

**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: | Основы статистического моделирования и исследование зависимостей |
| по направлению: | Инфокоммуникационные технологии и системы связи |
| профиль подготовки: | Телекоммуникационные сети и системы |
| | Физтех-школа Радиотехники и Компьютерных Технологий |
| | кафедра проблем передачи информации и анализа данных |
| курс: | 1 |
| квалификация: | магистр |

Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен

Аудиторных часов: 30 всего, в том числе:

лекции: 0 час.

семинары: 30 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Подготовка к экзамену: 30 час.

Всего часов: 90, всего зач. ед.: 2

Программу составили:

М.Е. Панов, канд. физ.-мат. наук

Ю.А. Янович

Программа обсуждена на заседании кафедры проблем передачи информации и анализа данных 14.03.2024

Аннотация

Математическая статистика — это быстро развивающаяся и обширная область прикладной математики. Классические методы математической статистики важны для её глубокого понимания, но на практике зачастую применяются современные подходы, например, бутстреп и методы планирования экспериментов. Понимание теории, лежащей в основе этих методов, а также умение применить их на практике, абсолютно необходимо каждому, кто занимается математической статистикой и анализом данных.

В курсе рассматриваются задачи моделирования и исследования зависимостей классическими и современными статистическими методами, включающими бутстреп, непараметрическое оценивание, методы планирования экспериментов, анализа чувствительности, фильтрацию выбросов, методы снижения размерности и т.д.

1. Цели и задачи

Цель дисциплины

Дать представление об основных современных методах прикладной математической статистики и способах ее применения для решения прикладных задач анализа и обработки данных.

Задачи дисциплины

- изучение основных методов прикладной математической статистики;
- практическое применение основных методов прикладной математической статистики.

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 |
| УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия | УК-4.4 Способен использовать современные средства информационно-коммуникационных технологий для академического и профессионального взаимодействия |
| 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 представлять современную научную картину мира, выявлять естественнонаучную сущность проблем своей профессиональной деятельности, определять пути их решения и оценивать эффективность сделанного выбора | ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные и прикладные научные знания в области естественных наук |
| | ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности |
| ОПК-2 Способен реализовывать новые принципы и методы исследования современных инфокоммуникационных систем и сетей различных типов передачи, распределения, обработки и хранения информации | ОПК-2.2 Владеет навыками реализации новых принципов и методов исследования в современных инфокоммуникационных системах и сетях |

| | |
|--|---|
| ОПК-3 Способен приобретать, обрабатывать и использовать новую информацию в своей предметной области, предлагать новые идеи и подходы к решению задач своей профессиональной деятельности | ОПК-3.1 Умеет использовать современные информационные и компьютерные технологии, средства коммуникаций при поиске научно-технической информации в своей профессиональной деятельности |
|--|---|

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные методы прикладной математической статистики и возможности их практического использования;
- численные алгоритмы, реализующие изученные методы прикладной математической статистики.

уметь:

- применять на практике методы прикладной математической статистики и/или разрабатывать их модификации для решения поставленных задач;
- писать программы, реализующие алгоритмические процедуры анализа и обработки данных, на языке системы MatLab.

владеть:

- навыком освоения большого объема информации;
- навыками постановки научно-исследовательских задач.

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 | Методологические основы прикладной математической статистики | | 2 | | 2 |
| 2 | Основы теории статистических выводов | | 15 | | 10 |
| 3 | Статистические модели и методы | | 8 | | 6 |
| 4 | Примеры применения методов прикладной математической статистики | | 5 | | 12 |
| Итого часов | | | 30 | | 30 |
| Подготовка к экзамену | | 30 час. | | | |
| Общая трудоёмкость | | 90 час., 2 зач.ед. | | | |

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

Семестр: 1 (Осенний)

1. Методологические основы прикладной математической статистики

Прикладная математическая статистика как самостоятельная научная дисциплина. Связь прикладной математической статистики с теорией вероятностей, теоретической математической статистикой и анализом данных.

Теоретико-вероятностный способ рассуждения в прикладной математической статистике. Математические модели в прикладной математической статистике. Робастность статистических процедур.

2. Основы теории статистических выводов

Основные задачи и методы теории статистических выводов. Параметрические и непараметрические модели. Основные задачи: точечное оценивание, доверительные множества, тестирование гипотез, исследование зависимостей.

Оценка распределения и статистические функционалы. Эмпирическая функция распределения. Статистические функционалы.

Бутстреп. Моделирование Монте-Карло, бутстреп. Оценка дисперсии на основе бутстрепа. Оценка доверительных интервалов на основе бутстрепа. Метод складного ножа.

Параметрическое оценивание. Метод моментов. Метод максимального правдоподобия и его свойства. Дельта-метод. Параметрический бутстреп.

Проверка гипотез. Основные понятия теории проверки гипотез. Критерий Вальда. Р-значение. Критерий перестановок. Множественные тесты.

Байесовский подход к оцениванию. Философия байесовского подхода. Байесовское оценивание и свойства получаемых оценок. Типы априорных распределений. Проверка гипотез. Достоинства и недостатки байесовского подхода.

Статистическая теория решений. Функция риска. Байесовская оценочная функция. Минимаксный подход. Принятие решений на основе отношения правдоподобия, минимаксного и байесовского подходов.

3. Статистические модели и методы

Многомерные данные. Случайные вектора. Многомерное нормальное распределение. Оценка корреляций.

Линейная и логистическая регрессии. Стандартная линейная регрессия. Метод оценивания на основе минимизации невязок/максимизации правдоподобия. Свойства оценок метода наименьших квадратов. Выбор модели. AIC, BIC, Lasso, Bridge-регрессия, Elastic Net.

Непараметрическое оценивание сигналов. Выбор оптимального соотношения между смещением и дисперсией. Гистограммы. Ядерная оценка плотности. Непараметрическая регрессия.

Нелинейные методы построения регрессионных зависимостей. Аддитивные модели. Разложение по адаптивным сигмоидоподобным функциям. Разложение по адаптивным гауссоподобным функциям: кригинг, радиальные базисные функции и т.п.

Снижение размерности многомерных данных. Внутренняя размерность множества (фрактальная размерность, корреляционная размерность). Постановка задачи снижения размерности. Обзор линейных методов снижения размерности (метод главных компонент, целенаправленное проектирование и т.п.). Обзор локальных и нелинейных методов снижения размерности (метод нелинейных главных компонент, метод локального линейного вложения и т.п.). Снижение размерности данных в соболевской метрике. Выбор наиболее значимых признаков в задаче построения регрессии как задача снижения размерности.

4. Примеры применения методов прикладной математической статистики

Методы построения и способы использования моделей на основе данных. Методы построения матрицы плана. Методология построения моделей на основе данных. Использование моделей на основе данных для оптимизации сложных технических объектов.

Примеры применения моделей на основе данных для решения реальных промышленных задач.

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

Учебная аудитория, оснащенная мультимедийным оборудованием (проектор или плазменная панель), доской; компьютерный класс.

6. Перечень рекомендуемой литературы

Основная литература

1. Исследование операций [Текст] : Задачи, принципы, методология : учеб. пособие для вузов / Е. С. Вентцель .— 3-е изд., стереотип. — М. : Дрофа, 2004 .— 208 с.

Фонд литературы базовой кафедры (организации):

2. Hastie T., Tibshirani R., Friedman J. The elements of statistical learning: data mining, inference, and prediction. Springer, 2001. - 533 p.
3. Wasserman L. All of statistics. A concise course in statistical inference. Springer, 2004. - 442 p.

Дополнительная литература

1. Статистическое моделирование [Текст] : уч. пособие для вузов : доп. М-вом высш. и сред. спец. образов. СССР / С. М. Ермаков, Г. А. Михайлов .— 2-е изд., доп. — М. : Наука, 1982 .— 296 с.

Фонд литературы базовой кафедры (организации):

2. Bishop C.M. Pattern recognition and machine learning. Springer, 2006. - 101 p.
3. David Mackay J.C. Information theory, inference, and learning algorithms. Cambridge, 2007. - 640 p.
4. Grimmett G., Stirzaker D. Probability and Random Processes. Oxford University Press, 2001. - 448 p.

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

1. <http://lib.mipt.ru> – электронная библиотека Физтеха.
2. <http://www.edu.ru> – федеральный портал «Российское образование».
3. <http://benran.ru> – библиотека по естественным наукам Российской академии наук.

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 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 прикладной математической статистики и возможности их практического использования; численные алгоритмы, реализующие изученные методы прикладной математической статистики.

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, решение задач;
- подготовка к 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.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

| | |
|----------------------------|--|
| по направлению: | Инфокоммуникационные технологии и системы связи |
| профиль подготовки: | Телекоммуникационные сети и системы Физтех-школа Радиотехники и Компьютерных Технологий кафедра проблем передачи информации и анализа данных |
| курс: | <u>1</u> |
| квалификация: | магистр |

Семестр, формы промежуточной аттестации: 1 (осенний) - Экзамен

Разработчики:

М.Е. Панов, канд. физ.-мат. наук

Ю.А. Янович

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 |
| УК-4 Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия | УК-4.4 Способен использовать современные средства информационно-коммуникационных технологий для академического и профессионального взаимодействия |
| 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 представлять современную научную картину мира, выявлять естественнонаучную сущность проблем своей профессиональной деятельности, определять пути их решения и оценивать эффективность сделанного выбора | ОПК-1.1 Знает и способен использовать в профессиональной деятельности фундаментальные и прикладные научные знания в области естественных наук |
| | ОПК-1.2 Способен обобщать и критически оценивать опыт и результаты научных исследований в области профессиональной деятельности |
| ОПК-2 Способен реализовывать новые принципы и методы исследования современных инфокоммуникационных систем и сетей различных типов передачи, распределения, обработки и хранения информации | ОПК-2.2 Владеет навыками реализации новых принципов и методов исследования в современных инфокоммуникационных системах и сетях |
| ОПК-3 Способен приобретать, обрабатывать и использовать новую информацию в своей предметной области, предлагать новые идеи и подходы к решению задач своей профессиональной деятельности | ОПК-3.1 Умеет использовать современные информационные и компьютерные технологии, средства коммуникаций при поиске научно-технической информации в своей профессиональной деятельности |

2. Показатели оценивания компетенций

В результате изучения дисциплины «Основы статистического моделирования и исследование зависимостей» обучающийся должен:

знать:

- основные методы прикладной математической статистики и возможности их практического использования;
- численные алгоритмы, реализующие изученные методы прикладной математической статистики.

уметь:

- применять на практике методы прикладной математической статистики и/или разрабатывать их модификации для решения поставленных задач;
- писать программы, реализующие алгоритмические процедуры анализа и обработки данных, на языке системы MatLab.

владеть:

- навыком освоения большого объема информации;
- навыками постановки научно-исследовательских задач.

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 семинара или в ходе занятия по изучаемой теме.

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

Перечень контрольных вопросов для сдачи экзамена:

1. Эмпирическая функция распределения и ее свойства.
2. Статистические функционалы и их свойства.
3. Оценка дисперсии на основе бутстрепа.
4. Оценка доверительных интервалов на основе бутстрепа.
5. Метод складного ножа.
6. Метод моментов и его свойства.
7. Метод максимума правдоподобия и его свойства.
8. Дельта метод и его свойства.
9. Основные понятия теории проверки гипотез: ошибки первого и второго рода, функция мощности, несмещенность критерия, равномерно наиболее мощный критерий, Р-значение, и др.
10. Критерий Вальда. Критерий перестановок. Хи-квадрат.
11. Метод оценивания параметров линейной регрессии на основе минимизации невязок/максимизации правдоподобия.
12. Выбор модели. AIC, BIC, Lasso, Bridge-регрессия, Elastic Net.
13. Выбор оптимального соотношения между смещением и дисперсией.
14. Непараметрическая ядерная оценка гистограммы и ее свойства.
15. Непараметрическая ядерная оценка регрессии и ее свойства.
16. Регрессия на основе гауссовских процессов.
17. Приближение многомерных зависимостей на основе разложения по словарю параметрических функций.
18. Анализ главных компонент и его свойства.
19. Методы моделирования многообразий LLE, LTSA, ISOMAP.
20. Суррогатная оптимизация на основе гауссовских процессов.

Примеры билетов для проведения экзамена:

Билет 1.

1. Эмпирическая функция распределения и ее свойства.
2. Суррогатная оптимизация на основе гауссовских процессов.

Билет 2.

1. Статистические функционалы и их свойства.
2. Методы моделирования многообразий LLE, LTSA, ISOMAP.

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

Экзамен проводится в устной форме.

При проведении устного экзамена обучающемуся предоставляется 30 минут на подготовку. Опрос обучающегося по билету на устном экзамене не должен превышать двух астрономических часов.

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