

**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**

	<b>Рабочая программа дисциплины (модуля)</b>
<b>по дисциплине:</b>	Геометрические методы машинного обучения
<b>по направлению:</b>	Информатика и вычислительная техника
<b>профиль подготовки:</b>	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра проблем передачи информации и анализа данных
<b>курс:</b>	1
<b>квалификация:</b>	магистр

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

Аудиторных часов: 30 всего, в том числе:

лекции: 30 час.

семинары: 0 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 60 час.

Всего часов: 90, всего зач. ед.: 2

Программу составил: А.В. Бернштейн, д-р физ.-мат. наук, профессор

Программа обсуждена на заседании кафедры проблем передачи информации и анализа данных 09.01.2025

## Аннотация

Многие задачи машинного обучения имеют геометрическую природу. Общая цель машинного обучения - извлечь из данных ранее неизвестную информацию, которая отражается в структуре (базовой геометрии) данных. Поэтому понимание формы данных, особенно многомерных, играет важную роль в современной теории обучения и аналитике данных. Данные реального мира, полученные из естественных источников, обычно занимают лишь очень небольшую часть «пространства наблюдений» и концентрируются на структурах более низкой размерности, а геометрические методы позволяют обнаружить форму этих структур по заданным наборам данным.

Первоначально для исследования структур данных использовались линейные методы (такие как анализ главных компонент, анализ независимых компонент), основанные на различных методах проектирования на линейные подпространства меньшей размерности, а нелинейные геометрические методы использовались только для задач снижения размерности. Однако в настоящее время геометрические методы в машинном обучении становятся центральной методологией для раскрытия семантики информации из данных.

Цель курса — объяснить основные идеи и результаты использования современных геометрических подходов и методов для решения основных задач машинного обучения, таких как классификация, регрессия, снижение размерности, кластеризация и т. д. Значительная часть курса относится к наиболее популярной геометрической модели многомерных данных, называемой «моделью многообразия», в соответствии с которой многомерные данные лежат на (или вблизи) многообразия меньшей размерности. Раздел науки о данных, описываемых моделью многообразия, называют моделированием многообразий. Курс включает также топологический анализ данных, который активно используется в разведочном и интеллектуальном анализе данных и предоставляет набор топологических и геометрических инструментов для анализа многомерных, неполных и зашумленных данных. Курс включает в себя также результаты применения геометрических и топологических методов обучения к различным прикладным задачам. Необходимые краткие сведения по дифференциальной геометрии и топологии будут даны в курсе.

## 1. Цели и задачи

### Цель дисциплины

Цель курса — объяснить основные идеи, методы и результаты использования современных геометрических подходов для решения основных задач машинного обучения, таких как классификация, регрессия, снижение размерности, кластеризация и т. д.

### Задачи дисциплины

- геометрические подходы к анализу высокоразмерных данных в машинном обучении;
- описание и оценивание внутренней структуры высокоразмерных данных;
- современные модели и методы анализа высокоразмерных данных на основе моделирования многообразий;
- топологический анализ данных.

## 2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области информатики и вычислительной техники, способен на научном языке формулировать профессиональные задачи	ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности
	ОПК-2.2 Способен оценивать актуальность исследований в области информатики и вычислительной техники и их практическую значимость
ПК-2 Понимает и способен применить в научно-исследовательской и прикладной деятельности основные законы естествознания, современный математический аппарат и алгоритмы,	ПК-2.1 Знает основы научно-исследовательской деятельности в области информационных технологий, владеет знанием основ философии и методологии науки; знанием методов научных исследований и навыками их проведения

современные информационно-коммуникационные технологии	ПК-2.3 Имеет практический опыт научно-исследовательской деятельности в области информационно-коммуникационных технологий
ПК-1 Готов к включению в профессиональное сообщество; способен проводить под научным руководством локальные исследования на основе существующих методов в конкретной области профессиональной деятельности	ПК-1.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации; владеет навыками подготовки научных обзоров, публикаций, рефератов и библиографий по тематике проводимых исследований на русском и английском языке
	ПК-1.2 Умеет решать научные задачи с пониманием существующих подходов к верификации моделей программного обеспечения в связи с поставленной целью и в соответствии с выбранной методикой

### 3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- суть основных моделей, методов и алгоритмов геометрического анализа данных;
- основной доступный программный инструментарий, реализующий стандартные алгоритмы геометрического анализа данных и используемый для работы с массивами данных.

уметь:

- выявлять геометрическую природу различных задач машинного обучения и подходы к их решению;
- выбирать подходящие методы и алгоритмы при использовании геометрических подходов в решении задач машинного обучения;
- работать с научной литературой по использованию подходов и методов геометрического анализа данных для решения задач машинного обучения;
- предлагать и внедрять изученные методы для решения задач машинного обучения.

владеть:

- теоретическими основами и алгоритмическими приемами современного геометрического анализа данных;
- навыками работы с теоретическими и прикладными статьями, посвященных или использующих геометрическими методами анализа данных.

### 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	5			6
2	Линейные методы анализа данных	5			7
3	Традиционные эвристические методы нелинейного анализа данных	5			7
4	Внутренняя размерность массива данных и методы ее оценивания	2			8
5	Элементы дифференциальной геометрии и топологии	1			6
6	Математические модели нелинейных данных	2			7
7	Моделирование многообразий данных	3			5

8	Статистические задачи на многообразии данных	4			6
9	Элементы топологического анализа данных	3			8
Итого часов		30			60
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. Введение

Геометрические методы в машинном обучении: мотивация, примеры, основные задачи, под-ходы.

##### 2. Линейные методы анализа данных

Линейные модели данных. Проекционные методы в машинном обучении: анализ главных компонент, анализ независимых компонент, целенаправленное проектирование, случайные проекции.

##### 3. Традиционные эвристические методы нелинейного анализа данных

Эвристические методы нелинейного анализа данных: многомерное масштабирование, репли-кативные нейронные сети, ядерный анализ главных компонент.

##### 4. Внутренняя размерность массива данных и методы ее оценивания

Внутренняя размерность наборов нелинейных данных, оценка внутренней размерности набо-ра данных.

##### 5. Элементы дифференциальной геометрии и топологии

Топологические пространства. Многообразия. Векторные поля на многообразии. Римановы многообразия.

##### 6. Математические модели нелинейных данных

Модель нелинейных данных в виде многообразия. Задачи моделирования многообразий.

##### 7. Моделирование многообразий данных

Основные методы и алгоритмы моделирования многообразий (локально линейное вложение, изометрическое вложение, спектральное разложение лапласиана графа, локальное выравнива-ние касательного пространства, анализ геодезических, римановы нормальные координаты, спектральное разложение Грассманна-Штифеля, стохастическое вложение соседей (SNE, t-SNE, UMAP).

##### 8. Статистические задачи на многообразии данных

Регрессия. Оценка плотности.

##### 9. Элементы топологического анализа данных

Топологический анализ данных. Симплициальные комплексы. Филترация. Персистентные диаграммы.

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

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

## **6. Перечень рекомендуемой литературы**

### Основная литература

1. Введение в методы машинного обучения с подкреплением, учебное пособие /А. И. Панов; Министерство науки и высшего образования Российской Федерации ; Московский физико-технический институт (национальный исследовательский университет). Москва, МФТИ, 2019
2. Математические основы машинного обучения и прогнозирования / В. В. Вьюгин, Москва, МЦНМО, 2014

### Дополнительная литература

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

В процессе обучения ресурсы информационно-телекоммуникационной сети "Интернет" непосредственно использоваться не будут, но обучающие получают информацию об основных ресурсах информационно-телекоммуникационной сети "Интернет", которыми можно пользоваться при изучении дисциплины и в дальнейшем использовании полученных знаний и навыков в практической работе

## **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 занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

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

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

Успешное освоение курса требует напряженной самостоятельной работы студента. В программе курса отведено минимально необходимое время для работы студента над темой. Самостоятельная работа включает в себя:

- чтение и конспектирование рекомендованной литературы;
- проработку учебного материала (по конспектам занятий, учебной и научной литературе), подготовку ответов на вопросы, предназначенные для самостоятельного изучения, доказательство отдельных утверждений, свойств, решение задач;
- подготовка к дифференцированному зачёту.

Руководство и контроль за самостоятельной работой студента осуществляется в форме индивидуальных консультаций.

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



**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)**

<b>по направлению:</b>	Информатика и вычислительная техника
<b>профиль подготовки:</b>	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра проблем передачи информации и анализа данных
<b>курс:</b>	<u>1</u>
<b>квалификация:</b>	магистр
Семестр, формы промежуточной аттестации: 1 (осенний) - Дифференцированный зачет	
<b>Разработчик:</b>	А.В. Бернштейн, д-р физ.-мат. наук, профессор

## 1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Имеет представление об актуальных проблемах науки и техники в области информатики и вычислительной техники, способен на научном языке формулировать профессиональные задачи	ОПК-2.1 Имеет представление о современном состоянии исследований в рамках тематической области своей профессиональной деятельности
	ОПК-2.2 Способен оценивать актуальность исследований в области информатики и вычислительной техники и их практическую значимость
ПК-2 Понимает и способен применить в научно-исследовательской и прикладной деятельности основные законы естествознания, современный математический аппарат и алгоритмы, современные информационно-коммуникационные технологии	ПК-2.1 Знает основы научно-исследовательской деятельности в области информационных технологий, владеет знанием основ философии и методологии науки; знанием методов научных исследований и навыками их проведения
	ПК-2.3 Имеет практический опыт научно-исследовательской деятельности в области информационно-коммуникационных технологий
ПК-1 Готов к включению в профессиональное сообщество; способен проводить под научным руководством локальные исследования на основе существующих методов в конкретной области профессиональной деятельности	ПК-1.1 Знает принципы построения научной работы, методы сбора и анализа полученного материала, способы аргументации; владеет навыками подготовки научных обзоров, публикаций, рефератов и библиографий по тематике проводимых исследований на русском и английском языке
	ПК-1.2 Умеет решать научные задачи с пониманием существующих подходов к верификации моделей программного обеспечения в связи с поставленной целью и в соответствии с выбранной методикой

## 2. Показатели оценивания компетенций

В результате изучения дисциплины «Геометрические методы машинного обучения» обучающийся должен:

### знать:

- суть основных моделей, методов и алгоритмов геометрического анализа данных;
- основной доступный программный инструментарий, реализующий стандартные алгоритмы геометрического анализа данных и используемый для работы с массивами данных.

### уметь:

- выявлять геометрическую природу различных задач машинного обучения и подходы к их решению;
- выбирать подходящие методы и алгоритмы при использовании геометрических подходов в решении задач машинного обучения;
- работать с научной литературой по использованию подходов и методов геометрического анализа данных для решения задач машинного обучения;
- предлагать и внедрять изученные методы для решения задач машинного обучения.

### владеть:

- теоретическими основами и алгоритмическими приемами современного геометрического анализа данных;
- навыками работы с теоретическими и прикладными статьями, посвященных или использующих геометрическими методами анализа данных.

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

С целью контроля освоения обучающимися учебного материала проводится 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

1. Феномен пустого пространства: подробности и примеры.
2. Проклятие размерности: точные математические утверждения.
3. Анализ главных компонент как наилучшее линейное приближение ансамбля точек.
4. Анализ главных компонент на основе сингулярного разложения матриц.
5. Анализ главных компонент как решение задачи снижения размерности.
6. Анализ главных компонент как решение задачи метрического многомерного шкалирования.
7. Анализ главных компонент: как сохранение максимальной дисперсии.
8. Анализ независимых компонент: предположения и общий подход.
9. Анализ независимых компонент на основе характеристик островершинности распределения.
10. Анализ независимых компонент на основе энтропийного подхода.
11. Анализ независимых компонент на основе максимизации правдоподобия и минимизации взаимной информации.
12. Целенаправленное проектирование: подход и примеры.
13. Внутренняя размерность: математические определения.
14. Оценка внутренней размерности: геометрические методы.
15. Оценка внутренней размерности: регрессионный подход.
16. Оценка внутренней размерности: подход максимального правдоподобия.
17. Ядерный анализ главных компонент.
18. Автоассоциативные нейронные сети для снижения размерности.
19. Моделирование многообразий: локально-линейное вложение.
20. Моделирование многообразий: изометрическое отображение.
21. Моделирование многообразий: спектральное разложение Лапласиана.
22. Моделирование многообразий на основе анализа геодезических.
23. Моделирование многообразий на основе римановых нормальных координат.

##### Примеры билетов

###### Билет 1

1. Моделирование многообразий: обучение римановых многообразий.
2. Моделирование многообразий на основе стохастического анализа соседей.
3. Обучение многообразий: собственные карты Грассмана и Штифеля.

###### Билет 2

1. Обучение многообразий: регрессионные задачи.
2. Обучение многообразий: оценка плотности распределения.
3. Топологический анализ данных: основные элементы.

##### Критерии оценивания

Оценка "Отлично" (10) - 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) - 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) - Выставляется студенту, показавшему 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) - 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) - 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) - 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) - 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) - 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) - 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) - 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.

При проведении устного дифференцированного зачёта обучающемуся предоставляется 30 минут на подготовку. Во время проведения дифференцированного зачета обучающиеся могут пользоваться программой дисциплины, а также справочной литературой, вычислительной техникой.