

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

УТВЕРЖДЕНО

**и.о. директора физтех-школы
физики и исследований им.
Ландау**

А.А. Воронов

	Рабочая программа дисциплины (модуля)
по дисциплине:	Введение в машинное обучение
по направлению:	Прикладные математика и физика
профиль подготовки:	Общая и прикладная физика Физтех-школа физики и исследований им. Ландау Физтех-кластер академической и научной карьеры
курс:	1
квалификация:	магистр

Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет

Аудиторных часов: 45 всего, в том числе:

лекции: 30 час.

семинары: 15 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Всего часов: 90, всего зач. ед.: 2

Программу составил: Ф.Д. Ратников, канд. физ.-мат. наук

Программа обсуждена на заседании Физтех-кластера академической и научной карьеры 29.05.2020

Аннотация

Методы машинного обучения прочно вошли в набор инструментов современного физического исследования. Классификации типов исследуемых объектов, многомерная регрессия, быстрое обобщение генеративных моделей, мониторинг качества данных, поиск физических и технических аномалий - неполный перечень приложений этих методов для получения физических результатов высочайшего качества. В курсе будут рассмотрены базисные принципы и методики машинного обучения: обобщенные модели. Типовые задачи: классификация, регрессия, поиск аномалий в данных, тренировка генеративных моделей, анализ изображений, поиск оптимального управления и др. будут рассмотрены на примерах конкретных применений в ядерной физике и физике высоких энергий. Будут представлены подходы к решению задач с использованием детерминистических и вероятностных методов, с использованием решающих деревьев, ансамблей, нейронных сетей. В ходе семинарских занятий и выполнения самостоятельных работ студенты не только изучат примеры применения методов машинного обучения к конкретным задачам физических исследований, но и научатся сознательно выбирать наиболее подходящий под задачу подход, получают практический навык настраивания и оптимизации моделей, а также количественных оценок влияния используемых методов на точность получаемых физических результатов.

1. Цели и задачи

Цель дисциплины

обеспечить введение слушателей в предмет математического моделирования с использованием методов машинного обучения, получение ими практических навыков в области использования методов машинного обучения для решения различных задач.

Задачи дисциплины

- Ознакомление с основами современных подходов машинного обучения
- Получения теоретических и практических навыков решения типичных задач
- Изучение специфических аспектов использования машинного обучения в научных исследованиях
- Получение навыков практической работы с программными библиотеками средств машинного обучения

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
УК-2 Способен управлять проектом на всех этапах его реализации	УК-2.2 Способен прогнозировать результат деятельности и планировать последовательность шагов для достижения данного результата. Формирует план-график реализации проекта в целом и план контроля его выполнения
	УК-2.3 Способен организовать и координировать работу участников проекта, обеспечивать работу команды необходимыми ресурсами
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности

ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области профессиональной деятельности, учитывая особенности и ограничения различных методов решения	ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения
	ОПК-3.2 Способен использовать исследовательские методы при решении новых задач, применяя знания в различных областях науки (техники)
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, понимает и учитывает на практике границы применимости получаемых решений
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)
	ПК-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

содержание предмета курса «Введение в машинное обучение», соответствующую терминологию и понятийный аппарат. Понимать логику различных подходов машинного обучения в применении к различным типам задач. Иметь представление о сильных и слабых сторонах различных подходов, границы применимости.

уметь:

переводить практические задачи научного исследования на стандартный язык машинного обучения, переводить полученные результаты обратно в физический контекст и оценивать точность получаемых результатов.

владеть:

способами решения типичных задач с использованием стандартных программных пакетов средств и сред машинного обучения.

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	Машинное обучение, статистика, линейные методы регрессии	6	3		9
2	Линейные методы классификации, Особенности работы с реальными данными, Работа с признаками	6	3		9
3	Решающие деревья, Композиции алгоритмов	6	3		9

4	Нейронные сети, Обучение без учителя	8	4		12
5	Продвинутые методики: генеративные модели, поиска аномалий в данных	4	2		6
Итого часов		30	15		45
Подготовка к экзамену		0 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

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

Семестр: 1 (Осенний)

1. Машинное обучение, статистика, линейные методы регрессии

Неделя 1. История анализа данных. Постановки задач в машинном обучении: классификация, регрессия, ранжирование, кластеризация, латентные модели. Примеры задач. Виды данных. Признаки.

Неделя 2. Метод максимального правдоподобия и его свойства. Статистические гипотезы и статистические критерии.

Лемма Неймана-Пирсона. Критерий отношения правдоподобия. Аналитическое и численное решение задачи МНК.

Неделя 3. Градиентный спуск, методы оценивания градиента. Функции потерь. Регуляризация. Методы оценивания обобщающей способности, кросс-валидация. Метрики качества регрессии.

2. Линейные методы классификации, Особенности работы с реальными данными, Работа с признаками

Неделя 4. Аппроксимация эмпирического риска. Персептрон. Метод опорных векторов.

Неделя 5. Задача оценивания вероятностей, логистическая регрессия. Обобщённые линейные модели. Метрики качества в задачах классификации.

Неделя 6. Пропуски в данных. Предобработка признаков. Чистка данных. Категориальные признаки. Разреженные признаки.

Методы отбора признаков. Метод главных компонент.

3. Решающие деревья, Композиции алгоритмов

Неделя 7. Общий алгоритм построения, критерии информативности. Конкретные критерии для классификации и регрессии.

Неделя 8. Тонкости решающих деревьев: обработка пропущенных значений, стрижка, регуляризация.

Неделя 9. Общая идея bias-variance decomposition. Бэггинг, бустинг. Градиентный бустинг над решающими деревьями.

4. Нейронные сети, Обучение без учителя

Неделя 10. Структура нейронной сети. Обратное распространение ошибки. Полносвязные нейронные сети.

Неделя 11. Методы регуляризации. Примеры архитектур как наборов кубиков.

Неделя 12. Свёрточные сети. Рекуррентные сети.

Неделя 13. Задача кластеризации. K-Means, spectral clustering. Автокодировщики.

5. Продвинутые методики: генеративные модели, поиска аномалий в данных

Неделя 14. Принципы построения генеративных моделей. Вариативные автокодировщики, состязательные генеративные сети, нормализующие потоки.

Неделя 15. Постановка и подходы к задаче поиска аномалий в данных.

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

Семинары: студентам необходимы вычислительные средства с предустановленным программным обеспечением

- Python – часть установки Anaconda
- Jupyter - часть установки Anaconda
- Scikit-learn - часть установки Anaconda
- Pytorch
- Tensorflow

6. Перечень рекомендуемой литературы

Основная литература

1. Deep learning, Goodfellow, I., Bengio, Y., ISBN: 9780262035613, 2016
2. Pattern recognition and machine learning, Bishop, C. M., ISBN: 9780387310732, 2006

Дополнительная литература

1. An introduction to statistical learning : with applications in R, James G., Witten D., ISBN: 978-1-461-47137-0, 2013

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

- <https://www.deeplearningbook.org/> - Deep learning, Goodfellow, I., Bengio
- <http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf> - Pattern recognition and machine learning, Bishop, C. M.

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

На лекционных и семинарских занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

- Python – часть установки Anaconda
- Jupyter - часть установки Anaconda
- Scikit-learn - часть установки Anaconda
- Pytorch
- Tensorflow

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;
- подготовку уточняющих вопросов по изученному материалу;
- решение задач и заданий, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины и решать типичные задачи.

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:	Общая и прикладная физика Физтех-школа физики и исследований им. Ландау Физтех-кластер академической и научной карьеры (Фундаментальные взаимодействия и физика элементарных частиц)
курс:	1
квалификация:	магистр
Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет	
Разработчик:	Ф.Д. Ратников, канд. физ.-мат. наук

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними
	УК-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
	УК-1.3 Разрабатывает стратегию достижения поставленной цели как последовательность шагов, предвидя результат каждого из них и оценивая их влияние на внешнее окружение планируемой деятельности и на взаимоотношения участников этой деятельности
УК-2 Способен управлять проектом на всех этапах его реализации	УК-2.2 Способен прогнозировать результат деятельности и планировать последовательность шагов для достижения данного результата. Формирует план-график реализации проекта в целом и план контроля его выполнения
	УК-2.3 Способен организовать и координировать работу участников проекта, обеспечивать работу команды необходимыми ресурсами
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.3 Понимает междисциплинарные связи в области математики и физики и способен их применять при решении задач профессиональной деятельности
ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области профессиональной деятельности, учитывая особенности и ограничения различных методов решения	ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения
	ОПК-3.2 Способен использовать исследовательские методы при решении новых задач, применяя знания в различных областях науки (техники)
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, понимает и учитывает на практике границы применимости получаемых решений
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели
ПК-3 Способен профессионально работать с исследовательским и испытательным оборудованием (приборами и установками, специализированными пакетами прикладных программ) в избранной предметной области	ПК-3.1 Понимает принципы работы используемого оборудования (специализированных пакетов прикладных программ)
	ПК-3.2 Способен проводить эксперимент (моделирование) с использованием исследовательского оборудования (пакетов прикладных программ)
	ПК-3.3 Способен оценивать точность полученных экспериментальных (численных) результатов

2. Показатели оценивания компетенций

В результате изучения дисциплины «Введение в машинное обучение» обучающийся должен:

знать:

содержание предмета курса «Введение в машинное обучение», соответствующую терминологию и понятийный аппарат. Понимать логику различных подходов машинного обучения в применении к различным типам задач. Иметь представление о сильных и слабых сторонах различных подходов, границы применимости.

уметь:

переводить практические задачи научного исследования на стандартный язык машинного обучения, переводить полученные результаты обратно в физический контекст и оценивать точность получаемых результатов.

владеть:

способами решения типичных задач с использованием стандартных программных пакетов средств и сред машинного обучения.

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

Не предусмотрено.

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

- Математические модели
 - o Свойства
 - o Типы
 - o Виды данных
- Правдоподобие модели
 - o Метод максимального правдоподобия и его свойства.
 - o Статистические гипотезы и статистические критерии.
 - o Критерий отношения правдоподобия.
 - o Лемма Неймана-Пирсона (без доказательства).
- Метод наименьших квадратов
 - o Аналитическое решение задачи МНК.
- ЕМ алгоритм
 - o Пример для суперпозиции гауссианов
- Обобщающая способность модели
 - o Методы проверки
- Постановка формальной задачи машинного обучения
 - o Аппроксимация эмпирического риска
- Линейная регрессия
 - o Обобщённые линейные модели
- Методы численной оптимизации
- Оценка качества модели
 - o ROC, ROC AUC
- Отбор признаков. Метод главных компонент.
- Линейная классификация
 - o Метод опорных векторов
 - o Задача оценивания вероятностей, логистическая регрессия
- Метрики качества в задачах классификации.
- Отбор и деколениаризация признаков. Метод главных компонент
- Структура нейронной сети.
 - o Обратное распространение ошибки.
 - o Полносвязные нейронные сети
- Предобработка данных
 - o Нормализация
 - o Декорреляция
 - o Batch normalization
- Методы регуляризации
 - o впрыскивание шума
 - o L1, L2, ...
 - o connection dropping
 - o dropout

- Свёрточные сети
 - o Архитектура
 - o padding, strides, pooling
 - o предобученные сети
- Автокодировщики
 - o применения
- Общий алгоритм построения решающих деревьев
 - o критерии информативности
 - o Конкретные критерии для классификации и регрессии
 - o обработка пропущенных значений,
- Улучшение обобщающей способности решающих деревьев
 - o стрижка
 - o регуляризация
- Ансамбли моделей
 - o Бэггинг, бустинг
 - o AdaBoost
 - o случайный лес
- Градиентный бустинг над решающими деревьями

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 по билетам. В каждом билете представлено два теоретических вопроса. При проведении зачёта обучающемуся предоставляется 30 минут на подготовку.