

**АННОТАЦИИ РАБОЧИХ ПРОГРАММ ДИСЦИПЛИН (МОДУЛЕЙ)
ООП ВЫСШЕГО ОБРАЗОВАНИЯ – ПРОГРАММЫ МАГИСТРАТУРЫ
Направление подготовки 01.04.02 «Прикладная математика и информатика»**

Направленность программы (магистерская программа)
«Интеллектуальный анализ больших данных»

Английский язык

Задачи дисциплины:

- совершенствовать навыки чтения и понимания научной литературы по профессиональной тематике на английском языке;
- помочь развитию логического мышления учащихся, умения выделить основную и второстепенную информацию, аргументировать и резюмировать прочитанное;
- научить студентов магистратуры принципам написания реферата, академического эссе и аннотаций профессионального текста на английском языке;
- обучить представлению результатов исследования в виде презентаций и дискуссий профессиональной направленности на английском языке;
- совершенствовать навыки понимания публичной речи;
- познакомить студентов магистратуры с современными требованиями цитирования, оформления ссылок на источники и библиографического списка в собственных научных работах и статьях на английском языке;
- повысить общеобразовательный, культурный и политический кругозор студентов.

Правоведение

Целью освоения дисциплины «Правоведение» является получение студентами магистратуры общетеоретических и специальных знаний по основным отраслям российского права в период построения правового государства, формирования информационного общества и цифровой экономики в России. Главными задачами являются, во-первых, повышение общего уровня правовой подготовки студентов на основе формирования у них правосознания, умения и навыков, обеспечивающих использование методических приемов защиты прав и законных интересов, повышения правовой активности во всех сферах жизнедеятельности, в том числе, профессиональной информационной сфере; во-вторых, изучение приемов и методов раскрытия наиболее важных вопросов права с учетом инновационного развития российского государства.

Русский язык и культура речи

Целями освоения дисциплины являются: формирование умения устанавливать связь между языковыми знаками русского языка и явлениями отражаемой этими знаками действительности; овладение сознательным умением извлекать полный и точный смысл из предъявленного речевого сообщения; формирование умения создавать речевые произведения разных стилей и жанров в соответствии с замыслом производителя речи, условиями общения и характером отношений с адресатом; совершенствование представления о русском языке как о культурной ценности, нуждающейся в сохранении и постоянном развитии в соответствии с динамикой жизни и потребностями российского общества.

Суперкомпьютерное моделирование и технологии

Суперкомпьютерное моделирование является определяющим фактором развития научно-технического прогресса. Решение прорывных задач современности невозможно без использования суперкомпьютеров. Курс посвящен изучению базовых основ суперкомпьютерного моделирования. В курсе рассматриваются вопросы современного состояния развития суперкомпьютерных технологий, включая суперкомпьютерные аппаратно-программные платформы, математические модели решаемых на суперкомпьютерных задач и алгоритмов их решения, параллельные технологии реализации таких задач на суперкомпьютерах. Неотъемлемой частью курса является выполнение студентами практических заданий на суперкомпьютерах МГУ и высокопроизводительных вычислительных системах ряда научных организаций. Особенностью курса является широкое участие в его проведении специалистов из различных научных областей, связанных с применением суперкомпьютерных технологий. Этот подход позволяет обеспечить квалифицированный междисциплинарный подход, являющийся основой суперкомпьютерного моделирования.

Современная философия и методология науки

Целью дисциплины является формирование у слушателя целостного видения науки, понимания ее специфики научной деятельности, характера исторического развития науки, ее взаимодействия с другими сферами человеческой деятельности. В курсе представлены основные темы философии. Рассматриваются основные положения учения о науке как познавательной деятельности, как социальном институте, как виде человеческой деятельности, как элементе культуры.

История и методология прикладной математики и информатики

В рамках курса рассматриваются основные факты, события и идеи многовековой истории развития математики в целом и одного из ее важнейших направлений – «прикладной» - вычислительной математики, зарождения и развития вычислительной техники и программирования. Показывается роль математики и информатики в истории развития цивилизации. Дается характеристика научного творчества наиболее выдающихся ученых – генераторов научных идей. Особое внимание уделяется развитию математики и информатики в России.

Курс нацелен на формирование математического мировоззрения будущих магистров, выстраивание общего контекста математического мышления как культурной формы деятельности, определяемой как структурными особенностями математического знания, так и местом математики в системе наук.

Модуль «Технологии анализа данных»

Анализ риска

Курс направлен на формирование у студентов понимания общих подходов к постановке и решению задач актуарной математики, относящихся к области теории риска. Подробно обсуждаются основные принципы формирования страховых портфелей и тарифов в зависимости от рискованных ситуаций страховщика. Исследуются вопросы, связанные с нахождением вероятности разорения и ее асимптотических аппроксимаций. Особое внимание уделяется факторизационной статической модели страхования, в основе которой лежит разложение индивидуальных исков страхователей на независимые компоненты. Подобный подход позволяет существенно расширить и обобщить классические постановки прикладных актуарных задач в целях получения адекватных решений.

Анализ временных рядов

В рамках курса рассматриваются основные классические и современные методы анализа временных рядов. Среди рассматриваемых методов: линейный и нелинейный регрессионный анализ, линейные стационарные модели временных рядов авторегрессии и скользящего среднего, модели рядов с гетероскедастичностью, модели многомерных нестационарных временных рядов, теория коинтеграции, Байесовские модели временных рядов, теория копул, применение нейронных сетей для анализа временных рядов. Весь теоретический материал курса сопровождается выполнением программ на языке R и интерактивными примерами с анимацией, позволяющим улучшить усвоение материала. Все материалы курса: тексты лекции, интерактивные фреймы с анимацией находятся на факультетском сервере и доступны при наличии доступа в сеть факультета.

Интеллектуальный анализ данных

В курсе рассматриваются современные алгоритмы и методы интеллектуального анализа данных для решения поиска ассоциативных правил, тематического моделирования, кластеризации, классификации и прогнозирования. В первой части курса, посвященной изучению методов обучения без учителя, рассматриваются: задача поиска ассоциативных правил и основные применяемые для этого алгоритмы - `apriori` и `fp-tree`; задача выявления скрытых структур в данных на основе тематического моделирования, в частности метод главных компонент, кластеризация переменных, самоорганизующиеся отображения, неотрицательная матричная факторизация; задача кластеризации данных на основе иерархических, метрических и вероятностных методов. Также обсуждаются методы предобработки данных для эффективного решения данных задач. Вторая часть курса посвящена изучению методов прогнозирования,

используемых в системах интеллектуального анализа данных, связанные с этим проблемы, алгоритмы и терминология. Рассматриваются следующие вопросы: понятие проклятия размерности и проблема переобучения; вопросы и критерии для оценки и выбора моделей с использованием валидации и кросс-валидации; алгоритмы и методы необходимой предобработки данных для решения задачи прогнозирования. Далее рассматриваются наиболее популярные и современные алгоритмы и модели машинного обучения и прикладной статистики для решения задач прогнозирования в системах интеллектуального анализа данных, в частности: линейные регрессионные модели; пошаговые методы отбора переменных, регуляризация, преобразование пространства признаков для решения задач прогнозирования; нелинейные регрессионные модели, сплайны, локальная взвешенная регрессия; нейронные сети, их типовые архитектуры RBF и MLP, алгоритмы ранней остановки обучения, методы оптимизации для обучения нейронных сетей; а метод опорных векторов для бинарной классификации, виды ядерных функций, алгоритмы оптимизации для обучения модели на основе опорных векторов; деревья решений, алгоритмы и критерии поиска разбиения при их построении, вопросы управление процессом роста и обрубания ветвей деревьев для борьбы с переобучением; ансамбли моделей на основе бустинга и бэггинга, случайный лес и градиентный бустинг. Демонстрация примеров использования изучаемых методов и процедур проводится преподавателями на каждой лекции и каждом семинаре. Также данная дисциплина поддерживается практическими заданиями (практическими самостоятельными работами), позволяющими студентам овладеть навыками построения прогнозных и описательных моделей интеллектуального анализа данных, а также навыками анализа результатов и оценки работы реализованных моделей. Обсуждение практических самостоятельных работ, а также их защита, проводятся на семинарах. Дополнительно, на семинарах студенты выполняют небольшие практические задания по тематике последней на момент данного семинара лекции. Темы семинаров соответствуют темам лекций. Семинары направлены на укрепление знаний, полученных на лекциях.

Модуль «Статистический анализ и прогнозирование»

Прикладной многомерный статистический анализ

В рамках данного курса будут рассмотрены основные задачи многомерного статистического анализа. А именно, будет дано описание математических моделей и методов таких разделов математической статистики как корреляционный анализ, регрессионный анализ, дисперсионный анализ, дискриминантный анализ, кластерный анализ. Предложенные методы и алгоритмы иллюстрируются с помощью более-менее реальных примеров.

Прикладные задачи теории случайных процессов

В рамках курса рассматриваются основные понятия и методы теории случайных процессов, которые используются в качестве математических моделей широкого круга явлений. Существенно используются многие разделы курсов «теория вероятностей и математическая статистика», «математический анализ», «алгебра и геометрия», «функциональный анализ». Особое внимание уделяется постановкам и методам решения задач для основных классов случайных процессов, в том числе: урновые модели; геометрическое, равномерное, показательное распределения, а также формированию умений и навыков использовать математический аппарат теории случайных процессов для решения прикладных задач.

Модуль «Методы обработки и анализа цифровой информации»

Современные методы обработки сигналов

В рамках курса рассматриваются различные задачи анализа и обработки сигналов и изображений. Основное внимание уделяется статистическим методам анализа временных выборок, методам спектрального Фурье-анализа, методам вейвлет-анализа. Излагается теория линейной фильтрации. Обсуждаются методы подавления шума и сжатия данных. Отдельное внимание уделяется методам решения обратных задач обработки и анализа изображений. Излагаются элементы теории преобразования Радона и методы реконструкции томографических изображений.

Обработка и распознавание изображений

Дисциплина посвящена изучению современных методов распознавания образов применительно к задачам компьютерного зрения. Общие подходы к распознаванию образов имеют специфические особенности, связанные с особой природой и форматами данных в цифровых изображениях. В

курсе рассматриваются модели и методы генерации признаков, построения и оценки классификаторов для распознавания изображений. Основной акцент делается на математическом содержании используемых моделей и методов

Модуль «Компьютерные технологии и программное обеспечение в задачах анализа данных»

Пакеты прикладных программ для статистической обработки и анализа данных

В курсе рассматриваются основные вопросы программирования для решения задач статистического анализа данных с использованием аналитической платформы SAS. В первой части курса, посвященной обучению программированию для решения задач подготовки данных и формирования отчетности, рассматриваются: основные принципы работы шага обработки данных; работа со структурированными наборами данных и массивами; форматы и типы данных языка SAS Base; процедуры преобразования форматов и типов; работа с внешними сложно структурированными наборами данных; алгоритмы и методы для организации поиска по ключу с помощью индексов, форматов, хэш-объектов; методы разработки и использования пользовательских процедур и функций; программирование с использованием макросов, макропеременных и макроподстановок; использованием языка SQL; формирование отчетов и работа подсистемой вывода; графические возможности и процедуры. Вторая часть курса посвящена изучению методов разработки программ для статистического анализа данных с использованием соответствующих библиотек аналитической платформы SAS. Рассматриваются следующие вопросы: процедуры и методы для проверки статистических гипотез; модели и процедуры для дисперсионного анализа данных; построение линейных регрессионных моделей; проблема мультиколлинеарности; методы пошагового отбора переменных, регуляризации, преобразования пространства признаков; процедуры поиска главных компонент и кластеризации переменных; процедуры и инструменты для поиска выбросов; процедуры построения нелинейных регрессий; анализ таблиц сопряженности; логистическая регрессия; обобщенные линейные модели, пуассоновская и гамма регрессии; методы сравнения и оценки моделей на тестовом наборе данных. Демонстрация примеров использования изучаемых методов и процедур проводится преподавателями на каждой лекции и каждом семинаре. Также данная дисциплина поддерживается практическими заданиями (практическими самостоятельными работами), позволяющими студентам овладеть навыками написания программ для статистической обработки данных, а также навыками анализа результатов работы реализованных алгоритмов.

Методы анализа и проектирования программного обеспечения

Курс с экзаменом и практическими заданиями по UML-моделированию. В курсе рассматриваются современные методы и средства анализа и проектирования программного обеспечения, основанные на применении объектно-ориентированного подхода и унифицированного языка моделирования, а также их практическое использование.

Современные методы распределенного хранения и обработки данных

В курсе рассматриваются основные вопросы аналитики больших данных. Рассматриваются проблемы работы современных перспективных СУБД с большими данными, особенности масштабируемой параллельной серверной бизнес-аналитики. Изучаются современные виды хранилищ данных, модель данных OLAP. Большое внимание уделено практическим сценариям использования больших данных. Изучаются наиболее часто используемые на практике языки программирования для обработки больших данных. Введение в распределенные файловые системы и распределенную обработку данных изучается на примере широко известных технологий Google File System и Google Map Reduce. Далее данная тема глубоко изучается с использованием технологий HDFS и Hadoop Map Reduce. Изучение Hadoop производится с использованием технологий Microsoft Azure, позволяющих более детально изучить влияние отдельных факторов на скорость распределенной обработки информации. Рассматриваются примеры распределенных алгоритмов, теория сложности Map Reduce, вопросы планирования и управления вычислениями. Также изучаются технологии Hadoop Streaming, Apache Spark. Большое внимание уделено системам хранения на базе HDFS: Pig, Hive, HBase, а также форматам хранения бинарных данных.

Анализ текстовых данных и информационный поиск