

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

УТВЕРЖДЕНО

**Директор физтех-школы
прикладной математики и
информатики**

А.М. Райгородский

	Рабочая программа дисциплины (модуля)
по дисциплине:	Продвинутое машинное обучение
по направлению:	Прикладная математика и информатика
профиль подготовки:	Комбинаторика и цифровая экономика центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" кафедра дискретной математики
курс:	2
квалификация:	магистр

Семестр, формы промежуточной аттестации: 4 (весенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 30 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Программу составил: Р.Г. Нейчев, профессор

Программа обсуждена на заседании кафедры дискретной математики 30.01.2025

Аннотация

Этот курс призван познакомить студентов с современным состоянием машинного обучения и искусственного интеллекта. Он сочетает в себе теоретические основы алгоритмов машинного обучения с комплексными практическими заданиями. Курс охватывает материалы от классических алгоритмов до подходов к глубокому обучению и последних достижений в области искусственного интеллекта.

1. Цели и задачи

Цель дисциплины

- Изучите основные теоретические основы машинного обучения и глубокого обучения.
- Ознакомьтесь с различными подходами к контролируемому и неконтролируемому обучению.
- Получите необходимый опыт в области предварительной обработки данных, разработки моделей, подгонки и проверки.

Задачи дисциплины

- Предварительная обработка данных, разработка моделей, подгонка и проверка.
- Навыки, необходимые для разработки продуктов и прикладных исследований.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов
ОПК-2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач	ОПК-2.1 Имеет представление о современном состоянии математических исследований в рамках тематической области своей профессиональной деятельности
ОПК-3 Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	ОПК-3.1 Умеет анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

- основные принципы и проблематику теории обучения машин;
- основные методы и алгоритмы решения задач обучения по прецедентам;
- основные области применения этих методов и алгоритмов;
- классификации, кластеризации и регрессии.

уметь:

- формализовать постановки прикладных задач анализа данных;
- использовать методы обучения по прецедентам для решения практических задач;
- оценивать точность и эффективность полученных решений.

владеть:

- основными понятиями теории машинного обучения;
- навыками самостоятельной работы при решении типовых задач;
- культурой постановки и моделирования практически значимых задач;
- навыками теоретического анализа реальных задач, решаемых с помощью алгоритмов обучения по прецедентам.

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, knn, наивный Байес	6	6		9
2	СВМ, ПКА	6	6		9
3	Повышение градиента	6	6		9
4	Оптимизация, регуляризация в DL	6	6		9
5	Векторизация текста, встраивание, автокодировщики	6	6		9
Итого часов		30	30		45
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

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

Семестр: 4 (Весенний)

1. Intro, knn, наивный байесовский

Линейная регрессия. Градиентный спуск Логистическая регрессия

2. SVM, PCA

Декомпозиция отклонения смещения, тестовая среда для проверки правильности смещения. Деревья и ансамбли.

3. Повышение градиента

Типы функций, отсутствующие значения, важность функций. Основы нейронных сетей

4. Оптимизация, регуляризация в DL

Рекуррентные нейронные сети Сверточные нейронные сети

5. Векторизация текста, вложения, автоэнкодеры.

Обучение без учителя. Кластеризация

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

Стандартная аудитория

6.Перечень рекомендуемой литературы

Основная литература

1. Машинное обучение [Текст]/Х. Бринк, Дж. Ричардс, М. Феверолф, Real-World Machine Learning, -СПб., Питер, 2017

2. Python и машинное обучение [Текст] = Python Machine Learning : крайне необходимое издание по новейшей предсказательной аналитике для более глубокого понимания методологии машинного обучения / С. Рашка; пер. с англ. А. В. Логунова. — М. : ДМК Пресс, 2017. — 418 с.: ил. - Предм. указ.: с. 408-417. - 200 экз. - ISBN 978-5-97060-409-0 (в пер.) .— Полный текст (Доступ из сети МФТИ / Удаленный доступ).

Дополнительная литература

1. Математические основы машинного обучения и прогнозирования [Текст] / В. В. Вьюгин ; Моск. физ.-техн. ин-т (гос. ун-т), Лаб. структурных методов анализа данных в предсказательном моделировании (ПреМоЛаб), Ин-т проблем передачи информации им. А. А. Харкевича РАН - М.МЦНМО,2013
2. Машинное обучение: новый искусственный интеллект [Текст]/Э. Алпайдин, -М., Изд. группа "Точка", 2017

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

<http://dm.fizteh.ru/>

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

Мультимедийные технологии можно использовать на лекциях и практических занятиях, в том числе на презентациях.

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

Студент, изучающий дисциплину, должен, с одной стороны, владеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения, понятия, аксиомы.

Для успешного освоения курса требуется напряженная самостоятельная работа студента. Программа курса предусматривает минимально необходимое время для работы студента над темой. Самостоятельная работа включает:

- чтение и запись рекомендованной литературы;
- изучение учебного материала (по конспектам, учебной и научной литературе), подготовка ответов на вопросы, предназначенные для самостоятельного изучения, подтверждение индивидуальных утверждений, свойств;
- лабораторные работы для понимания связи между теорией и практическими навыками;
- подготовка к экзамену.

Управление и контроль самостоятельной работы студента осуществляется в форме индивидуальных консультаций.

Важно понимать изучаемый материал, а не его механическое запоминание. Если сложно изучить отдельные темы, вопросы, следует обратиться к преподавателю.

ПРИЛОЖЕНИЕ

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению: Прикладная математика и информатика
профиль подготовки: Комбинаторика и цифровая экономика
центр дополнительного, дополнительного профессионального и
онлайн-образования "Пуск"
кафедра дискретной математики
курс: 2
квалификация: магистр
Семестр, формы промежуточной аттестации: 4 (весенний) - Экзамен
Разработчик: Р.Г. Нейчев, профессор

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов
ОПК-2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач	ОПК-2.1 Имеет представление о современном состоянии математических исследований в рамках тематической области своей профессиональной деятельности
ОПК-3 Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	ОПК-3.1 Умеет анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения

2. Показатели оценивания компетенций

В результате изучения дисциплины «Продвинутое машинное обучение» обучающийся должен:

знать:

- основные принципы и проблематику теории обучения машин;
- основные методы и алгоритмы решения задач обучения по прецедентам;
- основные области применения этих методов и алгоритмов;
- классификации, кластеризации и регрессии.

уметь:

- формализовать постановки прикладных задач анализа данных;
- использовать методы обучения по прецедентам для решения практических задач;
- оценивать точность и эффективность полученных решений.

владеть:

- основными понятиями теории машинного обучения;
- навыками самостоятельной работы при решении типовых задач;
- культурой постановки и моделирования практически значимых задач;
- навыками теоретического анализа реальных задач, решаемых с помощью алгоритмов обучения по прецедентам.

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

1. Постановка задачи контролируемого обучения. Проблемы регрессии и классификации. Какая разница?
2. Линейная модель для задачи регрессии в матричной записи. Функция потерь среднеквадратичной ошибки.
3. Что такое градиент? Как это используется в оптимизации?
4. Запишите шаг градиентного спуска для линейной модели и MSE для одномерного случая.
5. Что такое проверка? Перекрестная проверка?
6. Что такое регуляризация? Чем регуляризация L1 отличается от L2?
7. Что такое показатели точности и отзыва?
8. Как работает упаковка? Что такое случайный лес? В чем разница между Random forest и Bagging?
9. Чем параметры отличаются от гиперпараметров? Например, что такое параметры в линейных моделях и деревьях решений? Гиперпараметры?
10. Что такое повышение? Повышение градиента? Как следует обучать модель на шаге $t + 1$ в ансамбле повышения градиента?
11. Что такое обратное распространение? Как это работает?
12. Как градиент будет распространяться через линейный слой? Через ReLU?
13. Как работает сверточный слой? Какие ядра (фильтры) в сверточном слое? Они независимы?

14. Что такое отсев? Как это работает в нейронной сети? Меняет ли он свое поведение на этапе вывода (тестирования)?
15. Что такое пакетная нормализация? Как это работает? Как это влияет на скорость обучения? Меняет ли он свое поведение на этапе вывода (тестирования)?
16. Сформулируйте неконтролируемую формулировку проблемы. Что такое кластеризация? Как работает алгоритм k-средних?

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

Вопросы к экзамену

1. Докажите, что если m , n - два взаимно простых целых числа разной четности, то числа $m^2 - n^2$ и $2mn$ также взаимно просты.
2. Напишите и докажите общую формулу для количества различных представлений данного целого числа n в виде суммы двух квадратов. (Представители, которые не получены друг от друга путем изменения знаков и порядка слов, считаются разными.)
3. На основе полученной формулы выведите нижнюю границу максимального числа равных расстояний между заданными n точками на плоскости, используя правильную прямоугольную решетку.
4. Постройте правильный пятиугольник с помощью циркуля и линейки.
5. Постройте правильный 15-угольник, используя циркуль и линейку.
6. Вам дается один сегмент. Требуется построить с помощью циркуля и линейки отрезок длины x , удовлетворяющий уравнению
7. Основываясь на предыдущем задании, докажите, что правильный семиугольник нельзя построить с помощью циркуля и линейки.
8. Докажите, что трисекция угла невозможна.
9. Опишите все возможные комбинации количества черных и белых шаров в урне для голосования, чтобы при случайном вылове двух шаров в выборке без возврата, вероятность вылова двух белых шаров составляла точно 0,5.
10. Рассмотрим соотношение сторон a , b , c треугольника, в котором треугольник с вершинами в основании биссектрис равнобедренный. Предполагая, что стороны, сходящиеся на стороне с большого треугольника, равны, сведем это соотношение к следующему
11. Далее мы рассматриваем куб, определяемый первым из трех уравнений (отказ от требования, чтобы a , b , c были сторонами треугольника). Покажите, что полученный куб неразложим, то есть определяющий его многочлен не учитывается.
12. В дополнение к этому, покажите, что наш куб неособен, то есть на его проективизации нет ни одной точки, в которой каждое направление касалось бы (или того же самого, в котором все три первые частные производные многочлена, определяющего его, вырождаются.).

Примеры экзаменационных билетов

Билет №1

1. Напишите и докажите общую формулу для количества различных представлений данного целого числа n в виде суммы двух квадратов.
2. Докажите, что трисекция угла невозможна.

Билет №2

1. Рассмотрим соотношение сторон a , b , c треугольника, в котором треугольник с вершинами в основании биссектрис равнобедренный.
2. Опишите всевозможные комбинации чисел черных и белых шаров в урне для голосования так, чтобы, если два шара случайно выловлены в выборке и не вернулись, вероятность вылова двух белых шаров была ровно 0,5.

Критерии оценивания

Оценка «отлично (10)» выставляется студенту, который проявил всестороннее, систематическое и глубокое знание материала образовательной программы, самостоятельно выполнил все задачи, предусмотренные программой, глубоко изучил основную и дополнительную литературу, рекомендованную программой. , активно работает в классе и понимает основные научные концепции по изучаемой дисциплине, проявил творческий подход и научный подход в понимании и представлении материала образовательной программы, ответ на который характеризуется использованием богатых и адекватных терминов, а также последовательным и логичным изложением материала;

Оценка «отлично (9)» дается студенту, который продемонстрировал всестороннее систематическое знание материала образовательной программы, самостоятельно выполнил все задачи, предусмотренные программой, глубоко усвоил основную литературу и знаком с рекомендуемой дополнительной литературой. по программе, активно проработал на занятиях, показал системность знаний по дисциплине, достаточную для дальнейшего изучения, а также умение самостоятельно расширять ее, ответ которой отличается точностью используемых терминов, а изложение материала в нем последовательное и логичное;

Оценка «отлично (8)» выставляется студенту, который проявил полное знание материала образовательной программы, не допускает существенных неточностей в своем ответе, самостоятельно выполнил все задания, предусмотренные программой, изучил основную литературу, рекомендованную учебной программой. программа, активно проработанная на занятиях, показала системность его знаний по дисциплине, достаточных для дальнейшего изучения, а также способность самостоятельно их расширять;

Оценка «хорошо (7)» выставляется студенту, который проявил достаточно полное знание материала образовательной программы, не допускает существенных неточностей в ответе, самостоятельно выполнил все задания, предусмотренные программой, изучил основную рекомендованную литературу по программе, активно работал на занятиях, проявил системность своих знаний по дисциплине, достаточных для дальнейшего изучения, а также способность самостоятельно их усиливать;

Оценка «хорошо (6)» выставляется студенту, который проявил достаточно полное знание материала образовательной программы, не допускает существенных неточностей в своем ответе, самостоятельно выполнил основные задачи, предусмотренные программой, изучил основную литературу. рекомендован программой, показал систематичность своих знаний по дисциплине, достаточную для дальнейшего изучения;

Оценка «хорошо (5)» дается студенту, продемонстрировавшему знание материала основной образовательной программы в объеме, необходимом для дальнейшего обучения и будущей работы по профессии, который, не проявляя достаточной активности на уроках, тем не менее самостоятельно выполнял овладел основными задачами, предусмотренными программой, освоил основную литературу, рекомендованную программой, допустил ошибки в их выполнении и ответе во время тестирования, но имеет необходимые знания для исправления этих ошибок самостоятельно;

Оценка «удовлетворительно (4)» дается студенту, обнаружившему знание материала основной образовательной программы в объеме, необходимом для дальнейшего обучения и будущей работы по профессии, который, не проявляя достаточной активности на уроках, тем не менее самостоятельно выполнял основные задачи, предусмотренные программой, изучил основную литературу, но допустил ошибки в их выполнении и в своем ответе во время теста, но имеет необходимые знания для исправления этих ошибок под руководством преподавателя;

Оценка «удовлетворительно (3)» выставляется обучающемуся, проявившему знание материала основной образовательной программы в объеме, необходимом для дальнейшего обучения и будущей работы по профессии, не проявившего активности на занятиях, самостоятельно выполнившего основные задания, предусмотренные законодательством. программа, но допускающая ошибки в их выполнении и в ответе во время теста, но обладающая необходимыми знаниями для устранения под руководством преподавателя наиболее существенных ошибок;

Оценка «неудовлетворительно (2)» дается студенту, который показал пробелы в знаниях или недостаток знаний по значительной части материала основной образовательной программы, не выполнил самостоятельно основные задачи, требуемые программой, допустил принципиальные ошибки в выполнении предусмотренных программой задач, не имеющего возможности продолжить учебу или начать профессиональную деятельность без дополнительной подготовки по данной дисциплине;

Оценка «неудовлетворительно (1)» ставится студенту при отсутствии ответа (отказ от ответа), либо когда представленный ответ совсем не соответствует сути вопросов, содержащихся в задании.

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

Во время экзамена студенту разрешается использовать программу дисциплины.