

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

УТВЕРЖДЕНО
Заместитель директора

Ю.О. Соболев

	Рабочая программа дисциплины (модуля)
по дисциплине:	Современные модели машинного обучения
по направлению:	Прикладная математика и информатика
профиль подготовки:	Науки о данных
	центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
	центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
курс:	2
квалификация:	магистр

Семестр, формы промежуточной аттестации: 3 (осенний) - Экзамен

Аудиторных часов: 22 всего, в том числе:

лекции: 2 час.

семинары: 20 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 83 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Программу составили:

А.В. Филатов, преподаватель

С.В. Доронин, преподаватель

К.А. Лапин, старший методист

Программа обсуждена на заседании центра дополнительного, дополнительного профессионального и
онлайн-образования "Пуск" 01.03.2025

Аннотация

В курсе приводятся примеры прикладных задач классификации, регрессии, ранжирования. Рассматриваются методы прогнозирования временных рядов, поиск ассоциативных правил, нейронные сети глубокого обучения, эвристические, стохастические, нелинейные композиции.

Также кратко обсуждаются некоторые вопросы методологии машинного обучения: особенности реальных данных, межотраслевой стандарт CRISP-DM, организация вычислительных экспериментов. Выделяются многочисленные сходства и взаимосвязи между различными методами машинного обучения, обсуждаются идеи разнообразных гибридных подходов.

1. Цели и задачи

Цель дисциплины

- сформировать теоретические и практические знания в области обучения машин, современных методов восстановления зависимостей по эмпирическим данным, включая дискриминантный, кластерный и регрессионный анализ.

Задачи дисциплины

- освоить методы корректной формулировки задач в терминах машинного обучения;
- овладеть навыками практического решения задач интеллектуального анализа данных.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1 Умеет решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности
	УК-6.2 Оценивает свою деятельность, соотносит цели, способы и средства выполнения деятельности с её результатами
ОПК-1 Способен решать актуальные задачи фундаментальной и прикладной математики	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области прикладной математики и информатики и способен их применять при решении задач профессиональной деятельности
ОПК-3 Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, задач, понимает и учитывает на практике границы применимости получаемых решений
	ОПК-3.2 Владеет исследовательскими методами и способен использовать их при решении новых задач, применяя знания из различных областей науки (техники)
ОПК-4 Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности	ОПК-4.1 Умеет применять информационно-коммуникационные технологии для поиска и анализа профессиональной информации, выделения в ней главного, структурирования, оформления и представления в виде аналитических обзоров с обоснованными выводами и рекомендациями
	ОПК-4.3 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
	ОПК-6.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения

ОПК-6 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области информатики и вычислительной техники, учитывая особенности и ограничения различных методов решения	ОПК-6.2 Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
	ОПК-6.5 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте

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	Диффузионные модели	1	7		33
2	Мультимодальные и мультизадачные модели	1	6		32
3	Работа с современными языковыми моделями		7		18
Итого часов		2	20		83
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

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

Семестр: 3 (Осенний)

1. Диффузионные модели

Что такое генеративные модели. Диффузионные модели. Модификации диффузионных моделей. Современные диффузионные модели. Применения диффузионных моделей в различных задачах.

2. Мультимодальные и мультизадачные модели

Мультимодальность и мультизадачность. Fusion-механизмы. Построение картиночно-текстовых моделей. Современные мультимодальные модели.

3. Работа с современными языковыми моделями

Zero-/few-shot learning языковых моделей. Дообучение языковых моделей. Техники для улучшения языковых моделей.

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

Обучение проходит с использованием дистанционных образовательных технологий.

6. Перечень рекомендуемой литературы

Основная литература

1. Методы распознавания [Текст] : учеб. пособие / А. Л. Горелик, В. А. Скрипкин .— 4-е изд., испр. — М. : Высшая школа, 2004 .— 261 с.
2. Машинное обучение [Текст], / Х. Бринк, Дж. Ричардс, М. Феверолф, СПб., Питер, 2017

Дополнительная литература

1. Теория распознавания образов : Статистические методы [Текст] : учеб. пособие для вузов / А. А. Натан ; М-во высш. и средн. спец. образования РСФСР, Моск. физ.-техн. ин-т .— М. : МФТИ, 1988 .— 84 с. : ил. - 400 экз.

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

1. <http://www.machinelearning.ru> – профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных.
2. <http://shad.yandex.ru> – сайт школы анализа данных Яндекса.
3. http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%28%D0%BA%D1%83%D1%80%D1%81_%D0%BB%D0%B5%D0%BA%D1%86%D0%B8%D0%B9%2C_%D0%9A.%D0%92.%D0%92%D0%BE%D1%80%D0%BE%D0%BD%D1%86%D0%BE%D0%B2%29

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

Образовательная платформа

Webinar.ru

Zoom

Google Drive

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 дисциплины, 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:	Прикладная математика и информатика
профиль подготовки:	Науки о данных центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
курс:	<u>2</u>
квалификация:	магистр

Семестр, формы промежуточной аттестации: 3 (осенний) - Экзамен

Разработчики:

А.В. Филатов, преподаватель
С.В. Доронин, преподаватель
К.А. Лапин, старший методист

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1 Умеет решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности
	УК-6.2 Оценивает свою деятельность, соотносит цели, способы и средства выполнения деятельности с её результатами
ОПК-1 Способен решать актуальные задачи фундаментальной и прикладной математики	ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности
	ОПК-1.3 Понимает междисциплинарные связи в области прикладной математики и информатики и способен их применять при решении задач профессиональной деятельности
ОПК-3 Способен разрабатывать математические модели и проводить их анализ при решении задач в области профессиональной деятельности	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, задач, понимает и учитывает на практике границы применимости получаемых решений
	ОПК-3.2 Владеет исследовательскими методами и способен использовать их при решении новых задач, применяя знания из различных областей науки (техники)
ОПК-4 Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности	ОПК-4.1 Умеет применять информационно-коммуникационные технологии для поиска и анализа профессиональной информации, выделения в ней главного, структурирования, оформления и представления в виде аналитических обзоров с обоснованными выводами и рекомендациями
	ОПК-4.3 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
ОПК-6 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области информатики и вычислительной техники, учитывая особенности и ограничения различных методов решения	ОПК-6.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения
	ОПК-6.2 Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
	ОПК-6.5 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте

2. Показатели оценивания компетенций

В результате изучения дисциплины «Современные модели машинного обучения» 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

С целью контроля освоения обучающимися учебного материала проводится устный опрос в начале занятия по теме прошлой лекции или в конце занятия по пройденной теме.

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

Вопросы для сдачи экзамена:

1. Задача прогнозирования временных рядов. Примеры приложений.
2. Экспоненциальное скользящее среднее. Модель Хольта. Модель Тейла-Вейджа. Модель Хольта-Уинтерса.
3. Адаптивная авторегрессионная модель.
4. Следящий контрольный сигнал. Модель Тригга-Лича.
5. Адаптивная селективная модель. Адаптивная композиция моделей.
6. Локальная адаптация весов с регуляризацией.
7. Понятие ассоциативного правила и его связь с понятием логической закономерности.
8. Примеры прикладных задач: анализ рыночных корзин, выделение терминов и тематики текстов.
9. Алгоритм APriori. Два этапа: поиск частых наборов и рекурсивное порождение ассоциативных правил. Недостатки и пути усовершенствования алгоритма APriori.
10. Алгоритм FP-growth. Понятия FP-дерева и условного FP-дерева. Два этапа поиска частых наборов в FP-growth: построение FP-дерева и рекурсивное порождение частых наборов.
11. Общее представление о динамических и иерархических методах поиска ассоциативных правил.
12. Свёрточные нейронные сети (CNN). Свёрточный нейрон. Pooling нейрон. Выборка размеченных изображений ImageNet.
13. Свёрточные сети для сигналов, текстов, графов, игр.
14. Рекуррентные нейронные сети (RNN). Обучение рекуррентных сетей: Backpropagation Through Time (BPTT).
15. Сети долгой кратковременной памяти (Long short-term memory, LSTM).
16. Рекуррентная сеть Gated Recurrent Unit (GRU).
17. Автокодировщики. Векторные представления дискретных данных.
18. Перенос обучения (transfer learning).
19. Самообучение (self-supervised learning).
20. Генеративные состязательные сети (GAN, generative adversarial net).
21. Стохастические методы: бэггинг и метод случайных подпространств.
22. Простое голосование (комитет большинства). Алгоритм ComBoost. Идентификация нетипичных объектов (выбросов).
23. Преобразование простого голосования во взвешенное.
24. Обобщение на большое число классов.
25. Случайный лес.
26. Анализ смещения и вариации для простого голосования.
27. Смесь алгоритмов (квазилинейная композиция), область компетентности, примеры функций компетентности.

28. Выпуклые функции потерь. Методы построения смесей: последовательный и иерархический.
29. Построение смеси алгоритмов с помощью ЕМ-подобного алгоритма.
30. Постановка задачи обучения ранжированию. Примеры.
31. Признаки в задаче ранжирования поисковой выдачи: текстовые, ссылочные, кликовые. TF-IDF. PageRank.
32. Критерии качества ранжирования: Precision, MAP, AUC, DCG, NDCG, pFound.
33. Ранговая классификация, ОС-SVM.
34. Попарный подход: RankingSVM, RankNet, LambdaRank.
35. Задачи коллаборативной фильтрации, транзакционные данные и матрица субъекты - объекты.
36. Корреляционные методы user-based, item-based. Задача восстановления пропущенных значений. Меры сходства субъектов и объектов.
37. Латентные методы на основе би-кластеризации. Алгоритм Брегмана.
38. Латентные методы на основе матричных разложений. Метод главных компонент для разреженных данных (LFM, Latent Factor Model). Метод стохастического градиента.
39. Неотрицательные матричные разложения. Метод чередующихся наименьших квадратов ALS.
40. Модель с учётом неявной информации (implicit feedback).
41. Рекомендации с учётом дополнительных признаков данных. Линейная и квадратичная регрессионные модели, libFM.
42. Измерение качества рекомендаций. Меры разнообразия (diversity), новизны (novelty), покрытия (coverage), догадливости (serendipity).
43. Задача тематического моделирования коллекции текстовых документов.
44. Вероятностный латентный семантический анализ PLSA. Метод максимума правдоподобия. ЕМ-алгоритм. Элементарная интерпретация ЕМ-алгоритма.
45. Латентное размещение Дирихле LDA. Метод максимума апостериорной вероятности. Сглаженная частотная оценка условной вероятности.
46. Небайесовская интерпретация LDA и её преимущества. Регуляризаторы разреживания, сглаживания, частичного обучения.
47. Аддитивная регуляризация тематических моделей. Регуляризованный ЕМ-алгоритм, теорема о стационарной точке (применение условий Каруша–Куна–Таккера).
48. Рациональный ЕМ-алгоритм. Онлайн-алгоритм и его распараллеливание.
49. Мультимодальная тематическая модель.
50. Регуляризаторы классификации и регрессии.
51. Регуляризаторы декоррелирования и отбора тем.
52. Внутренние и внешние критерии качества тематических моделей.
53. Задача о многоруком бандите. Жадные и эпсилон-жадные стратегии. Метод UCB (upper confidence bound). Стратегия Softmax.
54. Среда для экспериментов.
55. Адаптивные стратегии на основе скользящих средних. Метод сравнения с подкреплением. Метод преследования.
56. Постановка задачи в случае, когда агент влияет на среду. Ценность состояния среды. Ценность действия.
57. Жадные стратегии максимизации ценности. Уравнения оптимальности Беллмана.
58. Метод временных разностей TD. Метод Q-обучения.
59. Градиентная оптимизация стратегии (policy gradient). Связь с максимизацией log-правдоподобия.
60. Постановка задачи при наличии информации о среде в случае выбора действия. Контекстный многорукий бандит.
61. Линейная регрессионная модель с верхней доверительной оценкой LinUCB.
62. Оценивание новой стратегии по большим историческим данным.
63. Постановка задачи машинного обучения. Основные стратегии: отбор объектов из выборки и из потока, синтез объектов.
64. Сэмплирование по неуверенности. Почему активное обучение быстрее пассивного.
65. Сэмплирование по несогласию в комитете. Сокращение пространства решений.

66. Сэмплирование по ожидаемому изменению модели.
67. Сэмплирование по ожидаемому сокращению ошибки.
68. Синтез объектов по критерию сокращения дисперсии.
69. Взвешивание по плотности.
70. Оценивание качества активного обучения.
71. Введение изучающих действий в стратегию активного обучения. Алгоритмы ϵ -active и EG-active.
72. Применение обучения с подкреплением для активного обучения. Активное томпсоновское сэмплирование.

Примерный перечень билетов:

Билет №1

1. Метод временных разностей TD. Метод Q-обучения.
2. Градиентная оптимизация стратегии (policy gradient). Связь с максимизацией log-правдоподобия.

Билет №2

1. Постановка задачи при наличии информации о среде в случае выбора действия. Контекстный многорукий бандит.
2. Линейная регрессионная модель с верхней доверительной оценкой LinUCB.

Билет №3

1. Адаптивная авторегрессионная модель.
2. Выпуклые функции потерь. Методы построения смесей: последовательный и иерархический. Привести примеры.

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

Во время проведения экзамена обучающиеся могут пользоваться программой дисциплины, а также справочной литературой, вычислительной техникой, конспектами лекций.

Экзамен проводится письменно на lms. На экзамен отведено 4 академических часа.