

6.Перечень рекомендуемой литературы

Основная литература

1. Python для сложных задач: наука о данных и машинное обучение [Текст], [учеб. пособие для вузов] /Дж. Вандер Плас ; [пер. с англ. И. Пальти]. -СПб., Питер, 2018

Дополнительная литература

Дополнительная литература в открытом доступе:

1. Stevens E., Antiga L, Viehmann T, Deep Learning with PyTorch,
<https://pytorch.org/assets/deep-learning/Deep-Learning-with-PyTorch.pdf>
2. Giuseppe Carleo et al, Machine learning and the physical sciences,
https://www.researchgate.net/publication/332010806_Machine_learning_and_the_physical_sciences

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

<https://en.pelican.study/>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

Библиотеки машинного обучения pytorch, pytorch-lightning, fast-ai, scikit-learn, система дистанционного обучения Alemira, системы для проведения онлайн консультации в виде вебинаров - Zoom, MS teams.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, изучающий дисциплину, должен, с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения и понятия, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Прикладные математика и физика
профиль подготовки:	Физика и педагогика Физтех-школа физики и исследований им. Ландау Физтех-кластер академической и научной карьеры
курс:	<u>4</u>
квалификация:	бакалавр

Семестры, формы промежуточной аттестации:

7 (осенний) - Дифференцированный зачет

8 (весенний) - Дифференцированный зачет

Разработчики:

А.Е. Устюжанин, канд. физ.-мат. наук

М.И. Гущин, канд. физ.-мат. наук

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Анализирует задачу, выделяя этапы ее решения, действия по решению задачи
	УК-1.2 Находит, критически анализирует и выбирает информацию, необходимую для решения поставленной задачи
	УК-1.3 Рассматривает различные варианты решения задачи, оценивает их преимущества и недостатки
	УК-1.4 Грамотно, логично, аргументированно формирует собственные суждения и оценки
	УК-1.5 Определяет и оценивает практические последствия возможных вариантов решения задачи
УК-6 Способен управлять своим временем, выстраивать и реализовывать траекторию саморазвития на основе принципов образования в течение всей жизни	УК-6.2 Способен планировать самостоятельную деятельность в решении профессиональных задач; подвергать критическому анализу проделанную работу; находить и творчески использовать имеющийся опыт в соответствии с задачами саморазвития
ОПК-1 Способен применять фундаментальные знания, полученные в области физико-математических и (или) естественных наук, и использовать их в профессиональной деятельности	ОПК-1.1 Способен анализировать поставленную задачу, намечать пути ее решения
	ОПК-1.2 Способен строить математические модели, производить количественные расчеты и оценки
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.1 Владеет методами научного поиска и интеллектуального анализа информации при решении задач профессиональной деятельности
	ОПК-4.4 Владеет навыками работы с компьютером и компьютерными сетями с целью получения, хранения и обработки научной (технической, технологической) информации
ПК-2 Способен анализировать полученные в ходе научно-исследовательской работы данные и делать научные выводы (заключения)	ПК-2.1 Владеет методами статистической обработки и анализа научных данных
ПК-4 Способен критически оценивать применимость используемых методик и методов	ПК-4.1 Знает численные порядки величин, характерных для соответствующей профессиональной области

2. Показатели оценивания компетенций

В результате изучения дисциплины «Машинное обучение для физиков» обучающийся должен:

знать:

ключевые понятия, цели и задачи использования машинного обучения; методологические основы применения алгоритмов машинного обучения.

уметь:

самостоятельно выбирать метод машинного обучения, соответствующий исследовательской задаче, интерпретировать полученные результаты.

владеть:

навыками самостоятельного обучения, чтения и анализа академической литературы по применению методов машинного обучения, построения и оценки качества моделей.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Задания для самостоятельной работы будут оформлены в виде задач по анализу данных по разработке алгоритмов классификации/регрессии/классификации с автоматической проверкой результатов.

Разбор вопросов студентов по решению домашнего задания будет происходить на регулярных онлайн-консультациях.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Перечень контрольных вопросов (примеры на практику и теорию)

1. Выведите аналитически формулу для W : $W X = y$ при заданных X и y методом наименьших квадратов.
2. Приведите примеры алгоритмов кластеризации с заданным и не заданным количеством кластеров.
3. Для чего нужна регуляризация. Какие виды регуляризации существуют и чем они отличаются?
4. В чем преимущество сверточных сетей по сравнению с обычной нейронной сетью?

Примеры контрольных заданий

1. Найдите коэффициенты S_i в разложении матрицы плотности.
2. Что будет выходом алгоритма бинарной классификации для вектора признаков размера 10?
 - действительное число
 - 10-ти мерный булевый вектор
 - булево значение
 - два булевых значения
 - 10-ти мерный вещественный вектор
 - два вещественных числа в интервале $[0; 1]$, дающие в сумме 1
3. Рассмотрим минимизацию некоторой функции $f(x)$ с начальной точкой x_0 . Напишите, как будут выглядеть первые две итерации градиентного спуска.
4. Рассмотрим минимизацию некоторой функции $f(x)$ с начальной точкой x_0 . Напишите, как будут выглядеть первые две итерации метода Ньютона.
5. Вычислите среднюю квадратичную ошибку для целевого вектора и рассчитанного вектора.
6. Сколько всего весов будет иметь полносвязная нейронная сеть с 10 входами, двумя скрытыми слоями с 5 нейронами в каждом и одним выходом.

Критерии оценивания

Оценка отлично 10 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 9 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично 8 баллов - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо 7 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо 6 баллов - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо 5 баллов - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно 4 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно 3 балла - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно 2 балла - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно 1 балл - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Дифференцированный зачёт проводится в устной форме собеседования по заданиям и теории. Итоговая оценка учитывает баллы за работу в семестре при выполнении заданий с автоматической проверкой результатов работы.