|  |
| --- |
| **М 01.04.02 М 2014 очная Науки о данных** |
| **Tips&Tricks в анализе данных** |
| Данный предмет поможет взглянуть на машинное обучение больше с инженерной точки зрения. На курсе разберутся некоторые конкретные задачи и технические трюки, которые используются для того, чтобы машинное обучение работало. На примере конкретных практик для решения некоторых задач. |
| **Автоматическая обработка текста** |
| Цель курса - ознакомление студентов с задачами и современными технологиями Natural Language Processing. Планируется ознакомление со статистическими и нейросетевыми подходами в NLP, приобретение навыков их использования в реальных практических задачах, в частности, с помощью библиотеки tensorflow. Особое внимание будет уделено приложениям к задаче машинного перевода |
| **Алгоритмы и структуры данных** |
| &quot;Курс дает базовые знания в области алгоритмов и структур данных. В явном виде они, может быть, не пригодятся, но очень важны для понимания работы библиотек, алгоритмов и языков программирования. Домашние задания по курсу закрепляют полученные знания и воспитывают хороший стиль написания кода, который позволяет избежать стандартных, но от этого ничуть не менее распространенных даже у опытных разработчиков, ошибок.&quot; |
| **Анализ и прогнозирование временных рядов: методы и приложения** |
| How to forecast rates of a national currency? How to identify an oncoming heart attack in good time? – the answers to these questions are associated with the problems (1) to predict a chaotic time series and (2) to reveal typical sequences in an observed time series, respectively. All these problems along with many others comprise the field of time series prediction. The course combines real-world applications with a strong theoretical background: the authors selected mathematical topics required to solve complex problems of actual practice. On the other hand, several topics focus on “mathematics of future” that is theories that will have become the basis of applications in the decades to come. The course starts with simple concepts and gradually works in more advanced applications. To be specific, the course deals with main models to examine and predict regular time series (exemplified by ARIMA and GARCH models), chaotic time series (predictive clustering, constructive neural networks, deep learning models and others) as well as with state-of-the-art approaches used to distinguish regular and chaotic time series, using observations of the time series at hand only; particular topics deal with the concepts of forecasting (time) and stationarity horizons. Applications considered range from econometrics problem to mobile health. |
| **Анализ изображений** |
| Курс познакомит слушателей с основными методами анализа изображений. Будут рассматриваться различные практические задачи: сегментация изображений, обнаружение дубликатов, поиск по подобию, обнаружение текста, распознавание объектов, классификация и другие. В рамках курса у студентов будет возможность узнать и опробовать на практике классические алгоритмы; прочесть множество научных статьей; самостоятельно реализовать алгоритмы, описанные в статьях; поставить ряд экспериментов; предложить и реализовать свои собственные идеи по решению практических задач. |
| **Анализ изображений и видео** |
| &quot;Курс посвящен методам и алгоритмам компьютерного зрения, т.е. извлечения информации из изображений и видео. Рассмотрим основы обработки изображений (шумоподавление, тональную коррекцию, выделение краёв), классификации изображений (основные признаки), поиску изображений по содержанию (сжатие дескрипторов, приближенные методы сравнения дескрипторов), распознавание лиц, сегментацию изображений. Затем поговорим про алгоритмы обработки и анализа видео. Последняя часть курса посвящена трёхмерной реконструкции. Для большинства задач будем обсуждать сущестующие нейросетевые модели. В курсе мы стараемся уделять внимание только наиболее современным методам, которые используются в настоящее время при решении практических и исследовательских задач. Курс в большей степени является практическим, а не теоретическим. Поэтому все лекции снабжены лабораторными и домашними работами, которые позволяют попробовать на практике большинство из рассматриваемых методов. Работы выполняются на языке Python, с использованием различных библиотек.&quot; |
| **Анализ сетевых структур** |
| The course “Network Science” introduces students to new and actively evolving interdisciplinary field of network science. Started as a study of social networks by sociologists, it attracted attention of physicists, computer scientists, economists, computational biologists, linguists and others and become a truly interdisciplinary field of study. In spite of the variety of processes that form networks, and objects and relationships that serves as nodes and edges in these networks, all networks poses common statistical and structural properties. The interplay between order and disorder creates complex network structures that are the focus of the study. In the course we will consider methods of statistical and structural analysis of the networks, models of network formation and evolution and processes developing on network. Special attention will be given to the hands-on practical analysis and visualization of the real world networks using available software tools and modern programming languages and libraries. |
| **Беспроводные сети для Интернета вещей** |
| В данном курсе студенты познакомятся с концепцией Интернета вещей и основными технологиями беспроводной связи, используемыми для её реализации. Студенты изучат базовые аспекты беспроводных сетей для Интернета вещей, а также получат подробные знания об устройстве таких технологий, как Wi-Fi HaLow, LoRa/LoRaWAN, SigFox, RPMA, ZigBee, EC-GSM-IoT, NB-IoT. Много внимания также будет уделяться вопросам математического моделирования данных технологий, а также открытым научным задачам, касающихся беспроводных сетей для Интернета вещей. |
| **Введение в методы машинного обучения и майнинга данных** |
| The course “Machine Learning and Data Mining” introduces students to new and actively evolving interdisciplinary field of modern data analysis. Started as a branch of Artificial Intelligence, it attracted attention of physicists, computer scientists, economists, computational biologists, linguists and others and become a truly interdisciplinary field of study. In spite of the variety of data sources that could be analyzed, objects and attributes that from a particular dataset poses common statistical and structural properties. The interplay between known data and unknown ones give rise to complex pattern structures and machine learning methods that are the focus of the study. In the course we will consider methods of Machine Learning and Data Mining. Special attention will be given to the hands-on practical analysis of the real world datasets using available software tools and modern programming languages and libraries. |
| **Введение в мобильную робототехнику** |
| В курсе будут рассмотрены алгоритмы, составляющие основу современных мобильных роботов: алгоритмы локализации (фильтр Калмана, многочастичный фильтр), построения пути (алгоритм A\* и его модификации), алгоритмы управления (PID-регулятор, контроллер следования по траектории, контроллер поддержания заданной скорости). Также будут рассмотрены кинематические модели различных колесных роботов и модели измерений распространенных сенсоров (видеокамер, ультразвуковых датчиков, лазерных дальномеров). Часть рассмотренных алгоритмов будет реализована студентами на языке Python на практических занятиях и в качестве домашних заданий. Курс посвящен изучению самого популярного фреймворка для создания программного обеспечения для роботов - Robot Operating System (ROS). В курсе будут рассмотрены основные понятия и концепции программирования с использованием ROS, а также входящие в состав ROS инструменты для визуализации, симуляции и отладки различных аспектов программного обеспечения роботов. На практических занятиях мы разработаем несколько модулей ПО (управление перемещением с клавиатуры, модуль локализации) и протестируем их с симуляциях роботов, входящих в состав ROS. |
| **Введение в представление знаний** |
| (пусто) |
| **Введение в технологию блокчейн** |
| Технология блокчейн была впервые реализована в криптовалюте Биткоин и в последствии нашла применение во многих областях: государственные реестры, цепочки управления поставками, биомедицина, финансовый сектор и пр. В курсе будет рассказано, что такое блокчейн, какие возможности и ограничения есть у данной технологии, а также, о существующих и перспективных приложениях. Целью освоения дисциплины является изучение технологии блокчейн (распределенного реестра) с акцентом на её математические и технические основы. Курс предназначен для новичков, желающих познакомиться с данной технологией. |
| **Вероятностно-статистические методы в теории принятия решений** |
| &quot;За последние двадцать лет существенно возросла потребность в решении ряда практических задач таких как автоматическое обнаружение неисправностей (разладок, сбоев, аномалий и т.п.), обслуживание оборудования на основе автоматического контроля его состояния, обеспечение безопасности сложных технических и информационных систем (самолетов, судов, ракет, ядерных электростанций, различных интернет сервисов, и т.д.), автоматический контроль качества выпускаемой продукции, предсказание естественных катастрофических явлений (землятресения, цунами, и т.д.), мониторинг в биомедицине и финансовой сфере. Основная черта вышеперечисленных задач состоит в том, что по сути все они сводятся к выявлению момента резкого изменения (разладки) некоторых характеристик рассматриваемого объекта на основе статистических данных о других характеристиках этого объекта и/или детектированию наблюдений, которые по тем или иным статистическим свойствам значительно отличаются от большинства наблюдений из исследуемой генеральной совокупности. С развитием информатики появилась возможность построения автоматизированных информационных систем для статистической обработки огромного объема реальных данных с целью вынесения тех или иных суждений об истинных характеристиках процесса. Для создания таких систем с привлечением программных средств требуется прежде всего разработка соответствующих фундаментальных математических методов обработки поступающей и поступившей информации исходя из естественных критериев оптимальности. Именно поэтому в данном курсе: Рассматривается общая теория оптимальной остановки стохастических процессов, позволяющая строить оптимальные методы обработки поступающей и поступившей информации, в том числе и методы скорейшего обнаружения разладки. Рассматриваются особенности использования методов скорейшего обнаружения разладки и детектирования аномалий для решения практически задач в различных областях науки и техники. Предлагаемый курс позволит студентам с одной стороны ознакомиться с основными теоретическими свойствами алгоритмов скорейшего обнаружения разладки, а с другой стороны познакомит их с особенностями практической реализации и примерами применения современных алгоритмов скорейшего обнаружения разладки и детектирования аномалий |
| **Восстановление функциональных закономерностей из эмпирических данных** |
| Восстановление функциональных закономерностей из эмпирических данных&quot;&quot; рассматриваются вопросы восстановления (поиска, оценки) функциональных закономерностей по данным наблюдений. Вся наука построена на наблюдении и обобщении результатов в виде зависимостей, законов и т.д. В современном мире в качестве наблюдений выступает информация, обрабатывая которую исследователи стремятся получить знания либо о конкретных объектах (обнаружение брачных аферистов на сайтах знакомств), либо о процессах (прогноз риска развития заболевания по генетическим признакам), либо о явлениях (предсказание землетрясений). Всё это восстановление зависимости из эмпирических данных. При этом необходимо ответить на три фундаментальных вопроса: Какую зависимость восстанавливать Как это сделать Насколько хорошо получилось &quot;&quot;Обычные&quot;&quot; ответы на эти вопросы следующие: Надо восстанавливать линейные зависимости: минимальное число параметров + часто находится аналитическое решение, что гарантирует быстрые и точные вычисления. Если теория описывает явление нелинейно, то его надо линеаризовать, например прологарифмировать. Использовать метод наименьших квадратов (либо метод максимального правдоподобия) для оценки параметров линейной зависимости - модели. Считая, что данные получены в результате независимых испытаний из некоего вероятностного распределения (часто нормального) вычисляют (оценивают) доверительный интервал, покрывающий с заданной вероятностью (обычно 95%) &quot;&quot;истинное&quot;&quot; значение параметра модели. Если этот интервал &quot;&quot;мал&quot;&quot;, и достаточно оснований считать, что линейная модель адекватна описываемому процессу, то задача решена хорошо. В рамках методов машинного обучения, которые расширенно трактуются как методы анализа данных, каждый из этих пунктов получил развитие. А именно: Какую зависимость восстанавливать Переменной у может быть двоичной у={0,1} - это задача классификации (распознавания образов). Переменная у может принимать любые действительные значения - это задача регрессии. |
| **Выпуклый анализ и оптимизация** |
| Оптимизация является одной из основных рабочих лошадок, используемых в машинном обучении. Наш курс состоит из трех частей: Вводная часть о выпуклых множествах и выпуклых функциях, условиях оптимальности в задачах оптимизации, теории двойственности; Часть, посвященная алгоритмам для численного решения оптимизационных задач. В этой части мы попытаемся проиллюстрировать основные идеи, используемые при разработке методов оптимизации. В этой части мы будем не столько разбирать известные методы, сколько на их примере рассматривать основные концепты; Часть, посвященная задачам оптимизации с неопределенностью. Эта часть интересна сама по себе. Задачи оптимизации с неопределенностью встречаются повсеместно, при этом этой неопределенности может быть самый разный. В стандартных курсах по оптимизации обычно эти вопросы или не затрагиваются вовсе, или же затрагиваются лишь частично. Мы попытались в нашем курсе уделить задачам с неопределенностью больше внимания (конечно, в пределах временных рамок курса). |
| **Глубинное обучение** |
| Предлагаемый курс посвящён методам “глубинного обучения” - нового поколения нейросетевых методов машинного обучения, вызвавших бурный всплеск развития в ряде прикладных областей. В первую очередь курс направлен на формирования у студентов навыков решения прикладных задач при помощи глубоких нейронных сетей. |
| **Глубинное обучение: дополнительные главы** |
| Курс посвящен &quot;глубинному обучению&quot;, т.е. новому поколению методов, основанному на многослойных нейронных сетях и позволившему радикально улучшить работу систем распознавания образов и искусственного интеллекта. В данном курсе будут покрыты все основные темы, но преимущественное внимание будет посвящаться наиболее современным техникам и методам. |
| **Дискретная математика для разработки алгоритмов и программ** |
| This course includes the basics of mathematical logic, graph theory, combinatorics, and formal language theory. The emphasis is put upon the algorithmic side: mathematical results act as a support for effecient algorithms operating on graphs, strings, and, finally, parsing algorithms for regular expressions and contextfree grammars. The course is actually twofold: besides usual «chalk-and-blackboard» mathematical part, it also includes a practical one, i.e., implementing the algorithms discussed in the course. The students are supposed and encouraged to (but not restricted to) use the Python language. For the last part, parsing algorithms, PyBison is also welcome. |
| **Дополнительные главы статистической теории обучения** |
| Курс фокусируется на теории, стоящей за алгоритмами классификации и реграссии. Обсуждаются общие вероятностные подходы к задачам предсказания, после чего вводятся элементы теории Вапника-Червоненкиса, предоставляющие теоретическии гарантии минимизации эмпирического риска в контексте бинарной классификации. Также обсуждаются линейные методы и техники регуляризации. Последняя часть посвящена нелинейным алгоритмам: деревьям, ансамблям, SVM, нейронным сетям. |
| **Игры и решения в задачах анализа данных и моделирования** |
| The aim of the course &quot;Games and Decisions in Data Analysis and Modeling&quot; is to familiarize students with modern models of game theory and decision theory, their applications in modeling and analyzing socio-economic problems, as well as their use in analytical and decision support systems. The course covers fundamental topics in decision theory: individual preferences modelling using binary relations and choice functions, social choice theory, especially the theory of local voting procedures and the theory of majority rule-based solutions. Also, students will consider the problems of decision-making in the network models of participants interaction and models of strategic behavior of players, taking into account the network structure of connections. Game theory studies the strategic interaction of rational agents and plays a central role in the economics, but it is also widely used in biology, political science, military affairs, etc. In this course, we will study non-cooperative and cooperative games, as well as the mechanism design, which is the reverse task, i.e. the development of rules for the interaction of agents that leads to the desired result. Successful completion of the course contributes to the development of useful strategic thinking in life. |
| **Инжиниринг данных и сервисов для автоматизации бизнес процессов** |
| Machine learning is changing the world rapidly and dramatically, every modern enterprise is now eyeing machine learning as one of the top instruments to improve business KPIs. Yet, behind any successful application of machine learning is a large chunk of work that is done by engineers, which includes Data Engineering functions such as data cleaning, wrangling, integration, etc. And the models must be deployed in production as reliable services. And finally, advanced analytics will need to take place in order to understand how the service is operating. In this course you will learn the basics of these engineering and analytic disciplines. We won’t focus on machine learning algorithms in this course, its a prerequisite. |
| **Информационный поиск** |
| &quot;В курсе рассматриваются общие вопросы построения информационно-поисковых систем: задачи информационного поиска и архитектура поисковых систем, машинное обучение в поиске и компьютерная лингвистика, построение поискового индекса и обнаружение дубликатов, поисковый робот и оценка качества. Решение предлагаемых практических заданий связано со знакомством с широким спектром технологий и алгоритмов, применяемых на практике при построении компонентов поисковой системы .&quot; |
| **Компьютерная лингвистика и анализ текста** |
| Настоящая дисциплина охватывает изучение различных моделей автоматической обработки текста на естественном языке (ЕЯ), применяемых в современных информационных системах и затрагивающих несколько языковых уровней обрабатываемого текста, включая уровни морфологии, синтаксиса, дискурса и семантики. Изучаются также виды лингвистических ресурсов, используемых при обработке текстов, и методы их создания. Рассматриваются прикладные задачи, требующие многоуровневого анализа и синтеза текста (такие как машинный перевод, классификация и кластеризация текстов, извлечение информации и знаний из текста). |
| **Машинное обучение** |
| Теория обучения машин (machine learning, машинное обучение) находится на стыке прикладной статистики, численных методов оптимизации, дискретного анализа, и за последние 50 лет оформилась в самостоятельную математическую дисциплину. Методы машинного обучения составляют основу ещё более молодой дисциплины — интеллектуального анализа данных (data mining). В курсе рассматриваются основные задачи обучения по прецедентам: классификация, кластеризация, регрессия, понижение размерности. Изучаются методы их решения, как классические, так и новые, созданные за последние 10–15 лет. Упор делается на глубокое понимание математических основ, взаимосвязей, достоинств и ограничений рассматриваемых методов. Отдельные теоремы приводятся с доказательствами. Все методы излагаются по единой схеме: исходные идеи и эвристики; их формализация и математическая теория; описание алгоритма в виде слабо формализованного псевдокода; анализ достоинств, недостатков и границ применимости; пути устранения недостатков; сравнение с другими методами. примеры прикладных задач. |
| **Машинный перевод** |
| В данном курсе будут рассматриваться современные подходы к решению задачи машинного перевода. Среди тем, которые будут изучены в рамках курса, будут посвященные важным базовым вопросам перевода, а также такие актуальные, как encoder-decoder на основе нейросетевых методов и работа с low resource языками. |
| **Медицинская информатика** |
| Medical Informatics (MI) is a new, exponentially-growing field, where information sciences meet modern clinical applications. The main goal of this class in to introduce HSE students to the broad spectrum of MI problems and applications, and to provide the students with the skills necessary for conduction professional MI work. |
| **Методы автоматической верификации программ** |
| The course acts as introduction to some of the most successful logic based concepts, tools and techniques used today in CS and IT, which are behind a major breakthrough in the practical applications in verification of systems and software. These methods are proven to be of great theoretical and practical potential in CS and IT. The topics of this course fall under the umbrella of what is called verification (a) Verification means to verify that a system satisfies some property. (b) The system can be a physical or software system. (c) The property is expressed using specifications within a certain logical language. The course has been designed in accordance with the most recent trends in the program verification. The challenge has been twofold: (1) to select the material and design the course so that to make it meet the actual needs of the CS applications, and (2) to do that so that to allow the students to learn and digest all necessary ideas to efficiently prove that the programs behave in the correct way. |
| **Методы и системы обработки больших данных** |
| Курс посвящен архитектуре инфраструктурных распределенных систем. В курсе студенты познакомятся с такими понятиями, как: scalability, fault-tolerance, consensus, distributed storage, map-reduce, zookeeper, etc. У них будет возможность познакомиться с тем, какие проблемы стоят перед разработчиками распределённых систем и как они решаются в промышленных системах. |
| **Методы машинного обучения и майнинга данных** |
| The course &quot;Machine Learning and Data Mining&quot;; introduces students to new and actively evolving interdisciplinary field of modern data analysis. Started as a branch of Artificial Intelligence, it attracted attention of physicists, computer scientists, economists, computational biologists, linguists and others and become a truly interdisciplinary field of study. In spite of the variety of data sources that could be analyzed, objects and attributes that from a particular dataset poses common statistical and structural properties. The interplay between known data and unknown ones give rise to complex pattern structures and machine learning methods that are the focus of the study. In the course we will consider methods of Machine Learning and Data Mining. Special attention will be given to the hands-on practical analysis of the real world datasets using available software tools and modern programming languages and libraries. |
| **Методы оптимизации в машинном обучении** |
| Методы оптимизации лежат в основе решения многих задач компьютерных наук. Например, в машинном обучении задачу оптимизации необходимо решать каждый раз при настройке какой-то модели алгоритмов по данным. Причём от эффективности решения соответствующей задачи оптимизации зависит практическая применимость самого метода машинного обучения. Данный курс посвящен изучению классических и современных методов решения задач непрерывной оптимизации (в том числе невыпуклой), а также особенностям применения этих методов в задачах оптимизации, возникающих в машинном обучении. Основной акцент в изложении делается на практические аспекты реализации и использования методов. Целью курса является выработка у слушателей навыков по подбору подходящего метода для своей задачи, наиболее полно учитывающего её особенности. |
| **Методы теоретической информатики** |
| Учебная дисциплина «Методы теоретической информатики» предназначена для обучения студентов специальности «Теоретическая информатика» основным приемам и методам решения и анализа задач в этой области. Среди всего разнообразия методов теоретической информатики для изучения в данной дисциплине выбраны те, которые наиболее активно используются в современных исследованиях по актуальным направлениям теоретической информатики. |
| **Научно-исследовательский семинар "Анализ Интернет-данных"** |
| Основной целью научно-исследовательского семинара &quot;Анализ интернет-данных&quot; является ознакомление студентов с последними достижениями и тенденциями машинного обучения и формирование методом разбора научных статей на заданные тематики. |
| Основной целью научно-исследовательского семинара &quot;Анализ интернет-данных&quot; является ознакомление студентов с последними достижениями и тенденциями машинного обучения. Обзор основных фреймворков для глубинного обучения: PyTorch, tensorflow, MXNet. Анализ выложенного в открытый доступ кода современных статей. Обзор лучших практик при написании воспроизводимого кода для научных экспериментов. |
| **Научно-исследовательский семинар "Интеллектуальные системы и структурный анализ"** |
| The discipline goal is to develop students' professional skills in applied fields of the computer science. As part of the Scientific Seminar, the course &quot;Introduction to the Number Theory and its Applications&quot; will be read during the first semester and the course “Introduction to the Semantic Web Technologies” during the second semester. The course &quot;Introduction to the Number Theory and its Applications&quot; is aimed at the formation and development of theoretical-numerical thinking, as well as the understanding of the fundamental concepts of number theory and the ways in which they can be used in cryptography. The course “Introduction to the Semantic Web Technologies” is a gentle introduction to the theory and practice of the Semantic Web, an extension of the current Web that provides an easier way to find, share, reuse and combine information. It is based on machine-readable information and builds on XML technology's capability to define customized tagging schemes, RDF's (Resource Description Framework) flexible approach to representing data, the OWL (Web Ontology Language) schema language and SPARQL query language. The Semantic Web provides common formats for the interchange of data (where on the Web there is only an interchange of documents). It also provides a common language for recording how data relates to real world objects, allowing a person or a machine to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing. Important applications of the Semantic Web technologies include Healthcare (SNOMED CT), Supply Chain Management (Biogen Idec), Media Management (BBC), Data Integration in the Oil &amp; Gas industry (Chevron, Statoil), Web Search and E-commerce. |
| The discipline goal is to develop students' professional skills required for independent analytical work in applied fields of the computer science. Also, this course aims to improve skills of students in developing their research projects related with dialogue systems and chat bots. This course focuses on analysis of scientific and industrial linguistic system developing and motivates visiting different scientific colloquium at the university, especially at the faculty of computer science. |
| **Научно-исследовательский семинар "Теоретическая информатика"** |
| Целями освоения научно-исследовательского семинара «Теоретическая информатика» являются освоение основ теоретической информатики, вычислительной логики и искусственного интеллекта, студентов основным понятиям и методам дискретной математики, необходимым как в дальнейшем обучении, так и в работе по специальности. Это даст участникам семинара общее представление об указанных выше областях, а также позволит в дальнейшем заниматься более продвинутыми разделами этих областей. |
| Целями освоения научно-исследовательского семинара «Теоретическая информатика» являются освоение студентами основ теоретической информатики, вычислительной логики и искусственного интеллекта, основным понятиям и методам дискретной математики, необходимым как в дальнейшем обучении, так и в работе по специальности. Это даст участникам семинара общее представление об указанных выше областях, а также позволит в дальнейшем заниматься более продвинутыми разделами этих областей. Курс организован в виде семинара, на котором планируются выступления как участников семинара, так и гостей факультета - ведущих специалистов в указанных выше областях науки. |
| **Научно-исследовательский семинар "Технологии моделирования сложных систем"** |
| Курс направлен на ознакомление студентов специализации «Технологии моделирования сложных систем» с тематикой научных исследований, проводимых в ИППИ РАН, а также формирование у студентов навыков самостоятельной исследовательской работы |
| Курс Научно-исследовательский семинар &quot;Технологии моделирования сложных систем направлен на формирование и улучшение навыков самостоятельной исследовательской работы студентов и совершенствование навыков представления полученных в ходе исследования результатов |
| **Нейробайесовские методы в машинном обучении** |
| Курс посвящен применению байесовских методов в глубинном обучении. На лекциях будет рассказано о применении вероятностного моделирования для построения порождающих моделей данных, использованию состязающихся сетей для приближенного вывода, моделированию неопределенности в параметрах нейронной сети и о некоторых открытых проблемах глубинного обучения. |
| **Неопределённость и нечёткость при анализе данных и принятии решений** |
| Дисциплина «Неопределенность и нечеткость при анализе данных и принятии решений» структурно состоит из двух частей. Первая часть посвящена обучению навыкам работы с нечеткими данными. Во второй части рассматриваются некоторые модели описания неопределенности, возникающие при принятии решений и анализе данных. В целом данный курс позиционируется в рамках сравнительно нового научного направления «обобщенной теории информации» (G.J. Klir). Обобщение классической теории информации осуществляется в двух направлениях. Первое связано с обобщением понятия множества – от классического понятия множества, где каждый элемент некоторого универсума либо принадлежит данному множеству, либо – нет, до понятия нечеткого множества, в котором все элементы универсума принадлежат данному множеству с некоторой степенью. Второе – обобщением понятия меры: от классического понятия меры множества (например, вероятностной меры) с сильной аксиомой аддитивности, до понятия монотонной (в общем случае неаддитивной) меры. Например, одной из популярных (и сложных) задач финансовой аналитики является прогнозирования курса валют или акций на валютной или фондовой бирже соответственно. Классическим инструментом прогнозирования является регрессионный анализ. При этом если для анализа используются точечные данные курса (например, цены закрытия) за некоторый предшествующий период времени, то могут быть применены классические методы регрессионного анализа. Но вместо точечных данных можно рассматривать интервалы курсов за периоды торгов или более сложные – нечеткие данные. В этом случае могут быть использованы методы нечеткого регрессионного анализа. С другой стороны, поведение курса валют или акций характеризуется большой степенью неопределенности, зависящей от ряда трудно формализуемых и трудно предсказуемых факторов: политических, психологических, макроэкономических и пр. В этом случае чаще всего приходится полагаться на экспертные оценки. Но небольшое количество экспертных оценок, их возможная зависимость, противоречивость, различная степень надежности, исключают построение надежных статистических оценок. Поэтому такую информацию удобно представлять с помощью монотонных мер или с помощью других моделей в рамках более общей теории неточных вероятностей (imprecise probabilities). Монотонные меры не требуют своего задания на множестве (алгебре) всех подмножеств элементарных событий в отличие от вероятностной меры. Поэтому с их помощью можно моделировать ситуации отсутствия информации, ситуации, когда субъект (эксперт, классификатор, метод кластеризации и т.д.) готов определять степень принадлежности истинной альтернативы только некоторым (но не всем!) подмножествам (например, принадлежность образа объединению некоторых классов). Все эти и многие другие вопросы будут рассмотрены в анонсируемой дисциплине. Основными темами курса являются: - нечеткие множества и операции над ними; - нечеткие отношения; - принцип обобщения Заде и нечеткие числа, сравнение нечетких чисел; - принятие решений при нечетких данных; - нечеткая классификация и кластеризация; - нечеткая регрессия; - виды неопределенности при принятии решений и анализе данных; - монотонные меры и основные классы монотонных мер; - элементы теории свидетельств (функций доверия); - моделирование неопределенности при принятии решений. |
| **Обработка естественного языка на Python** |
| The course will cover the main tools and libraries of the Python language for natural language processing (NLTK - natural language toolkit). Tokenization, stemming, morphological and syntactic parsing, working with semantics, access to corpora and lexical resources, text classification, search for named entities. |
| **Обучение представлений и глубокое обучение** |
| &quot;Курс посвящен &quot;&quot;глубинному обучению&quot;&quot;, т.е. новому поколению методов, основанному на многослойных нейронных сетях и позволившему радикально улучшить работу систем распознавания образов и искусственного интеллекта. В начале курса мы рассмотрим основы глубокого обучения (и, отчасти, &quot;&quot;крупномасштабного&quot;&quot; машинного обучения вообще). Следующие четыре лекции будут посвящены детальному рассмотрению сверточных нейросетей и их применениям в области компьютерного зрения (на данный момент сверточные нейросети в компьютерном зрении &amp;mdash;&amp;mdash; самый бесспорный и громкий успех глубокого обучения). Затем мы рассмотрим рекуррентные нейросети и то, как они применяются в задачах, связанных с естественным языком, а также совсем недавно появившиеся модели, направленные на моделирование более универсального искусственного интеллекта и делающие первые шаги в этом направлении.&quot; |
| **Обучение с подкреплением** |
| Цель курса - ознакомление студентов с основными подходами и алгоритмами обучения с подкреплением. Планируется дать понимание того, какие существуют подходы к решению задач обучения с подкреплением, научить выбирать подход и алгоритм, наиболее подходящий для рассматриваемой студентом задачи, научить обучать модели с использованием современных нейросетевых библиотек. |
| **Обучение современному C++** |
| С++ - мощный язык с богатым наследием. Тем, кто только ступил на путь освоения этого языка, очень просто заблудиться в изобилии техник и приёмов, созданных за последние 30 лет. Курс учит &quot;Modern C++&quot; - современному подмножеству языка (стандарты 14, 17 и 20). Много внимания уделяется инструментам и библиотекам - вещам которые не являются частью языка, но без которых не получится построить большой и сложный проект. |
| **Основы стохастики. Стохастические модели** |
| &quot;Принятие рациональных решений может быть осуществлено только на основе исторической информации о поведении рассматриваемого объекта. Известная историческая информация об объекте агрегируется в модели объекта, призванной отразить основные (важные для исследователя) особенности его поведения. Полный (&quot;&quot;производственный&quot;&quot;) цикл построения модели объекта состоит из: сбора данных; собственно построения математической модели объекта, задаваемой системой соотношений между математическими величинами, характеризующими основные особенности объекта; идентификации математической модели по собранным данным; валидации модели и последующего уточнения модели. Далее на основе модели объекта происходит прогнозирование его поведения и принятие соответствующих рациональных решений. При принятии рациональных решений в любой из сфер человеческой деятельности необходимо учитывать случайную природу изменений, происходящих с окружающей средой. По этой причине именно стохастические модели широко используются в социальных, физических, инженерных и других науках. Построение стохастических моделей и их использование на практике невозможно без твердых знаний основных математических инструментов, позволяющих работать со случайностью. Именно поэтому предлагаемый вероятностно-статистический цикл &quot;&quot;Стохастика&quot;&quot;, состоящий из лекций и семинарских занятий, призван дать изложение основных понятий теории вероятностей и теории случайных процессов, необходимых для последующего курса лекций и семинарских занятий (весенний семестр) по &quot;&quot;Вероятностно-статистическим методам в теории принятия решений&quot;&quot;, являющихся, по существу, &quot;&quot;стохастическими&quot;&quot;. Изложение лекций будет следовать исторической схеме становления стохастических дисциплин, что дает лучшее представление об эволюции и значении основных вероятностно-стохастических концепций, моделей, методов, оперирующих с понятием &quot;&quot;случайность&quot;&quot;. Также курс позволит студентам ознакомиться с основными типами стохастических процессов и позволит овладеть всеми необходимыми инструментами для того, чтобы разрабатывать методы скорейшего обнаружения разладки и детектирования аномалий и использовать эти методы в задачах мониторинга состояния различных систем.&quot; |
| **Основы физического уровня: лаборатория по аналоговым и цифровым методам** |
| Технологии программно-определяемого радио (SDR) позволяют свести разработку устройств радиосвязи к программированию. При выполнении этого курса вы превратите свой компьютер в рацию с помощью программно-определяемого приёмо-передатчика. Данный курс покрывает работу с аналоговыми сигналами. Цифровые сигналы рассматриваются в следующем курсе. Современные системы связи и хранения данных в своей основе используют среды, по которым всё ещё распространяются аналоговые величины: амплитуды радиоволн, намагниченности жестких дисков и другие. В данном курсе кратко описано, какие они, как много и чего в них общего, какие есть отличия, какие перспективы развития, как сформировать сигнал, чтобы его передать и принять. |
| **Параллельные и распределённые вычисления** |
| Целью освоения дисциплины «Параллельные и распределённые вычисления» является знакомство с параллельными и распределенными вычислениями, различными классами высокопроизводительных систем, принципами реализации параллельных алгоритмов и используемыми моделями программирования, а также получение навыков практического использования соответствующих технологий и систем при решении при-кладных задач. |
| **Практическая аналитика** |
| Целью курса &quot;Практическая аналитика&quot; является получение опыта решения реальных задач: нечётко поставленных, требующих дополнительного уточнения, в которых необходимо самостоятельно собирать данные, подбирать метрики и т.п. Основная часть курса - это разбор кейсов из реальных задач Яндекса. |
| **Предсказательное моделирование** |
| Сутью методологии предсказательного моделирования является построение «облегченных» математических моделей поведения сложных многокомпонентных систем, позволяющих сочетать простоту вычислений по модели с достаточной точностью и надежностью. Этот подход является альтернативой традиционному подходу в математическом моделировании, когда модель строится «из первых принципов» как адекватное математическое описание собственной динамики системы. Предсказательное моделирование опирается на такие области математики, как теория аппроксимации, обучение машин, математическая статистика, теория информации, теория игр. |
| **Прикладная количественная логистика** |
| Basic quantitative logistics models and algorithms for Allocation, Transportation, and Vehicle Routing Problems including the sensitivity and stability analysis applied to the Minimum Spanning Tree Problem (MSTP) and its variations, Shortest Path, Traveling Salesman and Vehicle Routing Problems (Capacitated, Time Windows, Pick Up and Delivery, etc.), Preempted Single Machine Scheduling Jobs (Operations) with Arbitrary Processing Times, Release and Due Dates will be studied in this course. |
| **Прикладная линейная алгебра** |
| Линейная алгебра является одной из базовых математических дисциплин. Несмотря на её фундаментальный характер, у нее есть достаточно много «прямых», но вполне эффектных приложений. Наш курс посвящен знакомству магистрантов с такими приложениями и методами, которые, при всей своей элементарности, обычно находятся за пределами стандартных курсов бакалавриата. |
| **Проектный семинар "Анализ Интернет-данных"** |
| Основная тема семинара - алгоритмы работы с большими данными. Одно из наиболее бурно растущих направлений информационных технологий в настоящее время - это работа с большими объемами данных. Множество известных IT-компаний (к примеру, Яндекс, Google, Facebook, Twitter, LinkedIn) ежедневно обрабатывают огромные массивы информации для решения своих бизнес-задач. Необходимость работы с большими данными существенно изменяет принципы построения всех аспектов информационных систем: хранения и анализа данных, планирования задач, координации сервисов, эффективности использования ресурсов. Среди руководителей и участников семинара - сотрудники компании Яндекс с многолетним опытом построения и эксплуатации различных распределенных систем. В начале осеннего семестра мы расскажем о некоторых актуальных задачах области. Далее, в течение семестра на семинаре студенты будут выступать с докладами по современным достижениям в области, также будут приглашаться внешние докладчики с рассказами об их практической деятельности. Слушатели семинара смогут получить представление об актуальных проблемах и задачах индустрии, определиться с темой их научной работы. |
| **Проектный семинар "Интеллектуальные системы и структурный анализ"** |
| The course consists of student presentations on the recent research in various fields of computer science and data mining. The course is divided into two sections: the first part is organized to expand the horizons of students in the field of computer science and to test students' readiness to present scientific reports, and the second part is making a review on a scientific report. This course focuses on analysis of interdisciplinary research papers and motivates visiting different scientific colloquium at the university, especially at the faculty of computer science. In addition, it is assumed that the course will help in choosing the topic of scientific research for those students who have not decided yet. |
| **Проектный семинар "Теоретическая информатика"** |
| Целью семинара &quot;Теоретическая информатика&quot; является изучение дискретной математики и информатики. В ходе семинара участники разбираются в современных статьях по указанным темам. Возможными темами для докладов может быть модальная логика, графы-экспандеры, задача о равенстве нулю многочлена, параметрическая сложность задачи о выполнимости булевой формулы. |
| **Проектный семинар "Технологии моделирования сложных систем"** |
| Математика - это такой же язык, как естественный русский или английский: иногда, одни и те же слова разными людьми интерпретируются по-разному; иногда, абсолютно разные слова могут разными людьми интерпретироваться одинаково. Вообще, интерпретируемость - очень сложная вещь, которую не так просто формализовать (если вообще возможно). Тем не менее, можно сформулировать некоторые концепты, в рамках которых пытаться объяснять предсказания тех или иных математических моделей, и снабдить людей инструментами для этого. Этот курс знакомит слушателей с идеями зарождающегося (на мой взгляд) направления в машинном обучении, которое можно назвать “Интерпретируемое ML”. Лекции призваны дать обзор того, что сейчас происходит в данной области (скорее всего, этот обзор не будет полным, но, по крайней мере, он может послужить небольшим введением), и показать слушателям набор инструментов, с которыми можно работать в рамках предложенных идей. |
| **Робастные методы в статистике** |
| Учебная дисциплина «Робастные методы в статистике» является логическим продолжением и развитием курса «Теория вероятностей и математическая статистика», где будет показано, каким образом необходимо формулировать задачу, где предполагается использование статистических методов, а так же каким образом необходимо адаптировать классические статистические методы принятия решений в зависимости от условий статистического эксперимента, на базе которого принимается решение. Такой подход необходим в тех случаях, когда необходимо принять гарантийное решение, т.е. такое решение, риски которого могут быть надежно оценены. Будут рассмотрены необходимые модификации классических статистических методов тех случаях, когда вероятность ошибки очень мала и решение должно приниматься не на основе классических предельных теорем теории вероятностей, а на основании теории больших уклонений. Будут рассмотрены алгоритмы, позволяющие правильно учитывать «выбросы» в выборке, которые обычно рассматриваются как ошибочные наблюдения и отбрасываются. Сегодня грамотный практик обязан учитывать множество факторов, влияющих на процесс формирования данных наблюдений, и использовать статистические решения, устойчивые к негативному влиянию неконтролируемых негативных факторов, влияющих на результаты наблюдений и дисциплина «Робастные методы в статистике» позволит правильно выбирать методы обработки данных в зависимости от природы формирования данных и целей их обработки. |
| **Современные матричные вычисления** |
| Методы вычислительной линейной алгебры и матричные вычисления являются базовым компонентом при решении различных задач статистического анализа, обработки изображений, предсказательного моделирования. Владение основными алгоритмами матричных вычислений необходимо в любой сфере математического моделирования. Изучение курса «Современные матричные вычисления» требует предварительных знаний в объеме первых курсов стандартной бакалаврской программы по направлению Прикладная математика и информатика или другой смежной программы. |
| **Современные методы анализа данных** |
| This is a course in basic methods for modern Data Analysis. Its contents are heavily influenced by the idea that data analysis should help in enhancing and augmenting knowledge of the domain as represented by the concepts and statements of relation between them. This view distinguishes the subject from related courses such as applied statistics, machine learning, data mining, etc. Two main pathways for data analysis are: (1) summarization, for developing and augmenting concepts, and (2) correlation, for enhancing and establishing relations. Visualization, in this context, is a way of presenting results in a cognitively comfortable way. The term summarization is understood quite broadly here to embrace not only simple summaries like totals and means, but also more complex summaries: the principal components of a set of features and cluster structures in a set of entities. Similarly, correlation here covers both bivariate and multivariate relations between input and target features including classification trees and Bayes classifiers. Another feature of the class is that its main thrust is to give an in-depth understanding of a few basic techniques rather than to cover a broad spectrum of approaches developed so far. Most of the described methods fall under the same least-squares paradigm for mapping an “idealized” structure to the data. This allows me to bring forward a number of mathematically derived relations between methods that are usually overlooked. |
| **Современные методы принятия решений** |
| This course presents an introduction to the mathematical foundations of statistical learning theory. The presentation is oriented towards the most important algorithms and methods in the field. Topics studied include: empirical risk minimisation, local averaging, boosting and support vector machines. We will also provide an introduction to online and reinforcement learning methods. |
| **Теория вероятностей и математическая статистика** |
| The overwhelming majority of courses on data-mining and artificial intelligence implies a strong background in probability and statistics. The goal of this course is to provide those students who are not so keen on the subject matter with its fundamentals. |
| **Теория вычислений** |
| This course teaches a mathematical theory that helps to invent better algorithms. With “better” we mean that the algorithms use fewer resources such as time or memory. We also consider parallel computation, distributed systems and learning problems. In these settings we might also optimize other types of resources. For example, in distributed systems we might want to minimize the amount of communication. We focuss on worst case guarantees. A large part of our time is devoted to the study of what is not possible. In other words, we study fundamental barriers for the existence of programs that use fewer resources than a given bound. |
| **Теория статистического обучения** |
| We study a theory that inspired the development of two important families of machine learning algorithms: support vector machines and boosting. More generally, in a typical classification task we are given (1) a dataset for training and (2) a familyof classifiers, indexed by some parameters that need to be tuned. A learning algorithm uses the training set to select one of the classifiers, in other words, tune the parameters. For example, given a neural network, every choice of the weights specifies a classifier and the learning algorithm assigns values to the weights. On one side, we want to have a large set of classifiers to model all structure in the training data. This might lead to a large number of parameters to train. On the other hand a too large set of classifiers might lead to a decrease of accuracy on unseen examples. This is called overfitting. We study a theory that quantifies overfitting theoretically. Moreover, support vector machines and boosting algorithms can be seen as algorithms that optimize the trade-off discussed above under some additional assumptions. Moreover, the theory can determine good values for meta-parameters in machine learning algorithms that otherwise need to be tuned using cross-validation. We also study some recent deep boosting algorithms that were developed using the theory. These algorithms are currently among the best for classification tasks when the number of classes is high. Finally, we study the online mistake model. This model is more general but its mathematical analysis has many similarities with the theories above. |
| **Упорядоченные множества в анализе данных** |
| “Ordered sets for data analysis” is a first-semester master course within “Data Science” master program. Order relations are ubiquitous, we meet them when we consider numbers, Boolean algebras, partitions, multisets, graphs, logical formulas, and many other mathematical entities. On the one hand, order relations are used for representing data and knowledge, on the other hand they serve as important tools for describing models and methods of data analysis, such as decision trees, random forests, version spaces, association rules, etc. We start from the basic notions of binary relations, their representations with matrices and graphs, consider properties of relations and operations on them, giving examples on problems of data analysis. We come then to the notions of order, semilattice, and lattice. Very important data models like numbers, vectors of numbers, set algebras, partitions make lattices. We consider an important type of lattices, called concept lattices, which constitute the main subject of Formal Concept Analysis (FCA), a modern domain of applied lattice theory. Concept lattices provide very useful tools for data analysis and pattern mining. On the one hand, formal concepts, the elements of a concept lattice, describe sets of objects sharing sets of common attributes, thus presenting a nice model of (bi)clustering. Ordered sets of concepts give a hierarchy of classes of a subject domain, which makes a backbone of domain ontology. On the other hand, concept lattices represent dependencies on attributes, both precise (called implications), and imprecise, known in data mining as association rules. We consider most important algorithms and algorithmic problems related to applications of FCA. Concept lattices allow us to describe some fundamental models of data analysis and pattern mining such as decision trees, random forests, version spaces, and concept-based hypotheses. Since a serious limitation of many methods of pattern mining is computational complexity, complexity analysis makes an important leitmotif of the course. Throughout the course we consider applications of the introduced notions in various applied domains, such as text analysis, medicine, pharmacology, etc. Homework consists of both theoretical problems and practical problems in designing particular. |
| **Функциональный анализ для машинного обучения** |
| Классический, обычно читающийся на математических специальностях университетов, курс функционального анализа включает в себя общую теорию метрических и нормированных пространств, теорию операторов в нормированных пространствах. Будут затронуты аспекты спектральной теории операторов и преобразования Фурье. По возможности будет дано представление о приложениях функционального анализа к анализу данных, машинному обучению и т.д. |
| **Цифровая обработка сигналов** |
| Без применения Цифровой обработки сигналов не обходится ни одна современная система связи. Её использование позволяет значительно упростить и удешевить аналоговую часть многих устройств. В данном курсе изучаются основы Цифровой обработки сигналов: понятия аналогового, дискретного и цифрового сигнала, их спектры, системы их передачи и обработки. После прохождения данного курса студенты будут уметь: проектировать цифровые фильтры и учитывать особенности их реализации, рассчитывать влияние искажений, связанных с неидеальностью АЦП и ЦАП, шумами округления, использовать быстрое преобразование Фурье для повышения эффективности систем обработки сигналов, реализовать системы изменения частоты дискретизации сигнала. |
| **Элементы современной теории кодирования** |
| Современный мир тяжело представить без различных систем передачи и хранения данных. С ежегодно возрастающими требованиями как к скорости, так и к надежности передаваемых данных, все более актуальным становятся задачи теории кодирования и информации. Курс посвящен изучению методов теории помехоустойчивого кодирования, активно используемых в современных системах связи. В рамках курса будут даны элементы теории информации и теории кодирования. Будут описаны методы построения и анализа современных кодовых конструкций таких, как коды с малой плотностью проверок (МПП-коды, LDPC codes), полярные коды (polar codes), сверточные коды (convolution codes) и т.д., а также их алгоритмы кодирования и декодирования. Также будут приведены способны применения данных кодов как для защиты данных, передаваемых по каналам связи, так и для хранения данных в том числе и в распределенных хранилищах. Будут сформулированы открытые на данный момент вопросы и задачи теории кодирования. |
| **Язык Python** |
| Python - язык, ориентированный на скорость и простоту разработки за счет скорости выполнения программ. Это бывает полезно во многих задачах, например в исследовательском программировании, в аналитике, для написания прототипов и т.д. На курсе мы научимся во-первых писать простые программы, а далее будем изучать различные конструкции и библиотеки, позволяющие писать программы быстрее и с меньшим количеством ошибок; а также разрабатывать библиотеки и большие проекты. |