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

УТВЕРЖДЕНО
Проректор по учебной работе и довузовской подготовке
А.А. Воронов

Рабочая программа дисциплины (модуля)

по дисциплине:	Основы статистического моделирования и исследование зависимостей
по направлению:	Прикладные математика и физика
профиль подготовки:	Компьютерные технологии и интеллектуальный анализ данных Физтех-школа прикладной математики и информатики Кафедра проблем передачи информации и анализа данных
курс:	4
квалификация:	Бакалавр

Семестр, формы промежуточной аттестации: 8 (Весенний) - Простой зачёт

Программу составили: **Е.В. Бурнаев, кандидат физико-математических наук, доцент**
М.Е. Панов, кандидат физико-математических наук
Ю.А. Янович, кандидат физико-математических наук

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.

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

Семестр: 8 (Весенний)

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 при-

кладной 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; P-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 индустриальных задач.

Перечень рекомендуемой литературы

Основная литература

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

3. Bishop C.M. Pattern recognition and machine learning. Springer, 2006. - 101 p.
4. David Mackay J.C. Information theory, inference, and learning algorithms. Cambridge, 2007. - 640 p.
5. Grimmett G., Stirzaker D. Probability and Random Processes. Oxford University Press, 2001. - 448 p.
6. Forrester A., Sobester A., Keane A. Engineering Design via Surrogate Modelling. A Practical Guide. Wiley, 2008. - 228 p.
7. Lee J.A., Verleysen M. Nonlinear Dimensionality Reduction. Springer, 2007. - 328 p.
8. Wang G.G., Shan S. Review of Metamodeling Techniques in Support of Engineering Design Optimization // Journal of Mechanical Design. 2007. Vol. 129. № 4. P. 370-380.
9. Deconinck, Periaux, Giannakoglou (eds.). Optimization method and tools for multicriteria/multidisciplinary design. Applications to aeronautics and turbomachinary // Von Karman Institute for Fluid Dynamics. Lecture Series 2004-07. 2004.

Дополнительная литература

1. Лагутин М.Б. Наглядная математическая статистика // Учебное пособие. М.: БИНОМ. Лаборатория знаний, 2007. - 472 с.
2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных // Справочное издание. М.: Финансы и статистика, 1983. - 471 с.
3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей // Справочное издание. М.: Финансы и статистика, 1985. - 487 с.
4. Айвазян С.А., Бухштабер В.М., Енюков С.А., Мешалкин Л.Д. Прикладная статистика. Классификация и снижение размерности // Справочное издание. М.: Финансы и статистика, 1989. - 607 с.