

**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:	Применение методов машинного обучения для анализа экспериментальных данных
по направлению:	Прикладные математика и физика
профиль подготовки:	Общая и прикладная физика Физтех-школа физики и исследований им. Ландау кафедра электродинамики сложных систем и нанопотоники
курс:	1
квалификация:	магистр

Семестры, формы промежуточной аттестации:

- 1 (осенний) - Дифференцированный зачет
- 2 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

- лекции: 30 час.
- семинары: 30 час.
- лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Всего часов: 90, всего зач. ед.: 2

Программу составил: З.М. Шибзухов, д-р физ.-мат. наук, доцент

Программа обсуждена на заседании кафедры электродинамики сложных систем и нанопотоники 01.04.2024

Аннотация

Курс предназначен для освоения методов и алгоритмов машинного обучения и анализа данных (МОАД), ориентированных на решение прикладных задач. Он направлен на достижение понимания основных движущих идей и эвристик, освоение основных прикладных математических методов и алгоритмов МОАД, освоение базовых практических навыков решения прикладных задач с помощью методов МОАД на базе Python и его окружения.

Курс включает в себя лекции, излагающий теоретический материал и практические/лабораторные занятия.

1. Цели и задачи

Цель дисциплины

- освоение теоретических знаний и практических навыков применения алгоритмов МОАД, достижение понимания основных движущих идей и эвристик для осознанного применения на практике и способности к адаптации алгоритмов МОАД в случае практической необходимости. Курс направлен на освоение материала по принципу «от простого к более сложному» с учетом логики исторического развития и усложнения методов и алгоритмов МОАД. Курс включает в себя классические и современные методы решения задач регрессии, классификации, кластеризации, декомпозиции. Он охватывает широкий ряд моделей, начиная от простых линейных до сложных многослойных нейронных сетей и композиций алгоритмов.

Задачи дисциплины

- познакомить студентов с методами и задачами машинного обучения и анализа данных и научить применять полученные знания в будущих прикладных исследованиях.

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-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области своей профессиональной деятельности, способен на научном языке формулировать профессиональные задачи	ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации
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-3.2 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
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
ПК-1 Способен ставить, формализовывать и решать задачи, в том числе разрабатывать и исследовать математические модели	ПК-1.2 Способен выдвигать гипотезы, строить математические модели для описания изучаемых явлений и процессов, оценивать качество разработанной модели

изучаемых явлений и процессов, системно анализировать научные проблемы, получать новые научные результаты	ПК-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
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:

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

уметь:

- самостоятельно выбирать метод машинного обучения и анализа данных, соответствующий исследовательской задаче, интерпретировать полученные результаты.

владеть:

- навыками самостоятельного обучения, чтения и анализа академической литературы по применению методов машинного обучения и анализа данных, 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	Исторический экскурс и основные понятия	2	1		1
2	Линейная регрессия	4	4		4
3	Линейное разделение классов	4	4		4
4	Метод главных компонент и полная линейная регрессия	2	2		2
5	Кластеризация на основе поиска центров кластеров	2	2		2
6	Метрические алгоритмы регрессии и классификации	1	2		2
7	Логические методы и деревья решений	2	2		2
8	Оценка обобщающей способности и параметров алгоритмов обучения	2	2		2
9	Нелинейные модели и простые нейронные сети	2	2		2
10	Сигма-пи нейронные сети и конструктивное обучение	2	2		2
11	Многослойные нейронные сети	4	4		4
12	Методы бустинга и баггинга	3	3		3

Итого часов	30	30		30
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. Исторический экскурс и основные понятия

Исторический экскурс. Основные понятия. Обучение по конечным наборам прецедентов. Модели алгоритмов и методы обучения. Оценка качества алгоритма и метода обучения. Проблема переобучения. Агрегирующие функции и функционалы качества.

2. Линейная регрессия

Линейная регрессия. Применение принципа минимизации эмпирического риска. Применение принципа регуляризации для повышения устойчивости. Алгоритмы обучения для линейной регрессии. Робастные варианты линейной регрессии. Применение метода итеративного перевзвешивания.

3. Линейное разделение классов

Линейное разделение классов. Применение принципа максимизации отступа и минимизации эмпирического риска. Применение принципа регуляризации для повышения устойчивости. Алгоритмы линейного разделения классов. Робастные варианты линейного разделения классов. Применение метода итеративного перевзвешивания.

4. Метод главных компонент и полная линейная регрессия

Постановка задачи и алгоритм поиска главных компонент. Применение для понижения размерности. Алгоритм поиска полной линейной регрессии. Варианты обобщений. Робастные варианты метода главных компонент и полной линейной регрессии.

5. Кластеризация на основе поиска центров кластеров

Классический вариант метода поиска центров кластеров. Алгоритм k-means и его варианты. Использование расстояния Махаланобиса. Робастные варианты метода поиска центров кластеров. Модели смесей для поиска кластеров.

6. Метрические алгоритмы регрессии и классификации

Метод скользящего контроля и его варианты. Поиск параметров алгоритмов обучения.

Семестр: 2 (Весенний)

7. Логические методы и деревья решений

Логические решающие правила. Построение И-ИЛИ деревья решений. Леса деревьев решений.

8. Оценка обобщающей способности и параметров алгоритмов обучения

Метод K ближайших соседей для решения задач регрессии и классификации. Ядерный метод. LOWESS.

9. Нелинейные модели и простые нейронные сети

Квазилинейные нейроны и их обучение. Линейные и квазилинейные комбинации простых нейронов и их обучение. Робастные варианты обучения простых нейронных сетей.

10. Сигма-пи нейронные сети и конструктивное обучение

Сигма-пи нейроны и их варианты. Конструктивное обучение сигма-пи нейронов. Обучение ансамблей из сигма-пи нейронов. Корректные операции для построения ансамблей сигма-пи нейронов.

11. Многослойные нейронные сети

Многослойные сети из нейронов. Общий метод обратного распространения. Робастные варианты метода обратного распространения. Применение метода итеративного перевзвешивания.

12. Методы бустинга и баггинга

Методы градиентного бустинга для задач регрессии и классификации. Методы баггинга. Методы повышения робастности методов градиентного бустинга.

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

Для проведения лабораторных занятий необходим компьютерный класс. Необходимо иметь установленный вариант экосистемы Python на компьютерах.

6. Перечень рекомендуемой литературы

Основная литература

Фонд базовой кафедры:

1. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2014. — 739 p.
2. Bishop C.M. Pattern Recognition and Machine Learning. — Springer, 2006. — 738 p.
3. Мерков А.Б. Распознавание образов. Введение в методы статистического обучения. 2011. 256 с.
4. Мерков А.Б. Распознавание образов. Построение и обучение вероятностных моделей. 2014. 238 с.

Дополнительная литература

Фонд базовой кафедры:

1. Коэльо Л.П., Ричарт В. Построение систем машинного обучения на языке Python. 2016. 302 с.
2. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. 2023. 401 с.

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

1. Курс лекций Воронцова К.В. <https://bit.ly/ML-Vorontsov>
2. Курс лекций Золотых Н.Ю. <http://www.uic.unn.ru/~zny/ml/>

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

Графический планшет для использования в качестве «доски», google meet для трансляции в интернете, google collab для лабораторной работы в облачном формате.

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 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 кафедра электродинамики сложных систем и нанофотоники
курс:	<u>1</u>
квалификация:	магистр

Семестры, формы промежуточной аттестации:

- 1 (осенний) - Дифференцированный зачет
- 2 (весенний) - Дифференцированный зачет

Разработчик: З.М. Шибзухов, д-р физ.-мат. наук, доцент

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-1.2 Осуществляет поиск вариантов решения поставленной проблемной ситуации на основе доступных источников информации
ОПК-1 Владеет системой фундаментальных научных знаний в области физико-математических наук	ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные научные знания в области физико-математических наук
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области своей профессиональной деятельности, способен на научном языке формулировать профессиональные задачи	ОПК-2.3 Владеет профессиональной терминологией, используемой в современной научно-технической литературе, обладает навыками устного и письменного изложения результатов научной деятельности в рамках профессиональной коммуникации
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-3.2 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
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области физико-математических наук и информационно-коммуникационных технологий	ОПК-4.2 Способен применять знания в области физико-математических наук для решения поставленной задачи, формулирования выводов и оценки полученных результатов
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-1.3 Способен применять теоретические и (или) экспериментальные методы исследований к конкретной научной задаче и интерпретировать полученные результаты
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 «Применение методов машинного обучения для анализа экспериментальных данных» обучающийся должен:

знать:

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

уметь:

- самостоятельно выбирать метод машинного обучения и анализа данных, соответствующий исследовательской задаче, интерпретировать полученные результаты.

владеть:

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

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

С целью контроля освоения обучающимися учебного материала проводится устный опрос в начале занятия по теме прошлого занятия.

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

Перечень контрольных вопросов 1 семестр

1. Постановка задачи обучения по конечному набору прецедентов.
2. Метод наименьших квадратов и его регуляризованные версии для построения линейной регрессии.
3. Методы повышения робастности метода наименьших квадратов для построения линейной регрессии.
4. Применение принципа минимизации эмпирического риска для решения задачи классификации.
5. Метод скользящего контроля и его применение.
6. Метод главных компонент.
7. Алгоритм k-means для поиска центров кластеров.
8. Методы повышения робастности линейного разделения классов.
9. Конструктивный алгоритм для классификации по булевым признакам.
10. Модель смесей для поиска кластеров.
11. Метод обратного распространения для решения задачи регрессии при помощи нейронной сети с одним скрытым слоем.

Пример билета для дифф. зачета в 1 семестре:

Билет 1.

1. Постановка задачи обучения по конечному набору прецедентов.
2. Алгоритм k-means для поиска центров кластеров.

Примеры контрольных вопросов 2 семестр

1. Построить линейную регрессию для набора данных animals, идентифицировать выбросы, перестроить линейную регрессию по данным, не содержащим выбросы.
2. Разделить два класса из набора данных iris при помощи линейной гиперплоскости. Оценить качество разделения.
3. Оценить обобщающую способность линейного разделения двух классов для заданного метода обучения.
4. Найти главные компоненты и сократить размерность данных из набора iris и решить задачу линейного разделения двух классов.
5. Построить нелинейную поверхность (нейронная сеть с одним скрытым слоем), разделяющую классы для набора iris.
6. Построить линейную регрессию для набора данных starsCYG, идентифицировать выбросы, перестроить линейную регрессию по данным, не содержащим выбросы.
7. Разделить два класса из набора данных wine при помощи линейной гиперплоскости. Оценить качество разделения.
8. Оценить обобщающую способность линейного разделения двух классов для заданного метода обучения.
9. Найти главные компоненты и сократить размерность данных из набора wine и решить задачу линейного разделения двух классов.
10. Построить нелинейную поверхность (нейронная сеть с одним скрытым слоем), разделяющую классы для набора wine,

Пример билета для дифф. зачета во 2 семестре:

Билет 1.

1. Разделить два класса из набора данных iris при помощи линейной гиперплоскости. Оценить качество разделения.
2. Построить нелинейную поверхность (нейронная сеть с одним скрытым слоем), разделяющую классы для набора iris.

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 зачет проводится в устной форме по билетам. В каждом билете представлено два теоретических вопроса. При проведении зачета обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося не должен превышать одного астрономического часа.