

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

УТВЕРЖДЕНО

директор института -
заместитель директора ФАКТ
М.А. Кудров

	Рабочая программа дисциплины (модуля)
по дисциплине:	Анализ данных, суррогатное моделирование и оптимизация в прикладных задачах
по направлению:	Информатика и вычислительная техника
профиль подготовки:	Программная инженерия передовая инженерная школа радиолокации, радионавигации и программной инженерии кафедра компьютерного моделирования
курс:	4
квалификация:	бакалавр

Семестры, формы промежуточной аттестации:

7 (осенний) - Дифференцированный зачет
8 (весенний) - Экзамен

Аудиторных часов: 60 всего, в том числе:

лекции: 0 час.

семинары: 60 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 45 час.

Подготовка к экзамену: 30 час.

Всего часов: 135, всего зач. ед.: 3

Программу составил: А.А. Савельев, преподаватель

Программа обсуждена на заседании кафедры компьютерного моделирования 05.04.2022

Аннотация

Прогресс в сфере компьютерной техники и численных методов изменил задачи, с которыми сталкиваются современные учёные и инженеры, занимающиеся проектированием сложных технических систем.

Например, расчёт аэродинамики самолёта с требуемой для практики точностью в настоящее время может быть проведен менее чем за один час.

В результате, всё больше усилий исследователей направлено не на получение данных о характеристиках проектируемой системы, а на их анализ, интерпретацию и использование.

Предлагаемый курс посвящён методам интеллектуального анализа данных и машинного обучения, применяемым при проектировании технических систем, а также лежащим в основе этих методов разделам математики: линейной алгебре и математической статистики.

В программе содержится краткое изложение классических курсов по этим разделам, а также дополнительные главы, такие как сингулярное разложение матриц или случайные процессы.

Среди методов машинного обучения рассматриваются метод главных компонент, метод наименьших квадратов, линейная регрессия, гребневая регрессия и др.

Отдельный большой блок посвящён гауссовским процессам, а также основанным на них методам суррогатного моделирования и оптимизации.

Все рассматриваемые методы реализованы на языке Питон, а их применение проиллюстрировано на примере решения задач, взятых из реальной практики аэродинамического проектирования.

Курс ориентирован на студентов 4-го и 5-го курсов, обучающихся по специальностям "Прикладные математика и физика" и "Информатика и вычислительная техника".

1. Цели и задачи

Цель дисциплины

- изучение методов и технологий, применяющиеся в интеллектуальном анализе данных (ИАД, data mining) и базирующиеся на понятиях сходства, близости, аналогии.

Задачи дисциплины

- приобретение теоретических знаний в классификации, восстановлению регрессии, кластеризации, восстановлению пропущенных данных;
- освоение подходов для всех фундаментальных задач ИАД.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
ОПК-3 Способен составлять и оформлять научные и (или) технические (технологические, инновационные) отчеты (публикации, проекты)	ОПК-3.1 Знает основные правила оформления научных публикаций и научно-технической документации, в том числе с использованием прикладного программного обеспечения
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.2 Знает основные источники научно-технической и (или) технологической информации в области профессиональной деятельности

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

- теоретическая основа для построения, реализации и анализа широкого спектра моделей и методов ИАД, методы построения и вычисления функций сходства, согласование сходства на различных множествах объектов, синтез новых способов сравнения объектов на базе уже имеющихся;
- основной комплекс технологий, предназначенный для эффективного представления и обработки метрической информации вычислительными системами;
- основные области применения этих технологий.

уметь:

- применять методы и технологии к решению задач анализа данных.

владеть:

- навыками самостоятельной работы при решении типовых задач;
- культурой постановки и моделирования практически значимых задач;
- практикой исследования и решения теоретических и прикладных задач;
- навыками теоретического анализа реальных задач, решаемых с помощью методов 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	Основные подходы к заданию сходства. Классическое определение метрики и метрического пространства.		10		5
2	Локальные метрики и их продолжение на всё пространство. Геометрические подмножества общих метрических пространств. Примеры метрических пространств.		10		5
3	Классификация функций сходства. Характеристики метрик.		10		5
4	Преобразования метрик. Реализация метрик.		10		10
5	Принцип самоорганизации.		10		10
6	Метрики на конечных множествах. Разложение МК по конечным системам МК.		10		10
Итого часов			60		45
Подготовка к экзамену		30 час.			
Общая трудоёмкость		135 час., 3 зач.ед.			

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

Семестр: 7 (Осенний)

1. Основные подходы к заданию сходства. Классическое определение метрики и метрического пространства.

Основные подходы к заданию сходства. Функциональный подход: двуместные функции, удовлетворяющие аксиомам. Геометрический подход: определение в пространстве множеств точек. Табличный подход: матрицы попарного сходства над конечными множествами.

Классическое определение метрики и метрического пространства. Аксиоматическое задание метрики. Построение топологии по метрике. Пространства сходящихся последовательностей. Фундаментальные последовательности и полные пространства. Роль аксиомы треугольника и непрерывность метрики. Роль аксиомы сепарабельности и единственность предела сходящейся последовательности. Сопоставление метрик и отношений эквивалентности, 0,1-метрики. Различные модификации системы аксиом метрики и их интерпретация: расстояние, полуметрика, ультра-метрика, квази-метрика, неравенство Птолемея.

2. Локальные метрики и их продолжение на всё пространство. Геометрические подмножества общих метрических пространств. Примеры метрических пространств.

Локальные метрики и их продолжение на всё пространство. Формализация понятия «между» в метрическом пространстве. Выпуклость метрического пространства по Менгеру. Аксиомы существования и единственности точек между заданными точками. Аксиомы существования и единственности продолжения луча. Теорема о единственности продолжения локально совпадающих метрик. Практический пример проверки аксиом и использования локального продолжения метрики.

Геометрические подмножества общих метрических пространств. Понятия открытого и замкнутого шара, их согласованность с топологией метрического пространства. Понятия открытого и замкнутого обобщенного эллипсоида. Клетки Дирихле («сферы влияния»), автоматическое исправление ошибок. Геометрическое место точек, равноудаленных от заданных точек, проблема меры указанного подмножества. Понятие кривой в метрическом пространстве, длина кривой. Геодезическая линия, кривая наименьшей длины, сегмент. Свойство совпадения геодезических с множествами равноудаленных точек в обобщенных евклидовых пространствах.

Примеры метрических пространств. Пространство изолированных точек, дискретная топология. Метрики l_1 (городских кварталов), l_2 (евклидова), l_∞ (Чебышёва). Их физический смысл. Метрика l_p (Минковского). Форма шаров, вложенность единичных шаров. Зависимость объема шара от размерности пространства. Проблема сопоставления объема шаров в разных метриках с ростом размерности. Проблема единственности кратчайшего пути. Хаусдорфова метрика и другие метрики между подмножествами метрического пространства, индуцированные исходной метрикой между точками. Расстояния между функциями (графиками). Метрики на декартовом произведении метрических пространств. Случай конечного и бесконечного числа сомножителей, метрики на последовательностях.

3. Классификация функций сходства. Характеристики метрик.

Классификация функций сходства. Сопоставление значений: номинальные, порядковые, арифметические (интервальные, относительные, разностные, абсолютные) шкалы. Понятие о граничных объектах. Аксиомы сходства, главный и вспомогательный аргументы. Классификация мер сходства по одному свойству (признаку). Функции сходства на декартовом произведении пространств со значениями в различных шкалах.

Характеристики метрик. Инвариантность расстояния относительно сдвига, поворота. Инвариантность формы шаров относительно положения центра и направления на центр. Инвариантность объема шаров относительно положения центра и направления на центр. Ограниченность метрики. Ограниченность шаров. Понятие полностью абсолютных и полностью относительных метрик, промежуточные метрики. Выпуклость шаров. Односвязность шаров. Существование и единственность сегментов, непрерывность сегментов.

Семестр: 8 (Весенний)

4. Преобразования метрик. Реализация метрик.

Преобразования метрик. Изометрические преобразования пространств. Преобразования функций, сохраняющие метрические свойства. Некоторые достаточные условия преобразований, сохраняющих метрические свойства. Ограничение значений метрики (range comparands). Примеры универсальных компандеров. Возможность монотонного преобразования произвольной функции в метрику. Возможность линейного преобразования произвольной ограниченной функции в метрику. Нормализация метрик, зависимость от точки отсчета. Переход от булеанов конечных множеств к пространствам бинарных векторов, соответствие мощности множества и длины вектора.

Реализация метрик. Реализация конечных метрик точками ЛВП, точечные конфигурации. Алгоритмическая сложность решения задачи точного вложения в линейные пространства с метриками. Примеры МК, имеющих или не имеющих точную реализацию. Задача поиска оптимальной точечной конфигурации в пространстве малой размерности, методы метрического и неметрического многомерного шкалирования. Реализация многомерных данных элементами функциональных пространств. Методы визуализации многомерных данных: параллельные координатные оси, графики Эндрюса, шкалирование и иерархии, таблицы проекций, параметризованные глифы (звезды, лица Чернова).

5. Принцип самоорганизации.

Принцип самоорганизации. Принцип самоорганизации при построении эвристических информационных моделей. Понятие представителей, мера сходства между объектами и представителями. Функции представительства и назначений, структура метода. Самоорганизация в задаче кластеризации. Самоорганизация и задача факторного анализа, самоорганизация и задача дискриминантного анализа. Модификация прецедентной информации, понятие типологического дискриминантного анализа. Самоорганизация и задача восстановления пропусков.

6. Метрики на конечных множествах. Разложение МК по конечным системам МК.

Метрики на конечных множествах. Представление метрик таблицами попарных расстояний. Метрическая конфигурация (МК). Специальное линейное пространство метрических конфигураций. Система неравенств треугольника как определение полиэдрального конуса полуметрик. Грани и экстремальные лучи полуметрического конуса, проблема их определения. Векторное представление метрических конфигураций. Достаточные условия сохранения метрических свойств покомпонентными корректорами метрических конфигураций. Примеры использования достаточных условий. Несовместимость метрических свойств и ортогональности метрических конфигураций.

Разложение МК по конечным системам МК. Полные системы, базисы МК. Проблема использования переполненных систем МК. Гомогенные базисы, интерпретация коэффициентов разложения. Ранг МК. Ранговые и полуметрические ранговые базисы. Неполные системы, оптимальная аппроксимация МК. Разложение по системе «отдельных объектов», метрика попарных сумм, эффективное вычисление признака «общая удаленность» для индивидуальных объектов.

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

Необходимое оборудование: аудитория, компьютер и мультимедийное оборудование (проектор, звуковая система).

6.Перечень рекомендуемой литературы

Основная литература

1. Курс химической кинетики (гомогенные реакции) [Текст] / Н. М. Эмануэль, Д. Г. Кнорре - М. Высшая школа, 1962

Дополнительная литература

1. Методы анализа данных [Текст] = Optimisation en classification automatique, подход, основанный на методе динамических сгущений/пер. с фр. С. А. Айвазяна, В. Д. Конакова, С. Ю. Адамова, -М., Финансы и статистика, 1985

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

<http://www.machinelearning.ru/wiki/index.php>

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

На занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

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, понятия, аксиомы, алгоритмы.

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

- чтение и конспектирование рекомендованной литературы;
- проработку учебного материала (по учебной и научной литературе), подготовку ответов на вопросы, предназначенных для самостоятельного изучения, доказательство отдельных утверждений, свойств;
- подготовку к дифференцированному зачету и экзамену.

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.

ПРИЛОЖЕНИЕ

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Информатика и вычислительная техника
профиль подготовки:	Программная инженерия передовая инженерная школа радиолокации, радионавигации и программной инженерии кафедра компьютерного моделирования
курс:	4
квалификация:	бакалавр
Семестры, формы промежуточной аттестации:	
	7 (осенний) - Дифференцированный зачет
	8 (весенний) - Экзамен
Разработчик:	А.А. Савельев, преподаватель

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
ОПК-2 Способен использовать современные информационные технологии и программные средства при решении задач профессиональной деятельности, соблюдая требования информационной безопасности	ОПК-2.1 Способен применять современные вычислительную технику и сервисы сети Интернет в области (сфере) профессиональной деятельности
ОПК-3 Способен составлять и оформлять научные и (или) технические (технологические, инновационные) отчеты (публикации, проекты)	ОПК-3.1 Знает основные правила оформления научных публикаций и научно-технической документации, в том числе с использованием прикладного программного обеспечения
ОПК-4 Способен осуществлять сбор и обработку научно-технической и (или) технологической информации для решения фундаментальных и прикладных задач	ОПК-4.2 Знает основные источники научно-технической и (или) технологической информации в области профессиональной деятельности

2. Показатели оценивания компетенций

В результате изучения дисциплины «Анализ данных, суррогатное моделирование и оптимизация в прикладных задачах» обучающийся должен:

знать:

- теоретическая основа для построения, реализации и анализа широкого спектра моделей и методов ИАД, методы построения и вычисления функций сходства, согласование сходства на различных множествах объектов, синтез новых способов сравнения объектов на базе уже имеющихся;
- основной комплекс технологий, предназначенный для эффективного представления и обработки метрической информации вычислительными системами;
- основные области применения этих технологий.

уметь:

- применять методы и технологии к решению задач анализа данных.

владеть:

- навыками самостоятельной работы при решении типовых задач;
- культурой постановки и моделирования практически значимых задач;
- практикой исследования и решения теоретических и прикладных задач;
- навыками теоретического анализа реальных задач, решаемых с помощью методов анализа данных.

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

1. Описать подходы по сходству в оптимизации.
2. Определение метрики и метрического пространства, классификация метрик.
3. Локальные метрики и их продолжение на всё пространство.
4. Геометрические подмножества общих метрических пространств.
5. Привести примеры метрических пространств.
6. Что свойственно суррогатному моделированию? Какие его основные свойства?
7. Реализация метрик в программном описании процессов.
8. Классификация функций сходства.
9. Понятие о граничных объектах.
10. Аксиомы сходства, главный и вспомогательный аргументы.
11. Классификация мер сходства по одному свойству (признаку).

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

Вопросы к дифференцированному зачету

1. Классификация функций сходимости.
 2. Сопоставление значений: номинальные, порядковые, арифметические (интервальные, относительные, разностные, абсолютные) шкалы.
 3. Понятие о граничных объектах.
 4. Аксиомы сходимости, главный и вспомогательный аргументы.
 5. Классификация мер сходимости по одному свойству (признаку).
 6. Функции сходимости на декартовом произведении пространств со значениями в различных шкалах.
- Характеристики метрик.
7. Инвариантность расстояния относительно сдвига, поворота.
 8. Инвариантность формы шаров относительно положения центра и направления на центр. Инвариантность объема шаров относительно положения центра и направления на центр.
 9. Ограниченность метрики. Ограниченность шаров.
 10. Понятие полностью абсолютных и полностью относительных метрик, промежуточные метрики. Выпуклость шаров. Односвязность шаров.
 11. Существование и единственность сегментов, непрерывность сегментов.

Вопросы к экзамену:

1. Основные подходы к заданию сходимости.
2. Классическое определение метрики и метрического пространства.
3. Локальные метрики и их продолжение на всё пространство.
4. Геометрические подмножества общих метрических пространств.
5. Примеры метрических пространств.
6. Классификация функций сходимости.
7. Характеристики метрик.
8. Преобразования метрик.
9. Функции сходимости на декартовом произведении пространств со значениями в различных шкалах.
10. Ограниченность метрики. Ограниченность шаров.
10. Понятие полностью абсолютных и полностью относительных метрик, промежуточные метрики.

Билет 1

1. Реализация метрик.
2. Принцип самоорганизации.

Билет 2

1. Метрики на конечных множествах.
2. Разложение МК по конечным системам МК.

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 дифференцированного зачета и экзамена обучающиеся могут пользоваться программой дисциплины, а также справочной литературой, вычислительной техникой, конспектами лекций.

Дифференцированный зачет и экзамен может проводиться по итогам текущей успеваемости и сдачи заданий, или путем организации специального опроса, проводимого в устной форме.