



برنامه کارشناسی ارشد مهندسی کامپیوتر

گرایش علم داده

تهیه شده توسط دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف



برنامه‌ی کارشناسی ارشد مهندسی کامپیوتر – گرایش علم داده

مقررات عمومی گرایش علم داده

۱. طول دوره: طول دوره ۲ سال است.

۲. زیرگرایش‌ها: گرایش علم داده دارای دو زیرگرایش به قرار زیر است (دروس متعلق به زیرگرایش‌ها در جداول ۳ و ۴ آورده شده‌اند):

• زیرگرایش مهندسی داده

• زیرگرایش تحلیل داده

۳. واحدها: تعداد واحدهای این دوره ۳۰ واحد است. دانشجویان حق ندارند دروسی را که قبلاً در دوره‌ی کارشناسی گرفته‌اند مجدداً اخذ نمایند. گذراندن دروس زیرگرایش‌ها، سمینار و پروژه باید مطابق قواعد زیر انجام شود:

• **دروس:** همه دانشجویان گرایش علم داده باید درس‌های جبرانی را در دوران کارشناسی گذرانده باشند و در غیر این صورت باید این درس‌ها را در دوران کارشناسی ارشد اخذ نمایند. همه دانشجویان گرایش علم داده باید ۱۰ واحد دروس هسته گرایش علم داده (جدول ۲) را اخذ نمایند. برای سایر درس‌ها دانشجویان بایستی به شکل زیر عمل نمایند.

○ دانشجویان پژوهش‌محور: دانشجویان پژوهش‌محور باید حداقل ۳ درس معادل ۹ واحد را از دروس زیرگرایش خود اخذ نمایند (جدول ۳ و ۴). درس باقیمانده (۳ واحد باقیمانده) را می‌توانند از هر کدام از سبدها (شامل سید دروس زیرگرایش خود یا زیرگرایش دیگر یا دروس کاربردهای علم داده- جداول ۳، ۴، ۵) و یا با موافقت استاد راهنما از سایر گرایش‌های دانشکده یا دانشکده‌های دیگر اخذ نمایند.

○ دانشجویان آموزش‌محور: دانشجویان آموزش‌محور باید حداقل ۴ درس معادل ۱۲ واحد را از دروس زیرگرایش خود اخذ نمایند (جدول ۳ و ۴). ۲ درس باقیمانده را می‌توانند از هر کدام از سبدها (شامل سید دروس زیرگرایش خود یا زیرگرایش دیگر یا دروس کاربردهای علم داده- جداول ۳، ۴، ۵) اخذ نمایند. دانشجو می‌تواند با موافقت استاد راهنما یکی از این ۳ درس باقیمانده را از گرایش‌های دیگر دانشکده یا از دانشکده‌های دیگر اخذ نماید.

• **سمینار کارشناسی ارشد:** ۲ واحد - مطابق جدول ۶

• **پروژه‌ی کارشناسی ارشد:** ۶ واحد - مطابق جدول ۶ (مختص دانشجویان پژوهش‌محور)

۴. **تصویب پروژه‌ی کارشناسی ارشد:** مهلت تصویب تعریف پروژه‌ی کارشناسی ارشد، پایان نیمسال دوم است.

۵. **ارائه‌ی سمینار کارشناسی ارشد:** دانشجوی پژوهش‌محور باید در نیمسال سوم و پس از تصویب پروژه، سمیناری در زمینه‌ی پروژه ارائه دهد که گزارش پیشرفت پروژه محسوب می‌شود. سمینار دانشجویان آموزش‌محور نیز باید در نیمسال سوم و پس از تعیین موضوع و تأیید آن توسط گروه اخذ شود.

۶. **همکاری با دانشکده:** همه‌ی دانشجویان باید به صورت دستیار آموزشی یا با انجام امور ارجاعی دیگر (معادل ۲ واحد) با دانشکده همکاری کنند؛ همکاری با دانشکده برای دانشجویان آموزش‌محور نیز اجباری است.

۷. **استاد راهنما:** مدیر گروه تا قبل از انتخاب و تصویب پروژه، استاد راهنمای دانشجویان است. پس از تصویب تعریف پروژه، استاد راهنمای پروژه‌ی هر دانشجو مسئول تأیید و راهنمایی وی در اخذ واحدهای درسی نیز می‌باشد.

۸. **واحدهای جبرانی:** در صورتی که گذراندن واحدهای جبرانی به تشخیص مدیر گروه برای دانشجو ضروری باشد، دانشجو باید حداکثر تا پایان سال اول تحصیل با تأیید گروه حداکثر چهار درس از دروس جدول ۱ را بگذراند. اخذ ۸ واحد جبرانی یا بیشتر امکان افزایش حداکثر یک نیمسال تحصیلی را به سنوات تحصیلی دانشجو فراهم می‌کند. حداقل نمره‌ی قبولی برای دروس جبرانی ۱۲ است.



دروس جبرانی

جدول ۱. فهرست دروس جبرانی گرایش علم داده

ردیف	شماره	تعداد واحد	نام درس (فارسی)	نام درس (لاتین)
۱.	۴۰۳۵۴	۳	طراحی الگوریتم‌ها	Design of Algorithms
۲.	۴۰۳۸۴	۳	طراحی پایگاه داده‌ها	Database Design
۳.	۴۰۱۸۱	۳	آمار و احتمال مهندسی	Engineering Probability and Statistics
۴.	۴۰۲۸۲	۳	جبر خطی	Linear Algebra
۵.	۴۰۲۴۴	۳	برنامه‌سازی پیشرفته	Advanced Programming

دروس هسته

جدول ۲. فهرست دروس هسته گرایش علم داده

ردیف	شماره	تعداد واحد	نام درس (فارسی)	نام درس (لاتین)
۱.	---	۳	اصول و تکنیک‌ها در علم داده	Principles and Techniques in Data Science
۲.	۴۰۷۱۷	۳	یادگیری ماشین	Machine Learning
۳.	---	۳	زیرساخت‌های پردازشی داده	Data Processing Infrastructures
۴.	---	۱	آداب در علم داده	Ethics in Data Science

دروس زیر گرایش‌ها

جدول ۳. فهرست دروس زیر گرایش مهندسی داده

ردیف	شماره	تعداد واحد	نام درس (فارسی)	نام درس (لاتین)
۱.	---	۳	معماری سیستم‌های کلان‌داده	Architecture of Big-Data Systems
۲.	---	۳	اصول برنامه‌نویسی مقیاس پذیر	Principles of Scalable Programming
۳.	---	۳	امنیت و حریم خصوصی داده	Data Security and Privacy
۴.	۴۰۵۴۸	۳	سیستم‌های پشتیبانی تصمیم‌گیری	Decision Support Systems
۵.	۴۰۶۸۶	۳	الگوریتم‌های داده‌های حجیم	Massive Data Algorithms
۶.	---	۳	سیستم‌های موازی و توزیع شده	Parallel and Distributed Systems
۷.	---	۳	سامانه‌های تحلیل تعاملی داده	Interactive Data Analysis Systems
۸.	---	۳	تحلیل داده‌های مکانی	Spatial Data Analysis
۹.	---	۳	محاسبات ابری و لبه	Cloud and Edge Computing
۱۰.	۴۰۷۶۵	۳	الگوریتم‌های پیشرفته	Advanced Algorithms
۱۱.	۴۰۷۸۵	۳	بهینه‌سازی ترکیبیاتی	Combinatorial Optimization
۱۲.	۴۰۸۳۵	۳	نظریه‌ی الگوریتمی بازی‌ها	Algorithmic Game Theory
۱۳.	---	۳	روش‌های تصادفی و احتمالاتی	Randomized and Probabilistic Methods
۱۴.	۴۰۶۸۳	۳	سامانه‌های پیشرفته ذخیره‌سازی داده	Advanced Data Storage Systems



جدول ۴. فهرست دروس زیرگرایش تحلیل داده

ردیف	شماره	تعداد واحد	نام درس (فارسی)	نام درس (لاتین)
۱.	۴۰۷۱۹	۳	یادگیری ژرف	Deep Learning
۲.	---	۳	تحلیل آماری داده‌ها	Statistical Data Analysis
۳.	---	۳	تحلیل سری‌های زمانی	Time Series Analysis
۴.	۴۰۸۳۷	۳	بهینه‌سازی محدب	Convex Optimization
۵.	۴۰۷۶۸	۳	مدل‌های احتمالاتی گرافی	Probabilistic Graphical Models
۶.	---	۳	آمار در ابعاد بالا	High-Dimensional Statistics
۷.	۴۰۶۷۷	۳	پردازش زبان طبیعی	Natural Language Processing
۸.	---	۳	تحلیل داده‌های چندرسانه‌ای مقیاس بزرگ	Large Scale Multi-Media Data Analysis
۹.	---	۳	یادگیری ماشین مقیاس پذیر	Scalable Machine Learning
۱۰.	۴۰۷۱۸	۳	نظریه یادگیری ماشین	Machine Learning Theory
۱۱.	۴۰۷۲۹	۳	یادگیری ماشین آماری	Statistical Machine Learning
۱۲.	۴۰۶۹۵	۳	فرآیندهای تصادفی	Stochastic Processes
۱۳.	۴۰۶۷۶	۳	نظریه اطلاعات و کدینگ	Information Theory and Coding
۱۴.	---	۳	مبانی علم داده	Foundations of Data Science
۱۵.	---	۳	تحلیل داده‌های متنی	Text Data Analysis

دروس کاربردهای علم داده

جدول ۵. فهرست دروس کاربردهای علم داده

ردیف	شماره	تعداد واحد	نام درس (فارسی)	نام درس (لاتین)
۱	۴۰۶۴۲	۳	شبکه‌های اجتماعی و اقتصادی	Social and Economic Networks
۲	۴۰۵۵۵	۳	تحلیل داده‌های حجیم زیستی	Large-Scale Biological Data Analysis
۳	---	۳	علم داده در مهندسی نرم‌افزار	Data Science in Software Engineering
۴	---	۳	تحلیل داده‌های مالی	Financial Data Analysis
۵	---	۳	تحلیل داده‌های سلامت	Health Data Analysis
۶	---	۳	تحلیل داده‌های کسب‌وکارها	Data Analysis in Business
۷	---	۳	سیستم‌های توصیه‌گر	Recommender Systems

دروس سمینار و پروژه‌ی گرایش علم داده

جدول ۶. فهرست دروس سمینار و پروژه

ردیف	شماره	تعداد واحد	نام درس (فارسی)	نام درس (لاتین)
۱	۴۰۹۰۰	۲	سمینار کارشناسی ارشد	M.Sc. Seminar
۲	۴۰۶۴۰	۶	پایان‌نامه‌ی کارشناسی ارشد	M.Sc. Thesis



سیلابس دروس



طراحی الگوریتم‌ها

Design of Algorithms

اهداف درس

هدف از این درس، آشنایی دانش‌جویان با روش‌های متداول در طراحی الگوریتم‌های کارا برای مسائل مختلف است. در ارائه‌ی مطالب، بر تحلیل کارایی الگوریتم‌ها و اثبات درستی آن‌ها تأکید خواهد شد. همچنین، موضوعات مهمی از نظریه‌ی الگوریتم‌ها همچون پیچیدگی محاسباتی، شبکه‌های شار و الگوریتم‌های گراف در این درس ارائه خواهند شد.

ریز مواد

- مقدمات و مسائل نمونه
 - حل‌پذیری، تحلیل الگوریتم‌ها، زمان‌های اجرا
 - بزرگ‌ترین زیردنباله‌ی متوالی، مسئله‌ی ۳-مجموع
- الگوریتم‌های مبتنی بر استقرا
 - ارزیابی چندجمله‌ای‌ها، نگاشت یک‌به‌یک، ستاره‌ی مشهور
- تقسیم و حل
 - محاسبه‌ی توان، محاسبه‌ی روابط بازگشتی، نزدیک‌ترین زوج نقاط
 - الگوریتم استراسن برای ضرب ماتریس‌ها، تبدیل سریع فوریه
- الگوریتم‌های حریصانه
 - خرد کردن پول، مسائل زمان‌بندی، کوله‌پشتی کسری
 - فشردن‌سازی: کدگذاری هافمن
 - تطابق پایدار، الگوریتم گیل-شاپلی، قضایای مرتبط
- برنامه‌ریزی پویا
 - اعداد فیبوناچی، زمان‌بندی بازه‌های وزن‌دار، خرد کردن پول
 - ضرب زنجیره‌ی ماتریس‌ها، کوله‌پشتی، تراز دنباله‌ها
 - بزرگ‌ترین زیردنباله‌ی مشترک، بزرگ‌ترین زیردنباله‌ی افزایشی
 - محاسبه‌ی مجموعه‌ی مستقل روی درخت، درخت دودویی جست‌وجوی بهینه
- جست‌وجوی فضای حالت
 - روش پس‌گرد، مسئله‌ی هشت وزیر، مجموع زیرمجموعه‌ها
 - انشعاب و حد، فروشنده‌ی دوره‌گرد، درخت بازی، هرس آلفا-بتا
- الگوریتم‌های گراف
 - درخت فراگیر کمینه: الگوریتم‌های کروسکال و پریم
 - هرم فیبوناچی، تحلیل سرشکن برای کاهش کلید
 - کوتاه‌ترین مسیر بین تمام رأس‌ها: الگوریتم‌های فلوید-وارشال و جانسون
- تطابق رشته‌ها
 - روش مبتنی بر اثر انگشت، الگوریتم رابین-کارپ
 - تطابق رشته به وسیله‌ی اتوماتا: الگوریتم کنوت-موریس-پرت
- شبکه‌های شار
 - شار بیشینه و برش کمینه: الگوریتم فورد-فالکرسن
 - بهبود الگوریتم فورد-فالکرسن، بهبودهای ادموندز و کارپ
 - گونه‌ها و کاربردها: تطابق در گراف دوبخشی، مسیرهای مجزا، گرد کردن ماتریس



- برنامه‌ریزی خطی
 - فرم استاندارد، مدل‌سازی مسائل با برنامه‌ریزی خطی
 - الگوریتم سیمپلکس برای حل برنامه‌ریزی خطی
- پیچیدگی محاسبات
 - کاهش چندجمله‌ای، مسائل صدق‌پذیری
 - رده‌ی ان‌پی، اثبات ان‌پی-تمام بودن یک مسئله، قضیه‌ی کوک
 - دور همیلتنی، رنگ‌آمیزی گراف، مجموع زیرمجموعه‌ها
- الگوریتم‌های تقریبی
 - پوشش راسی، فروشنده‌ی دوره‌گرد، سختی تقریب
 - طرح‌های تقریبی چندجمله‌ای، مسئله‌ی کوله‌پشتی

ارزیابی

- سه تمرین نظری (۳ نمره)
- سه تمرین برنامه‌نویسی (۳ نمره)
- آزمون میان‌ترم (۷ نمره)
- آزمون پایانی (۷ نمره)
- یک مسابقه به سبک ای‌سی‌ام (۱+ نمره)

مراجع

- J. Kleinberg and E. Tardos. *Algorithm Design*. Addison Wesley, 2005.
- T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to Algorithms*. 3rd Edition, MIT Press, 2009.
- U. Manber. *Introduction to Algorithms: A Creative Approach*. Addison-Wesley, 1989.
- G. Brassard, P. Bratley. *Algorithmics: Theory and Practice*. Prentice-Hall, 1988.

اساتید

- دکتر ضرابی‌زاده
- دکتر آدام
- دکتر صدیقین



طراحی پایگاه داده‌ها

Database Design

اهداف درس

در این درس دانشجویان با مفاهیم مدل‌سازی معنایی داده‌ها و طراحی پایگاه داده‌ها آشنا می‌شوند و در انتهای ترم انتظار می‌رود که دانشجویان با مفاهیمی که در ریزموارد به آنها اشاره شده است، آشنایی کامل پیدا کرده باشند.

ریزمواد

- مشی پایگاهی و سیستم‌های پایگاهی
 - معرفی درس
 - تعریف پایگاه داده‌ها
 - مشی فابلینگ و مشی پایگاهی
 - عناصر محیط پایگاه داده
 - انواع معماری سیستم پایگاهی (متمرکز، مشتری-خدمت‌گزار، توزیع‌شده)
 - اجزای سیستم مدیریت پایگاه داده رابطه‌ای (OLTP, RDBMS)
- مدل‌سازی معنایی داده‌ها با روش ER
 - موجودیت
 - صفت
 - ارتباط
 - نمودار ER و EER
 - انواع دام‌ها
 - تکنیک‌های تخصیص، تعمیم، تجزیه، ترکیب و تجمیع
 - ویژگی‌های روش مدل‌سازی معنایی
- اصول طراحی پایگاه داده‌ها
 - آشنایی با ساختار جدولی و پایگاه داده جدولی
 - روش طراحی بالا به پایین (تبدیل مدل‌سازی معنایی به طراحی منطقی)
- مقدمات پیاده‌سازی و SQL
 - زبان پایگاه داده‌ی جدولی (SQL)
 - آشنایی مقدماتی با دستورات DDL و DCL
 - آشنایی با دستورات DML
 - SQL توکار در زبان‌های برنامه‌سازی
 - پیاده‌سازی تراکنش
 - پرس‌وجوی پارامتری (Query Parameterized)
- معماری سه‌سطحی پایگاه داده‌ها
 - معماری سه‌سطحی ANSI، دید (نمای) ادراکی
 - دید داخلی و دید خارجی
 - تبدیلات بین سطوح
 - انواع شاخص در سطح داخلی (Hash و Tree-B+, Tree-B)



- عملیات از دید خارجی در پایگاه داده‌ها و مشکلات مربوطه
- استقلال داده‌ای فیزیکی و منطقی
- مفاهیم اساسی مدل داده‌ای رابطه‌ای
 - مولفه‌های مدل داده‌ای در مدل رابطه‌ای
 - رابطه و مفاهیم مربوطه
 - کلید در مدل رابطه‌ای
 - مروری بر اصول طراحی پایگاه داده‌های رابطه‌ای
- جامعیت در مدل رابطه‌ای
 - قواعد جامعیت عام (C1, C2)
 - قواعد جامعیت یا محدودیت‌های کاربری
 - مکانیزم‌های اعمال قواعد جامعیت کاربری شامل اظهار (Assertion) و رهانا (Trigger)
- عملیات در پایگاه رابطه‌ای
 - جبر رابطه‌ای
 - حساب رابطه‌ای
- تئوری وابستگی و ارتقاء سطح نرمال‌سازی رابطه‌ها
 - مفاهیمی از تئوری وابستگی
 - شرح صورت‌های نرمال تا سطح BCNF
 - تجزیه مطلوب
- امنیت پایگاه داده‌ها
 - مدیریت کاربران
 - احراز اصالت
 - کنترل دسترسی
 - رمزنگاری داده‌ها
- سیستم‌های مدیریت پایگاه داده غیررابطه‌ای (DBMS NoSQL)
 - دلایل استفاده از پایگاه داده‌های غیررابطه‌ای (NoSQL)
 - تئوری CAP
 - پایگاه داده‌های کلید-مقدار، ستون-محور، گراف-محور و سند-محور
- مقدمه‌ای بر انبار داده (Warehouse Data)
 - معرفی انبار داده و OLAP

ارزیابی

- آزمون: آزمون‌های میان‌نیم‌سال (۳۰ درصد نمره) و پایان‌نیم‌سال (۳۵ درصد نمره)
- تمرین: ۱۷٪ کل نمره
- پروژه: ۱۳٪ کل نمره
- کوئیز و فعالیت‌های کلاسی ۵٪ کل نمره

مراجع

- سید محمد تقی روحانی رانکوهی، "مفاهیم بنیادی پایگاه داده‌ها"، ویراست چهارم، ۱۳۹۰.
- R. Elmasri, S. Navathe. *Fundamentals of Database Systems*. 7th Edition, Pearson, 2015.
- A. Silberschatz, H. F. Korth, S. Sudarshan. *Database System Concepts*. 6th Edition, McGraw-Hill, 2010.
- C. J. Date. *An Introduction to Database Systems* 8th Edition, Pearson, 2003.



- T. Connolly, C. Begg. *Database Systems*. 6th Edition, Pearson, 2014.
- R. Ramakrishnan, J. Gehrke. *Database Management Systems*. 4th Edition, McGraw-Hill, 2014.

اساتید

- دکتر حیدرنوری
- دکتر امینی
- دکتر فضلی



آمار و احتمال مهندسی

Engineering Probability and Statistics

اهداف درس

هدف از این درس، آشنایی دانش‌جویان با مفاهیم بنیادین تئوری احتمال و استنتاج آماری و کاربردهای آن در مهندسی کامپیوتر، مانند مساله‌ی مدل‌سازی داده‌ها مانند رگرسیون، است. این مفاهیم شامل تفسیر و اصول موضوعه‌ی آمار، توابع توزیع احتمال تک و چندمتغیره، احتمال شرطی و استقلال آماری، متغیرهای تصادفی و متوسط‌گیری، توابع تعریف شده روی متغیرهای تصادفی، خانواده توزیع‌های نمایی، قضیه‌ی حد مرکزی و قانون اعداد بزرگ، و تست فرضیه می‌شود.

ریز مواد

- تفسیرهای احتمال
- اصول موضوعه‌ی احتمال
- عملگرهای روی رخدادها
- استقلال آماری، احتمال شرطی و قاعده بیز
- متغیرهای تصادفی
- امید ریاضی و خواص آن
- ممان‌های مرتبه بالا و تابع مشخصه
- توابع تعریف شده روی یک متغیر تصادفی
- تابع توزیع احتمال مشترک (Distribution Probability Joint)
- تابع چگالی و قانون بیز
- توزیع پیشین مزدوج (Distribution Prior Conjugate)
- خانواده‌ی نمایی (Family Exponential)
- توزیع‌های خاص
- ممان‌های مشترک (Moments Joint)
- توابع تعریف شده روی دو یا چند متغیر تصادفی
- قضیه‌ی حد مرکزی (Theorem Limit Central)
- قانون اعداد بزرگ (Numbers Large of Law)
- همگرایی در احتمال (Probability in Convergence)
- تخمین‌گرهای بیشینه درست‌نمایی (Likelihood Maximum) و بیشینه احتمال پسین (Posteriori a Maximum)
- (Probability)
- تخمین بیز (Estimation Bayes)
- خصوصیات تخمین‌گرها
- تست‌های آماری و تست فرضیه (Tests Hypothesis and Statistical)
- ارزیابی
- تمرین‌های نظری: ۴ نمره
- آزمون میان‌ترم ۱: ۴ نمره
- آزمون میان‌ترم ۲: ۴ نمره



- آزمون پایان ترم: ۶ نمره
- آزمونک‌ها: ۲ نمره

مراجع

- A. Poppulis and S. Pillai. Probability, Random Variables and Stochastic Processes. 4th Edition, McGraw Hill, 2002 (Chapters 1 through 8).
- S. Ross. A First Course in Probability. 10th Edition, Prentice Hall, 2019.
- G. Casella and R. L. Berger. Statistical Inference. 2nd Edition, Wadsworth Press, 2002.

اساتید

- دکتر شریفی
- دکتر جعفری
- دکتر رهبان
- دکتر ربیعی
- دکتر موقر



جبر خطی

Linear Algebra

اهداف درس

هدف از این درس، آشنایی دانش‌جویان با مفاهیم پایه‌ای نظری جبر خطی و همچنین نحوه‌ی استفاده و پیاده‌سازی آن‌ها در بستر نرم‌افزاری مناسب است. آشنایی با مفاهیم این درس امکان تحلیل نگاشت‌ها و سیستم‌های خطی از طریق ماتریس‌ها و اعمال، اپراتورها و مفاهیم تعریف شده مرتبط با آنها را فراهم می‌کند. همین‌طور مساله بهینه‌سازی به عنوان یکی از کاربردهای پراستفاده جبر خطی مورد بررسی قرار می‌گیرد.

ریز مواد

- فضاهای برداری
 - نگاشت خطی و ماتریس
 - فضای برداری نگاشت خطی
 - ساختار جبری نگاشت خطی
 - ماتریس و رنک
 - معکوس نگاشت خطی
 - دوگانگی
 - دستگاه‌های خطی
 - حجم و دترمینان
- چندجمله‌ای‌ها
 - صفرهای چندجمله‌ای
 - فاکتورگیری چندجمله‌ای در میدان‌های مختلط و حقیقی
- مقادیر ویژه و بردارهای ویژه
 - زیرفضاهای ثابت
 - بردارهای ویژه و مقادیر ویژه
 - بردارهای ویژه مستقل خطی
 - فضاهای ویژه و قطری‌سازی ماتریس‌ها
- فضاهای ضرب داخلی
 - ضرب داخلی و تعریف فاصله
 - پایه‌های متعامد
 - عملگرهای فضاهای ضرب داخلی
- عملگرها و تجزیه
 - تجزیه‌ی قطبی
 - تجزیه‌ی مقدارهای منفرد
 - تجزیه‌ی چولسکی
 - تجزیه‌ی ال یو
 - تجزیه‌ی کیو آر
 - عملگرهای الحاقی
 - عملگرهای نرمال
 - عملگرهای یکانی و ایزومتري
 - عملگرهای مثبت



ارزیابی

- تمرین‌ها: ۶ نمره
- دو آزمون میان‌ترم: ۸ نمره
- آزمون پایانی: ۶ نمره
- آزمونک‌ها: ۱ نمره

مراجع

- Right Sheldon Axler. Linear Algebra. Springer, 2015.
- Gilbert Strang. Linear Algebra and Its Application. 4th Edition, Cengage Learning, 2006.
- David Clay. Linear Algebra and Its Application. 4th Edition, Pearson, 2011

اساتید

- دکتر ربیعی
- دکتر رهبان
- دکتر شریفی
- دکتر مطهری
- دکتر جعفری



برنامه‌سازی پیشرفته

Advanced Programming

اهداف درس

در این درس، مفاهیم برنامه‌نویسی شیء‌گرا در قالب زبان برنامه‌نویسی جاوا تدریس می‌شود. همچنین ویژگی‌های ذاتی زبان جاوا، امکانات برنامه‌سازی در جاوا، تفاوت رویکرد جاوا با زبان‌های مشابه در موارد مختلف، برنامه‌سازی همروند و کیفیت نرم‌افزار مورد توجه قرار می‌گیرد. فرض بر این است که دانشجویان در دروس قبلی، با یک زبان برنامه‌سازی آشنا هستند و برخی روش‌های حل مسئله با کمک برنامه‌سازی مانند جستجو، مرتب‌سازی، عملیات ریاضی و نظایر آن را می‌شناسند. تأکید در این درس بر مفاهیم شیء‌گرا است.

ریز مواد

- آشنایی با زبان جاوا
 - تاریخچه
 - ویژگی‌های زبان جاوا
 - اولین برنامه در جاوا
- مفاهیم اولیه برنامه‌سازی در جاوا
 - متغیر، متد، شرط، حلقه
 - داده‌های اولیه (Types Data Primitive)
 - رشته
 - آرایه
- مقدمه‌ای بر طراحی و برنامه‌سازی شیء‌گرا
 - تاریخچه تکامل رویکردها تا رویکرد شیء‌گرایی
 - مفاهیم اولیه شیء‌گرایی
 - طرز تفکر و طراحی شیء‌گرا
 - لفاف‌بندی، واسطه، رده، بسته، دسترسی
- برنامه‌سازی شیء‌گرا در جاوا
 - نحوه تعریف کلاس‌ها
 - اشیاء در حافظه و مدیریت حافظه
 - بارگذاری اولیه و مرگ اشیاء در حافظه
 - زباله‌روب (Collector Garbage)
 - روش‌های ارسال پارامتر در زبان‌های مختلف و جاوا
 - سازنده
 - this
 - اعضای ایستا
 - بسته (package)
 - آشنایی اولیه با نمودار Diagram Class UML
- وراثت
 - مفهوم وراثت
 - super abstract, Protected,



- وراثت چندگانه در سایر زبانها
- چندریختی
 - چندریختی با کمک وراثت
 - مفهوم virtual در زبان ++C
 - اعضای final
- واسط (interface)
 - کاربرد واسط
 - وراثت چندگانه با کمک واسط
- آزمون نرم‌افزار
 - مفهوم تضمین کیفیت نرم افزار، اهمیت، جایگاه و مصادیق آن
 - آزمون واحد
 - نوشتن آزمون واحد با استفاده از JUnit
 - چگونگی تست پرتاب استثناء با استفاده از JUnit
 - مفهوم بدل‌سازی (Mocking) و مقدمه ای بر یک کتابخانه بدل‌سازی در جاوا
- الگوهای طراحی
 - تعریف الگوها در مهندسی نرم‌افزار، جایگاه و اهمیت آنها
 - الگوهای طراحی GoF و دسته‌بندی آنها
 - شرح الگوهای طراحی Singleton, Sate, Strategy, Observer, Fecade
 - شرح الگوی معماری MVC
- بازآرایی کد (Refactoring)
 - چپستی بازآرایی و اهمیت نوشتن کد تمیز
 - نشانه‌های کد بد
 - الگوهای بازآرایی
 - ساختن متدها
 - انتقال ویژگی‌ها بین اشیاء
 - سازماندهی داده‌ها
 - ساده‌سازی عبارات شرطی
 - ساده‌سازی فراخوانی متدها
- مدیریت خطا و استثنا
 - مدل سنتی مدیریت خطا
 - چارچوب مدیریت خطا در جاوا
 - مزایای این مدل
- Exception Runtime Finally,
- داده‌های عام (Generics)
 - متدها و کلاس‌های عام
 - کاربردها
 - تولید و استفاده از کلاس‌های عام
 - کلاس‌های عام و وراثت



- ماهیت کلاس‌های عام و تفاوت آن‌ها با Template در زبان ++C
 - مجموعه‌ها و ظرف‌ها (Containers)
 - ساختمان داده‌های موجود در جاوا
 - Map Set, LinkedList, ArrayList, Collections, Iterator
 - فایل، جویبار و شبکه
 - ورودی و خروجی در فایل
 - Serialization
 - برنامه‌سازی تحت شبکه (socket)
 - برنامه‌سازی همروند
 - نیاز به همروندی
 - همروندی در جاوا
 - طول عمر یک thread
 - آشنایی اولیه با مفاهیم Synchronization و section critical
 - Reflection
 - نیاز به RTTI
 - RTTI در جاوا
 - کاربردها
 - مفاهیم پیشرفته و تکمیلی
 - کلاس‌های داخلی و کلاس‌های بی‌نام
 - حاشیه‌گذاری (Annotation)
 - Enumeration
- ارزیابی**
- آزمون‌ها (میان‌ترم، پایان‌ترم، و آزمونک‌ها): ۵۰٪ نمره
 - تمرین‌های برنامه‌سازی: ۲۵٪ نمره
 - پروژه (سه فاز در طول ترم): ۲۵٪ نمره

مراجع

- P. Deitel, H. Deitel. Java: How to Program. 11th Edition, Pearson Education, 2017.
- B. Eckel. Thinking in Java. 4th Edition, Prentice Hall, 2006.
- M. Fowler, K. Beck, J. Brant, W. Opdyke, D. Roberts. Refactoring: Improving the Design of Existing Code. Addison-Wesley, 1999.

اساتید

- دکتر فضلی
- دکتر ضرابی‌زاده



اصول و تکنیک‌ها در علم داده

Principles and Techniques in Data Science

اهداف درس

هدف این درس، توانمندسازی دانشجویان کارشناسی ارشد و دکترا برای ارائه راه‌حل‌های داده‌محور^۱ برای مسائل مختلف است. برای این منظور، دانشجویان با مفاهیم و پیش‌نیازهای ریاضی و آماری ارائه چنین رویکردهایی آشنا می‌شوند، با اصول و گام‌های ارائه راهکارهای داده‌محور هم‌چون تحلیل و مصورسازی داده‌ها، مدل‌سازی آماری و احتمالاتی، استنتاج آماری و تصمیم‌گیری در شرایط نامعین آشنا می‌شوند و با به‌کارگیری این روش‌ها به صورت عملی در مسائل واقعی، با چالش‌های به‌کارگیری این روش‌ها در دنیای واقعی آشنا می‌شوند. این درس از چهار بخش تشکیل شده است. در بخش اول این درس، دانشجویان با آنالیز داده و چالش‌های عملی آن آشنا می‌شوند. پس از آن، در بخش دوم، با مفاهیم پایه مدل‌سازی آماری و پیش‌نیازهای آن هم‌چون روش‌های بهینه‌سازی آشنا می‌شوند. در بخش سوم، دانشجویان با یک مسئله واقعی در حوزه دسته‌بندی روبرو می‌شوند و گام‌به‌گام مراحل ارائه یک روش مبتنی بر داده را با آن طی می‌کنند و در این مسیر با روش‌های مختلف استخراج و آماده‌سازی داده، یادگیری نمایش داده‌ها در فضای داده، دسته‌بندی داده و روش‌های ارزیابی مدل یادگیری شده آشنا می‌شوند. در بخش آخر، دانشجویان با چالش‌های مستقر کردن یک مدل مبتنی بر داده در محیط عملیات آشنا می‌شوند.

ریز مواد

- آنالیز داده
 - آشنایی با چرخه عمر علوم داده
 - تولید داده (پرسشنامه، سرشماری، آزمایش کنترل شده)
 - جمع‌آوری و تجمیع داده‌ها (یکسان‌سازی داده‌ها، نمایش جدولی داده‌ها، فیلتر کردن و تجمیع داده‌ها)
 - تمیز کردن داده (مدیریت داده‌های پرت، مقادیر نامعلوم داده‌ها، کدگذاری و نمایش در فضای برداری)
 - تحلیل اکتشافی داده
 - مصورسازی داده‌ها
 - تشخیص الگو و فرضیه پردازی به کمک مصورسازی داده‌ها
 - شناخت آفت‌های تحلیل‌های داده (بایاس در داده‌ها، ناکافی بودن ویژگی‌ها، تشخیص همزمانی از علّیت)
 - تست فرضیه و دستکاری p -value^۲
- مدل‌سازی آماری داده‌ها
 - آشنایی با گام‌های مدل‌سازی (تابع هزینه، یادگیری پارامترهای مدل، پیش‌بینی، نظریه تصمیم)
 - مفهوم قدرت تعمیم مدل و سنجش آن به کمک تابع هزینه
 - جداسازی داده‌های آموزش و اعتبارسنجی و تست
 - بیش‌برازش، اعتبارسنجی ضربدری، منظم‌سازی
 - روش‌های بهینه‌سازی (کاهش گرادیان، نیوتن، روش‌های مبتنی بر گشتاور)
 - مدل‌سازی احتمالاتی و بیزی
 - استنتاج آماری، یادگیری مدل به کمک نظریه تخمین، پیش‌بینی به کمک مدل‌های آموزش داده شده
 - نظریه تصمیم
 - مصالحه بین سوگیری و واریانس^۳
 - نفرین ابعاد^۴
- مدل‌سازی آماری در عمل

□ Data-Driven
□ P-Hacking
□ Bias-Variance Tradeoff
□ Curse of Dimensionality



- مصورسازی داده‌های با ابعاد بالا با استفاده از روش tSNE
- استخراج و انتخاب ویژگی
- کوانتیزاسیون ویژگی‌ها به کمک درخت تصمیم
- دسته‌بندی به کمک روش‌های خطی
- دسته‌بندی با استفاده از درخت تصمیم
- ارزیابی دسته‌بندها
- مهندسی یادگیری ماشین در محیط عملیاتی
 - مقدمه‌ای بر **mlops**: یادگیری انتهابه‌انتهاه، یادگیری مداوم، تغییر داده، تغییر مفهوم، فروشگاه ویژگی، خط‌لوله
 - چرخه عمر داده در محیط عملیاتی
 - چرخه عمر و خط‌لوله‌های یادگیری در محیط‌های عملیاتی
 - استقرار سامانه‌های یادگیری در محیط عملیاتی

ارزیابی

- آزمون: آزمون‌های میان‌نیمسال و پایان‌نیمسال (۴۰ درصد نمره)
- تمرین و پروژه: سه تمرین تئوری و یک پروژه عملی که در طول نیمسال تحویل داده می‌شوند (۴۰ درصد نمره).
- گزارش پژوهشی: موضوع پژوهش قبل از آزمون پایان‌نیمسال تعیین می‌شود. دانشجو کار پژوهش را با کمک استاد درس آغاز کرده و پس از انجام کار، نتیجه پژوهش را در قالب گزارش ارائه می‌دهد (۲۰ درصد نمره).

مراجع اصلی

- <https://ds100.org/fa21/syllabus/>
- J. Grus, Data Science from Scratch, O'Reilly, 2019.
- G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Springer, 2017.
- C. O'Neil, R. Schutt, Doing Data Science, O'Reilly, 2013.
- W. McKinney, Python for Data Analysis, O'Reilly, 2012

اساتید

- دکتر حسینی
- دکتر رهبان
- دکتر سلیمانی



یادگیری ماشین

Machine Learning

اهداف درس

در این درس مفاهیم یادگیری ماشین مطرح شده و آشنایی با شاخه‌های مختلف این زمینه صورت گرفته و جنبه‌های مهم عملی و نظری آن معرفی خواهد شد. در شاخه‌های مختلف تکنیک‌ها و الگوریتم‌های مهم بحث می‌شود. در حوزه‌ی یادگیری با ناظر، مسائل رگرسیون و دسته‌بندی مورد بررسی قرار خواهند گرفت و روش‌های حل این مسائل و ارزیابی مدل‌ها معرفی خواهد شد. برای مساله دسته‌بندی انواع دیدگاه‌ها و الگوریتم‌های مربوطه مطرح می‌شود. در بخش یادگیری بدون ناظر در مورد تخمین چگالی، کاهش ابعاد بدون ناظر و خوشه‌بندی صحبت خواهد شد. در نهایت آشنایی مختصری با شاخه‌ی یادگیری تقویتی صورت خواهد گرفت.

ریز مواد

- مقدمه‌ای بر یادگیری ماشین و مرور مباحث احتمال و جبرخطی
- روش‌های تخمین ML و MAP
- رگرسیون
 - رگرسیون خطی و غیرخطی
 - بیش‌برازش (overfitting)
 - تجزیه‌ی خطا به بایاس (bias)، واریانس (variance) و نویز
 - منظم‌سازی (regularization)
 - رگرسیون آماری (statistical): ارتباط توابع هدف مبتنی بر SSE با تخمین‌های احتمالی ML و MAP برای مساله‌ی رگرسیون
- ارزیابی (evaluation) و تنظیم کردن مدل‌ها
 - اعتبارسنجی (validation)
 - اعتبارسنجی متقابل (validation-Cross)
 - انتخاب مدل (selection model)
 - انتخاب ویژگی (selection feature)
 - دسته‌بندی (classification)
- دسته‌بندهای احتمالی (classifiers probabilistic)
 - تئوری تصمیم (theory decision) و دسته‌بند بهینه بیز (classifier optimal Bayes)
 - دسته‌بندی احتمالی جداساز (discriminative) و مولد (generative)
 - regression Logistic دو دسته‌ای و چند دسته‌ای (class-multi) و بیز ساده (Bayes Naïve)
- دسته‌بندی با استفاده از توابع جداسازی (functions discriminant)
 - پرسپترون (Perceptron)
 - جداساز خطی فیشر (Fisher)
 - ماشین بردار پشتیبان (SVM) و هسته (kernel)
 - شبکه‌های عصبی (networks neural)
- درخت تصمیم (Tree Decision)



- آنتروپی و بهره اطلاعاتی (Gain Information)
- الگوریتم ID۳
- توقف رشد و هرس درخت تصمیم
- روش‌های یادگیری مبتنی بر نمونه (based-instance) (دو جلسه)
- تخمین چگالی غیر پارامتری (estimation density parametric-Non)
- دسته‌بند k-نزدیکترین همسایه (Neighbors Nearest-k)
- رگرسیون خطی وزن‌دار محلی (Regression Linear Weighted Locally)
- تئوری یادگیری محاسباتی
 - learning-PAC
 - dimension VC
 - کمینه‌سازی ریسک ساختاری (minimization risk structural)
- یادگیری جمعی (learning ensemble)
 - Bagging و Boosting
 - AdaBoost
- کاهش ابعاد (reduction dimensionality) بدون ناظر
 - تحلیل مولفه اصلی (PCA)
 - تحلیل مولفه مستقل (ICA)
- خوشه‌بندی (clustering)
 - روش‌های افرازی (partitional): EM+GMM و k-means
 - روش‌های سلسله‌مراتبی (hierarchical)
- یادگیری تقویتی (learning reinforcement)
 - فرایند تصمیم مارکوف (MDP)
 - روش‌های یادگیری مبتنی بر مدل (based-model)
 - روش تکرار مقدار (iteration value) و تکرار سیاست (iteration policy)
 - روش‌های یادگیری بی‌مدل (free-model)
 - الگوریتم‌های learning-Q، SARSA، تفاضل زمانی (Difference Temporal)
 - مباحث پیشرفته در یادگیری ماشین

ارزیابی

- تمرین: ۲۰٪
- میان‌ترم: ۲۵٪
- پایان‌ترم: ۳۵٪
- امتحان‌های کوتاه: ۱۰٪
- پروژه: ۱۰٪

مراجع

- C. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.



- T. Mitchell. Machine Learning. MIT Press, 1998.
- K. Murphy. Machine Learning: A Probabilistic Perspective. MIT Press, 2012.
- T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning. 2nd Edition, 2008.

اساتید

- دکتر حسینی
- دکتر رهبان
- دکتر سلیمانی
- دکتر بیگی



زیرساخت‌های پردازشی داده

Data Processing Infrastructures

اهداف درس

هدف این درس، آشنا کردن دانشجویان کارشناسی ارشد و دکترا با مفاهیم و مسائل مطرح در زیرساخت‌های مرتبط با ذخیره سازی و مدیریت داده‌های بزرگ است. تمامی مفاهیم مطرح در این درس به صورت ملموس و عملی تدریس می‌شود و از دانشجویان خواسته می‌شود که زیرساخت‌های تدریس شده را راه‌اندازی کرده و با آن‌ها به صورت عملی کار کنند. به منظور همگرایی بهتر سعی شده است از هر مفهوم زیرساختی، یک فناوری عملی از آن نیز انتخاب شود و در حین آموزش مفاهیم تدریس گردد. به منظور راحتی بیشتر و همچنین یکپارچگی آسان‌تر، همه فناوری‌های از پشته Apache و Hadoop انتخاب شده‌اند.

ریز مواد

- مروری عملی بر سیستم عامل
- مروری بر پایگاه‌های داده به صورت عملی
- ماشین‌های مجازی و فناوری کانتینر
- مفاهیم مرتبط با سیستم عامل در تحلیل داده‌های حجیم
 - مفاهیم مقدماتی
 - معماری Hadoop
 - فایل سیستم توزیع شده و HDFS
 - محاسبه توزیع شده و MapReduce
 - ارسال کار نگاشت کاهش به Yarn
- جریان‌های کاری در Hadoop
 - Hadoop Streaming
 - مثال‌هایی از برنامه‌نویسی MapReduce با پایتون
 - MapReduce پیشرفته
- محاسبات داخل حافظه‌ای و Spark
 - مفاهیم Spark
 - استفاده از PySpark
 - پیاده‌سازی یک برنامه Spark
- انبارهای داده و کاوش آن‌ها
 - انبار داده و شمای داده‌ها
 - کوئری روی داده‌های ساختاریافته با Hive
 - پایگاه‌های داده ستون-محور و تحلیل بلادرنگ داده با HBase
- یکپارچگی داده
 - واردسازی داده‌های رابطه‌ای با استفاده Sqoop
 - واردسازی جریان‌های داده با استفاده از Flume
- تحلیل داده با API‌های لایه بالاتر
 - آشنایی با فناوری Pig
 - آشنایی با API‌های لایه بالاتر Spark مانند Spark SQL و DataFrame
- مقدمه‌ای بر یادگیری ماشین توزیع شده با Spark



ارزیابی

- **آزمون:** آزمون‌های میان‌نیمسال و پایان‌نیمسال (۴۰ درصد نمره)
- **تمرین و پروژه:** سه تمرین تئوری و یک پروژه عملی که در طول نیمسال تحویل داده می‌شوند (۴۰ درصد نمره).
- **گزارش پژوهشی:** موضوع پژوهش قبل از آزمون پایان‌نیمسال تعیین می‌شود. دانشجو کار پژوهش را با کمک استاد درس آغاز کرده و پس از انجام کار، نتیجه پژوهش را در قالب گزارش ارائه می‌دهد (۲۰ درصد نمره).

مراجع اصلی

- Bengfort, Benjamin, and Jenny Kim. Data analytics with Hadoop: an introduction for data scientists. " O'Reilly Media, Inc.", 2016.

اساتید

- دکتر فضلی
- دکتر ضرابی‌زاده
- دکتر بیگی



آداب در علم داده

Ethics in Data Science

اهداف درس

دانش داده دانشی قدرت‌آفرین است. این قدرت به همراه خود اثرات و مسئولیت‌هایی را به همراه می‌آورد. هرچه حجم داده‌ها بیشتر و بیشتر می‌شود نفوذ علوم داده در ابعاد بیشتری از زندگی انسان‌ها رخ می‌دهد و مسائل اخلاقی جدیدی در این عرصه مطرح می‌شوند. این درس سعی خواهد کرد ضمن ارائه یک چارچوب کلی برای تحلیل این مسائل، کاربردهای چنین تحلیلی را در قالب مطالعات موردی به دانشجویان نشان دهد. به همین دلیل شرکت موثر در بحث‌های کلاسی در خصوص هریک از موارد مورد مطالعه جزء لاینفک این درس خواهد بود.

ریزموارد

- معرفی آداب و مسائل اخلاقی در علوم داده
- تفاوت و مرز آداب با حقوق داده‌ها
- آداب خودمختاری و مسئولیت‌پذیری محتوایی سکوها
- مالکیت داده‌ها و حریم خصوصی
- شفافیت داده، رضایت، حق فراموشی
- استفاده هدفمند از داده‌ها (تجاری، سیاسی، پروفایل کردن اشخاص، توصیه گر‌ها و ...)
- روش‌های کشف و کاهش تعصب در داده‌ها و الگوریتم‌ها
- عدالت در یادگیری ماشین (داده و الگوریتم)
- تحلیل تبعیض یا نقض بی طرفی
- آداب در کارکرد موتورهای جست و جو و سیستم‌های توصیه گر
- مصالحه آداب، ریسک و نوآوری
- دستکاری ادراک

ارزیابی

- شرکت موثر در کلاس ۲۰٪
- تمرین‌ها ۲۵٪
- مقاله نهایی ۱۵٪
- آزمون پایانی ۴۰٪

منابع درس

به ازای هریک از جلسات تعدادی مقاله یا یادداشت برای خواندن تعیین می‌شوند. با این حال بخشی از درس از کتب زیر ارائه خواهد شد.

- Global Engineering Ethics (2017), by Heinz Luegenbiehl and Rockwell Clancy, Elsevier Press
- Davis, K. (2012). Ethics of Big Data: Balancing risk and innovation. " O'Reilly Media, Inc."
- Spinello, R. (2021). Cyberethics: Morality and law in cyberspace. Jones & Bartlett
- Learning. Latiff, R. H. (2018). Future war: Preparing for the new global battlefield. Vintage.

اساتید

- دکتر خرازی
- دکتر حبیبی



معماری سیستم‌های کلان داده

Architecture of Big-Data Systems

اهداف درس

هدف این درس، آشنا کردن دانشجویان کارشناسی ارشد و دکترا با مفاهیم و مسائل مطرح در طراحی و ساخت سامانه‌های بزرگ مبتنی بر داده هستند. عموم این سامانه‌ها با حجم بزرگی از داده‌ها سر و کار دارند که انواع مختلفی دارند. در این درس نحوه حل مسائل و چالش‌های مختلفی که در این سامانه‌ها اتفاق می‌افتند آموزش داده می‌شوند.

ریز مواد

- سیستم‌های قابل اطمینان، مقیاس پذیر و قابل نگهداری
- تفکر در مورد سیستم‌های داده
- قابلیت اطمینان (اشکالات سخت افزاری، خطاهای نرم افزاری، خطاهای انسانی)
- مقیاس پذیری (شرح بار، شرح عملکرد، رویکردهای مقابله با بار زیاد)
- قابلیت نگهداری (قابلیت استفاده: آسان کردن زندگی برای عملیات، سادگی: مدیریت پیچیدگی، تکامل پذیری: سهولت در تغییر)
- مدل‌های داده و زبانهای پرس و جو
 - مدل رابطه‌ای در مقابل مدل سند (تولد NoSQL، ناسازگاری رابطه و اشبا، روابط چند به یک و چند به چند، پایگاه‌های داده مبتنی بر سند و مقایسه آن‌ها در با پایگاه‌های داده رابطه‌ای)
 - پرس و جو برای داده‌ها (پرسش‌های اعلامی^۶ در وب، پرسش مبتنی بر نگاشت-کاهش^۷)
 - مدل‌های داده‌ای مبتنی بر گراف (گراف‌های ویژگی، زبان پرس و جو Cypher، پرس و جوهای گرافی در SQL، ذخیره‌های سه گانه و SPARQL، Datalog)
- ذخیره سازی و بازیابی
 - ساختارهای داده در پایگاه‌های داده (نمایه‌سازی هش، SSTables و درختان LSM، درختان B، سایر ساختارهای نمایه‌سازی، ذخیره همه چیز در حافظه)
 - پردازش تراکنش یا تجزیه و تحلیل^۸ (ذخیره سازی داده‌ها، ستاره‌ها و دانه‌های برف)
 - ذخیره سازی ستون محور (فشردن ستون، مرتب سازی بر اساس ذخیره ستون، نوشتن در محل ذخیره ستون محور، تجمیع: مکعب داده‌ها و نمای مادی شده^۹)
- کدگذاری و تکامل
 - فرمت‌های کدگذاری داده‌ها (فرمت‌های خاص زبان، JSON، XML و انواع باینری، Thrift و Protocol Buffer، Avro، مزایای طرحواره^{۱۰})
 - حالت‌های جریان داده (جریان داده‌ها از طریق پایگاه‌های داده، REST و RPC، جریان انتقال پیام، داده‌های توزیع شده)
- تکثیر^{۱۱}
 - رهبران و پیروان

⁶ Declarative

⁷ Map-Reduce

⁸ Analytics

⁹ Materialized View

¹⁰ Schemes

¹¹ Replication



- مشکلات مربوط به تأخیر تکثیر
- تکثیر با چند رهبر
- تکثیر بدون رهبر
- تقسیم‌بندی^{۱۲}
 - تقسیم‌بندی و تکثیر
 - تقسیم‌بندی داده‌های کلید-مقدار
 - تقسیم‌بندی و نمایه‌سازی‌های ثانویه
 - متعادل‌سازی مجدد تقسیم‌بندی‌ها
 - مسیریابی درخواست‌ها
- تراکنش‌ها
 - مفهوم تراکنش (ACID، عملیات تک‌شی‌ای و چندشی‌ای)
 - سطوح ضعیف ایزوله‌سازی
 - قابلیت سریال‌سازی (اجرای سریال واقعی، 2PL، SSI)
- مشکلات با سامانه‌های توزیع‌شده
 - شکست‌ها^{۱۳}
 - شبکه‌های غیرقابل اعتماد
 - ساعت‌های غیرقابل اعتماد
- ثبات و اجماع
 - ضمانت‌های ثبات
 - خطی‌سازی سیستم‌ها
 - ضمانت سفارش
 - تراکنش‌های توزیع‌شده و اجماع (2PC، اجماع مقاوم در برابر شکست و ...)
- پردازش دسته‌ای
 - پردازش دسته‌ای با استفاده از ابزارهای Unix
 - نگاشت-کاهش و فایل سیستم‌های توزیع‌شده
 - فراتر از نگاشت-کاهش (گراف و پردازش دنباله‌ای، API های سطح بالا، مادی‌سازی حالت‌ها)
- پردازش جریانی
 - ارسال جریان اتفاقات
 - جریان و پایگاه‌های داده
 - پردازش جریان‌های داده
- سایر مباحث مهم
 - آشنایی CI/CD و DevOps
 - آشنایی مجازی‌سازی و فناوری Container
 - آشنایی فناوری‌های Orchestration
 - آشنایی فناوری‌های Monitoring

¹² Partitioning

¹³ Failures



- آشنایی فناوری‌های Cache
- آشنایی فناوری‌های Load Balancing

ارزیابی

- **آزمون:** آزمون‌های میان‌نیمسال و پایان‌نیمسال (۴۰ درصد نمره)
- **تمرین و پروژه:** سه تمرین تئوری و یک پروژه عملی که در طول نیمسال تحویل داده می‌شوند (۴۰ درصد نمره).
- **گزارش پژوهشی:** موضوع پژوهش قبل از آزمون پایان‌نیمسال تعیین می‌شود. دانشجو کار پژوهش را با کمک استاد درس آغاز کرده و پس از انجام کار، نتیجه پژوهش را در قالب گزارش ارائه می‌دهد (۲۰ درصد نمره).

مراجع اصلی

- Kleppmann, Martin. Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. " O'Reilly Media, Inc.", 2017.

اساتید

- دکتر فضلی
- دکتر حبیبی
- دکتر ضرابی‌زاده



اصول برنامه‌نویسی مقیاس‌پذیر

Principles of Scalable Programming

اهداف درس

هدف این درس، آشنا کردن دانشجویان کارشناسی ارشد و دکترا با الگوها و پارادایم‌های برنامه‌نویسی مقیاس‌پذیر و با کارایی بالا برای ساخت سیستم‌های مقیاس بالا است. این الگوها باید به صورت عملی در یک یا چند زبان برنامه‌نویسی به دانشجویان تدریس شود که زبان پیشنهادی اسکالا است.

ریز مواد

- بحثی بر مقیاس‌پذیری در زبان‌های برنامه‌سازی
 - جمع‌آوری زباله
 - دسترسی مستقیم به حافظه و محاسبات اشاره‌گری^{۱۴}
 - بررسی نوع متغیر ایستا^{۱۵}
 - رسیدگی به استثنائات^{۱۶}
 - بررسی اشکالات در زمان اجرا^{۱۷}
 - قراردادهای و ادعاها^{۱۸}
 - پشتیبانی از انواع تجزیده‌ها (ماژول، شی‌گرایی، برنامه‌نویسی تابعی و)
 - و
- برنامه‌نویسی تابعی
 - مقدمه‌ای بر برنامه‌نویسی تابعی
 - طراحی تابعی برنامه و کتابخانه‌های ترکیبیاتی
 - ساختارهای معمول در طراحی برنامه‌نویسی تابعی
 - I/O در برنامه‌نویسی تابعی
- برنامه‌نویسی همزمان
 - مدل حافظه در برنامه‌نویسی همزمان
 - سنگ بناهای همزمانی
 - برنامه‌نویسی آسنکرون
 - جمع‌آوری دیتا موازی
- برنامه‌نویسی واکنشی
 - مقدمه‌ای بر برنامه‌نویسی واکنشی
 - برنامه‌نویسی توزیع‌شده
 - مدل اکتور و سکوی akka
 - پردازش جریان داده

ارزیابی

¹⁴ Pointer Arithmetic
¹⁵ Static Type Checking
¹⁶ Exception Handling
¹⁷ Runtime Error Checking
¹⁸ Assertion and Contracts
آخرین ویرایش: ۱۴۰۰/۱۰/۰۲



- **آزمون:** آزمون‌های میان‌نیمسال و پایان‌نیمسال (۴۰ درصد نمره)
- **تمرین و پروژه:** سه تمرین تئوری و یک پروژه عملی که در طول نیمسال تحویل داده می‌شوند (۴۰ درصد نمره).
- **گزارش پژوهشی:** موضوع پژوهش قبل از آزمون پایان‌نیمسال تعیین می‌شود. دانشجو کار پژوهش را با کمک استاد درس آغاز کرده و پس از انجام کار، نتیجه پژوهش را در قالب گزارش ارائه می‌دهد (۲۰ درصد نمره).

مراجع اصلی

- Vanier, Mike. "Scalable computer programming languages." (2008). Available on http://users.cms.caltech.edu/~mvanier/hacking/rants/scalable_computer_programming_languages.html
- Odersky, Martin, Lex Spoon, and Bill Venners. "Programming in Scala: Updated for Scala 2.12.", Third Edition, Artima Incorporation, USA, (2016).
- Prokopec, Aleksandar. Learning Concurrent Programming in Scala. Packt Publishing Ltd, 2017.
- Chiusano, Paul, and Runar Bjarnason. Functional programming in Scala. Simon and Schuster, 2014.

اساتید

- دکتر ایزدی
- دکتر فضلی



امنیت و حریم خصوصی داده

Data Security and Privacy

اهداف درس

این درس قصد دارد تا ابعاد مختلف ملاحظات مرتبط با امنیت و حریم خصوصی را در خصوص داده‌ها و بخصوص آنچه امروز تحت عنوان داده‌های کلان مطرح می‌شود به بحث گذاشته و روش‌های علمی در جهت تحلیل این ابعاد و یا کاهش اثرات سوء آنها به دانشجویان ارائه کند. بدین ترتیب بخش اعظم درس مرتبط با تعریف حریم خصوصی در عصر کلان داده‌هاست. بخش دیگری از درس مرتبط با مباحث امنیتی معماری‌های مختلفی است که از داده استفاده می‌کنند. این معماری‌ها شامل وب، معماری ابری، پردازش در لبه و سیستم‌های رایفیزیکی^{۱۹} می‌شود. علاوه بر این، با توجه به ارتباطات ماشین به ماشین در اینترنت اشیا و انواع حملات مرتبط با داده در این فضا بخشی از درس به این موضوع خواهد پرداخت. در نهایت به موضوع سیاست‌های امنیتی و قوانین برخاسته از آنها در فضای داده پرداخته می‌شود.

ریزموارد

- اثرات امنیتی اگزوز^{۲۰}های دیجیتال
- امنیت و حریم خصوصی در شبکه‌های اجتماعی
- انواع تعاریف و روش‌های اندازه‌گیری حریم خصوصی
 - حریم خصوصی تفاضلی
 - حریم خصوصی رفتاری
 - جامعیت زمینه‌ای^{۲۱}
 - رویکردهای فلسفی و اخلاقی به حریم خصوصی
 - چالش‌های تعریف استاندارد حریم خصوصی
- کشف نقض حریم خصوصی و تحدید اهداف^{۲۲} تحلیل داده از طریق وارسی^{۲۳}
- گمنام سازی و بازشناسایی داده
- شخصی‌سازی و انگشتنگاری با داده‌ها
- یادگیری ماشین حافظ حریم خصوصی
- انواع روش‌های انتقال/تبادل داده حافظ حریم خصوصی (مبتنی بر رمزنگاری)
- کشف و تحلیل استفاده از داده در وب
- دلالتی^{۲۴} داده و تبعات امنیت و حریم خصوصی
- اثرات امنیتی معماری‌های مبتنی بر ابر و پردازش در لبه
- امنیت و حریم خصوصی در اینترنت اشیا
- حملات داده در سیستم‌های رایفیزیکی
- ردگیری بین افزاره‌ای^{۲۵}
- سیاست‌های مرتبط با امنیت /حریم خصوصی و محدودیت‌های آن

¹⁹ Cyber-physical

²⁰ exhaust

²¹ Contextual integrity

²² Purpose restriction

²³ Audit

²⁴ Data broker

²⁵ Cross device tracking

آخرین ویرایش: ۱۴۰۰/۱۰/۰۲



- مقررات مرتبط با حریم خصوصی الزامات شخصی و سازمانی
- اثرات دو طرفه فناوری و قوانین مرتبط

ارزیابی درس

- فعالیت کلاسی (۱۰٪)
- گزارش پروژه گروهی (۲۰٪)
- تمرین (سه تمرین عملی) (۳۰٪)
- امتحان پایانی (۴۰٪)

مراجع

تعداد زیادی از یادداشت‌ها و مقالات موجود در خصوص هریک از موارد نامبرده شده در این درس مورد استفاده خواهند بود. بخش قابل توجهی از مفاهیم درس از طریق بحث در خصوص این منابع هستند بنابراین خواندن هریک از مقالات مورد اشاره در هر مورد برای دانشجویان اجباری بوده و خواندن منابع بعد از کلاس به تنهایی کفایت نخواهد کرد.

برخی از منابع استفاده شده در این درس عبارت اند از:

- Eric Horvitz, Deirdre Mulligan (2015) Data, privacy, and the greater good, Science.
- 1990 U.S. Census:
- Latanya Sweeney (2000) Simple Demographics Often Identify People Uniquely, CMU Data Privacy Working Paper 3.
- Yaniv Erlich, Tal Shor, Itsik Pe'er, Shai Carmi (2018) Identity inference of genomic data using long-range familial searches, Science.
- Cynthia Dwork & Aaron Roth (2014) Chapter 1: The Promise of Differential Privacy, in The Algorithmic Foundations of Differential Privacy, NOW Publishers.
- Zhanglong Ji, Zachary C. Lipton, Charles Elkan (2014) Differential Privacy and Machine Learning: a Survey and Review, arXiv.
- Logistic regression: Kamalika Chaudhuri, Claire Monteleoni (2009) Privacy-preserving logistic regression, NeurIPS.
- CBC Marketplace (Nov 24, 2017) How companies use personal data to charge different people different prices for the same product.
- Steven Englehardt, Arvind Narayanan (2016) Online Tracking: A 1-million-site Measurement and Analysis, SIGSAC.
- Elena Zheleva, Lise Getoor (2009) To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles, WWW.
- Michal Kosinski, David Stillwell, Thore Graepel (2013) Private traits and attributes are predictable from digital records of human behavior, PNAS.
- Bryce Goodman & Seth Flaxman (2016) European Union regulations on algorithmic decision-making and a “right to explanation”, 2016 ICML Workshop on Human Interpretability in Machine Learning.

اساتید

- دکتر خرازی
- دکتر جلیلی
- دکتر امینی



سیستم‌های پشتیبانی تصمیم‌گیری

Decision Support Systems

اهداف درس

سیستم‌های پشتیبانی تصمیم‌گیری واحد درسی است که طی آن فرایندهای تصمیم‌گیری و استفاده از ابزارهای کامپیوتری برای حل بهتر مسائل و غلبه بر پیچیدگی‌های موجود در آن‌ها معرفی می‌شود. در این درس دانشجویان با سیستم‌های پشتیبانی مدیریت (MSS)، پشتیبانی تصمیم (DSS) و مکانیزم‌های مرتبط با هوشمندی تجاری (BI) و توسعه و ساخت این سیستم‌ها آشنا می‌شوند. در این درس چهارچوبی تدوین شده که دانشجویان در آن با ساختار سیستم‌های اطلاعاتی و ابزارها و تکنیک‌هایی نظیر انبار داده‌ها در فرایندهای سیستم پشتیبانی تصمیم آشنا شده و در طی آن مهارت‌های لازم برای توسعه یک سیستم پشتیبانی تصمیم را فرا می‌گیرند

ریز مواد

- مفاهیم پایه فرایند تصمیم‌گیری
- نیاز به سیستم‌های پشتیبان تصمیم
- مفاهیم مطرح در هوشمندی تجاری (Intelligence Business) و رابطه آن با سیستم‌های پشتیبان تصمیم
- مدیران و تصمیم‌گیری، دورنمایی از سیستم‌های پشتیبانی مدیریت (MSS) شامل سیستم پشتیبانی تصمیم (DSS)
- سیستم پشتیبانی تصمیم گروهی (DSS-G)
- سیستم پشتیبانی اجرایی (EIS)
- سیستم خبره (ES)
- سیستم مدیریت دانش (KMS)
- مفهوم سیستم، تصمیم‌گیری، و فازهای آن
- دورنمایی از سیستم‌های پشتیبانی تصمیم (DSS) شامل زیر سیستم داده، زیرسیستم مدل و زیرسیستم واسط کاربر
- مدیریت داده‌ها شامل مخزن داده‌ها، بازیابی، پایگاه داده‌ها در DSS، پردازش تحلیلی برخط (OLAP)، داده کاوی و نمایش
- انبار داده‌ها و مفاهیم مرتبط با آن
- مدل‌سازی و تحلیل شامل انواع مدل‌های استاتیک و پویا، مدل‌سازی به وسیله صفحه گسترده، درخت تصمیم و جدول تصمیم، مدل‌های بهینه، و مدل‌های ابتکاری (Heuristic)، شبیه‌سازی، مدل‌های چند بعدی (OLAP)، مدل‌سازی و شبیه‌سازی بصری
- توسعه DSS شامل چرخه سنتی عمر نرم‌افزار، و متدولوژی‌های مختلف برای توسعه DSS
- سیستم‌های پشتیبانی تصمیم‌گیری گروهی و تکنولوژی‌های پشتیبانی گروهی
- سیستم‌های پشتیبانی اجرایی EIS شامل سیستم‌های اطلاعاتی اجرایی نقش مدیران و اطلاعات، مشخصه‌های EIS، مقایسه EIS و DSS، زنجیره تامین و پشتیبانی تصمیم)
- سیستم‌های خبره و هوش مصنوعی شامل مفهوم هوش مصنوعی و اجزای آن، مفهوم سیستم‌های خبره، ساختار سیستم‌های خبره، زمینه‌های سیستم‌های خبره، و انواع سیستم‌های خبره
- پشتیبانی تصمیم‌گیری شبکه‌ای مثل اینترنت، اینترانت و اکسترانت
- پیاده‌سازی DSS و استراتژی‌های آن
- DSS‌های هوشمند و آثار اجتماعی و سازمانی سیستم‌های پشتیبانی مدیریت (MSS)
- انبارهای داده و داده کاوی
- مدیریت ارتباطی با مشتری (CRM) و برنامه‌ریزی منابع سازمان (ERP)



ارزیابی

- تمرین‌های پژوهشی: (۱۰ درصد نمره)
 - تمرین‌هایی که باید با استفاده از اینترنت و متن کتاب تجزیه و تحلیل شده و پاسخ داده شوند.
 - تمرین‌هایی که در ارتباط با استفاده از مفاهیم و ابزارهای موجود در دنیای واقعی است که تحت عنوان مطالعه موردی به دانشجویان داده می‌شود.
- تمرین‌های عملی: در این بخش دانشجویان موظف به اجرا و پیاده سازی پروژه‌های مرتبط با این درس و استفاده از ابزارهای متنوع متن-باز موجود در این زمینه هستند (۱۰ درصد نمره).
- سمینار و ارائه کارهای پژوهشی: در این درس دانشجویان به صورت گروهی و یا انفرادی باید کارهای پژوهشی خود را ارائه دهند و از سمینارهای مذکور ارزشیابی به عمل می‌آید (۳۰ درصد نمره).
- آزمون‌های میانی و پایانی: این درس شامل یک آزمون میانی است که در اواسط ترم برگزار شده و شامل مطالب ۶ فصل ابتدایی است. آزمون پایانی شامل کلیه مطالب تدریس شده است (۵۰ درصد نمره).

مراجع

- E. Turban, R. Sharda, and D. Delen. Decision Support Systems and Business Intelligent Systems. 9th edition, Prentice-Hall, 2010.
- I. Linden, S. Liu, and C. Colot. ICDSST Decision Support Systems VII. 1st edition, Springer-Verlag, 2017.
- G.M. Marakas. Decision Support in the 21st Century. 2nd edition, Prentice-Hall, 2003.
- C. Todman. Designing a Data Warehouse: Supporting Customer Relationship Management. Prentice-Hall Professional Technical Reference, 2000.
- V. Kumar and W. Reinartz. Customer Relationship Management. 3rd edition, Springer-Verlag, 2018.
- S.A. Brown. Customer Relationship Management: A Strategic Imperative in the World of e-Business. Wiley, 1999.
- E. Berner. Clinical Decision Support Systems: Theory and Practice 3rd edition, Springer-Verlag, 2016.

اساتید

- دکتر حبیبی



الگوریتم‌های داده‌های حجیم

Massive Data Algorithms

اهداف درس

با افزایش چشمگیر حجم داده‌های ارزشمند و محدودیت پردازشی و ظرفیتی حافظه‌های سریع، نیاز به طراحی داده‌ساختارها و الگوریتم‌های ویژه برای داده‌های حجیم روزبه‌روز محسوس‌تر می‌شود. در این درس، مدل‌های رایج برای داده‌های حجیم مطرح و برای برخی مسائل پایه‌ای، الگوریتم‌های بهینه ارائه می‌شود.

ریز مواد

• Models aware-Cache

- معرفی مدل
- الگوریتم‌های مرتب‌سازی و کران پایین آن
- داده‌ساختارهای جست‌وجو: tree-B Persistent، tree-B Weighted، tree-B، tree Buffer و tree-KD
- داده‌ساختارهای هندسی: tree search Priority، tree Interval، tree Range و tree-KD
- یک نمونه الگوریتم جاروب
- الگوریتم‌های گراف: MST و ranking List، trees on Algorithms، BFS، DFS

• Models oblivious-Cache

- معرفی مدل
- مرتب‌سازی
- داده‌ساختارهای جست‌وجو
- ضرب ماتریس‌ها

• Models Streaming

- معرفی مدل
- پیدا کردن اعداد با تکرار زیاد
- پیدا کردن تعداد اعداد متمایز
- معرفی تکنیک sketching
- بررسی مسائل پایه‌ای گراف
- بررسی مسائل هندسی و معرفی sets-core
- خوشه‌بندی
- نظریه ارتباطات و حد پایین

• Algorithms Sublinear

- معرفی مدل
- محاسبه میانگین درجه‌ی گراف
- اشتراک دو گوز محدب

ارزیابی

- آزمون: آزمون‌های میان ترم و پایان ترم (۱۴ نمره)
- ارائه و گزارش پژوهشی (۳ نمره).
- تمرین تئوری (۳ نمره)



مراجع

- L. Arge. External memory geometric data structures. Lecture notes, 2005.
- A. Chakrabarti and D. College. Data streams algorithms. Lecture notes, 2011.
- A. Czumaj and C. Sohler. Sublinear-time algorithms. Lecture notes.
- Erik Demaine. Cache oblivious algorithms and data structures. Lecture notes.
- U. Meyer, P. Sanders, and J. Sibeyn. Algorithms for memory hierarchies. Lecture notes, 2003.
- Norbert Zeh. I/O efficient graph algorithms. Lecture notes

اساتید

- دکتر آبام



سیستم‌های موازی و توزیع شده

Parallel and Distributed Algorithms

اهداف درس

این دوره مفاهیم مقدماتی کلی در طراحی و پیاده سازی سیستم‌های موازی و توزیع شده را پوشش می دهد که تمامی شاخه های اصلی مانند محاسبات ابری، محاسبات شبکه ای، محاسبات خوشه ای، ابررایانه ها و محاسبات چند هسته ای را پوشش می دهد.

ریز مواد

- مدل‌های محاسبات توزیعی
- محاسبات با کارایی بالا
- محاسبات شبکه‌ای
- محاسبات ابری
- محاسبات چند-هسته‌ای
- محاسبات چند-کاری
- سیستم‌ها و مدل‌های برنامه‌سازی
- فرآیندها و ریسها
- نگاشت-کاهش
- سیستم‌های جریان کاری
- مجازی‌سازی
- فایل سیستم‌ها و ذخیره‌سازهای توزیع شده
- محاسبات داده-محور
- جدول‌های هش توزیع شده
- مدل‌های ثبات و توافق
- تحمل پذیری خطا
- تحلیل کارایی
- معماری‌های موازی
- برنامه‌سازی چندریسه‌ای
- معماری GPU و برنامه‌سازی
- واسط ارسال پیغام

مراجع

- Hwang, Kai, Jack Dongarra, and Geoffrey C. Fox. Distributed and cloud computing: from parallel processing to the internet of things. Morgan kaufmann, 2013.
- Andrew S. Tanenbaum and Maarten van Steen. 2006. Distributed Systems: Principles and Paradigms (2nd Edition). Prentice-Hall, Inc., USA



اساتید

- دکتر سربازی آزاد
- دکتر قدسی



سامانه‌های تحلیل تعاملی داده

Interactive Data Analysis Systems

اهداف درس

هدف این درس آشناسدن دانشجویان با سامانه‌های تحلیل تعاملی داده است. این سامانه‌ها داده‌های مورد نیاز کاربران را جمع‌آوری و تجمیع کرده و اطلاعات و دانش مورد نیاز کاربران را با تحلیل این داده‌ها و ارائه نمایش‌های مختلف، عرضه می‌کنند. برای تولید این سامانه‌ها بایستی نکاتی در نظر گرفته شود که این درس به بیان این نکات می‌پردازد.

ریز مواد

- مقدمه
 - استخراج نیازمندی‌ها
 - مشخص‌سازی داده‌های مورد نیاز برای پاسخگویی به نیازمندی‌ها
 - تبدیل داده‌ها به پاسخ‌ها
- جمع‌آوری داده‌ها
 - انواع مختلف ورودی‌ها
 - API‌های وب و OAUTH
 - قالب‌های مشترک داده‌ها: XML, JSON, CSV و ...
 - نمونه‌برداری و جانبداری در جمع‌آوری داده
- تمیزسازی داده
 - فهم داده
 - کیفیت داده: کامل بودن، قابلیت اتکا، مطابقت، صحت
 - مشکلات عمده با داده‌ها
- ابزارها برای تحلیل داده
 - EDA
 - تبدیل‌ها و ویژگی‌های آن‌ها
 - یادگیری ماشین قابل استفاده
- چارچوب نمایش
 - چگونه نباید نمایش داد؟ چه چیزهایی را نباید نمایش داد؟ چرا نباید بعضی چیزها را نمایش داد؟
 - مسائل ادراکی در نمایش
 - چه چیزی یک نمایش را خوب می‌کند؟
 - نمایش داده‌های بزرگ
 - فرآیند طراحی نمایش
- توسعه نمایش طراحی شده
 - نمایش داده
 - تعامل
 - نشان‌گذاری
 - ترکیب بندی

مراجع



- L. Arge. External memory geometric data structures. Lecture notes, 2005.
- Kirk, Andy. Data visualisation: A handbook for data driven design. Sage, 2016.
- Knaflic, Cole Nussbaumer. Storytelling with data: A data visualization guide for business professionals. John Wiley & Sons, 2015.

اساتید

- دکتر حیدرنوری
- دکتر حبیبی



تحلیل داده‌های مکانی

Spatial Data Analysis

اهداف درس

هدف از این درس، آشنایی دانش‌جویان با داده‌ساختارها و الگوریتم‌های کارا برای ذخیره‌سازی، پردازش و تحلیل داده‌های مکانی است. موضوعات ارائه‌شده در این درس در کاربردهای مختلف از جمله سامانه‌های مکان‌محور، سامانه‌های ناوبری و سامانه‌های اطلاعات جغرافیایی مورد استفاده قرار می‌گیرند.

ریز مواد

- داده‌های مکانی
 - نقاط، سطوح، چندضلعی‌ها
 - اندازه‌گیری، دقت، خطا
- ذخیره‌سازی و بازیابی
 - ساختمان داده‌های مکانی
 - جستجوی بازه‌ای
 - مکان‌یابی و پنجره‌بندی
- پردازش داده‌های مکانی
 - اندازه‌گیری فاصله، محیط و مساحت
 - محاسبه‌ی پوسته‌ی محدب
 - یافتن نزدیک‌ترین نقاط
- تحلیل داده‌های مکانی
 - یافتن مرکز، میانه، عمق توکی
 - خوشه‌بندی
 - برازش اشکال
 - روش‌های برنامه‌ریزی خطی
- نقشه‌های جغرافیایی
 - ذخیره‌سازی نقشه
 - عملیات پایه روی نقشه
 - ترکیب لایه‌های مختلف نقشه
 - تحلیل داده‌های نقشه
- نقشه‌های سه‌بعدی و رویه‌ها
 - شبکه‌های نامنظم مثلث‌بندی‌شده
 - نمونه‌برداری، درون‌یابی، تخمین

ارزیابی

- آزمون (میان‌ترم و پایان‌ترم): ۵۵ درصد نمره
- تمرین: ۲۵ درصد نمره
- پروژه: ۲۰ درصد نمره

مراجع



- R. P. Haining and R. Haining. Spatial Data Analysis: Theory and Practice. Cambridge University Press, 2003.
- C. Lloyd. Spatial Data Analysis: An Introduction for GIS users. Oxford University Press, 2010.
- M. de Berg, O. Cheong, M. van Kreveld, and M. Overmars. Computational Geometry: Algorithms and Applications. 3rd edition, Springer-Verlag, 2008.
- C. McGrew, A. Lembo, and C. Monroe. An introduction to statistical problem solving in geography. 3rd edition, Waveland Press, 2014.

اساتید

- دکتر ضرابی‌زاده
- دکتر قدسی



محاسبات ابری و لبه

Cloud and Edge Computing

اهداف درس

محاسبات (رایانش) ابری، یک فناوری مبتنی بر بستر اینترنت است، که به منظور ارائه‌ی خدمات رایانه‌ای فراگیر، از جمله سرورهای ذخیره‌سازی، شبکه، نرم‌افزار، تجزیه و تحلیل، پایگاه داده و ارتباطات مورد استفاده قرار می‌گیرد. در این فناوری، کاربران می‌توانند در ازای پرداخت هزینه، با استفاده از منابع سریع، منعطف و دسترس‌پذیر، هزینه‌های عملیاتی کسب‌وکار خود را کاهش داده و بهره‌وری تجارت خود را افزایش دهند. از جمله خصوصیات بارز زیرساخت‌های ابری، استفاده از سامانه‌های توزیع‌شده‌ی جغرافیایی است، که ممکن است شامل ده‌ها مرکز داده و میلیون‌ها رایانه باشد. از جمله مهم‌ترین دلایل استفاده از محاسبات ابری در دنیا می‌توان به هزینه‌ی کم، سرعت بالا، توسعه‌پذیری، بهره‌وری، کارایی، قابلیت اطمینان و امنیت آن اشاره نمود. در این درس، دانشجویان علاوه بر آشنایی با ساختار محاسبات ابری، بینش کافی در مورد اصول اساسی نحوه‌ی عملکرد این سامانه‌ها را به دست خواهند آورد. از طرف دیگر، با رشد فزاینده‌ی اینترنت اشیاء و فناوری‌های مبتنی بر نسل پنجم ارتباطات، حجم داده‌های تولید شده روز به روز در حال افزایش است، که در بسیاری کاربردها، پردازش داده‌ها به زمان پاسخ کمی نیازمند هستند. محاسبات لبه با نزدیک کردن عملیات ذخیره‌سازی و پردازشی به کاربران، سعی در رفع یکی از مهم‌ترین چالش‌های زیرساخت‌های ابری، یعنی زمان پاسخ‌گویی، برآمده است. در همین راستا، از جمله سرفصل‌های این درس، معرفی و بررسی ساختار محاسبات لبه و ارتباط آن با ابر خواهد بود.

ریز مواد

- معرفی محاسبات ابری و پردازش لبه، تاریخچه، آمار و ارقام، ساختار، چالش‌ها، کاربردها
- روش‌های خدمت‌رسانی در ابر
- معرفی مدل Azure IoT، قواعد حاکم بر Azure IoT Edge، تحلیل داده‌های حجیم^{۲۶}
- مروری بر سخت‌افزارهای استفاده شده در زیرساخت‌های ابری
- توسعه‌پذیری و پارتیشن‌بندی^{۲۷} کلید-مقدار، معرفی تکنیک‌های توسعه
- بررسی سلسله مراتب، مراکز داده‌ی کوچک و کامل، مفهوم انتشار و عضویت^{۲۸} و کاربرد DHT، آشنایی با مفاهیم تئوری صف
- مسیریابی در شبکه‌های ابری، مسیریاب‌های با دسترس‌پذیری بالا
- جستجو و اشتراک‌گذاری فایل‌ها به صورت نقطه به نقطه، پروتکل‌های انتقال فایل
- خدمات ابری یک طبقه^{۲۹} و دو طبقه، فناوری‌های پوششی انطباقی برای معماری دو طبقه
- برنامه‌نویسی برای فضای ابری، ذخیره‌سازی و دسترسی به داده‌ها
- پروتکل Torrent و نقش آن در شبکه‌های نقطه به نقطه و مراکز داده‌ی ابری
- مفهوم شایعه و ساخت شبکه‌های پیچیده‌ی پوششی با استفاده از آن
- محاسبات جغرافیایی، نواحی دسترس‌پذیری، رونوشت^{۳۰} در شبکه‌های بی‌سیم محلی، قرینه‌سازی^{۳۱}
- انواع خطا و تحمل‌پذیری اشکال، رونوشت و کدهای محوکننده^{۳۲}، معرفی پروتکل‌های تعهد^{۳۳} تک مرحله‌ای و دو مرحله‌ای
- بررسی موازنه‌های موجود در سامانه‌های ابری، آشنایی با تئوری CAP
- آشنایی با اصل BASE و بررسی آن در تقابل با مدل ACID

²⁶ Big Data

²⁷ Sharding

²⁸ Publish-Subscribe

²⁹ One-Tier

³⁰ Replication

³¹ Mirroring

³² Erasure Codes

³³ Commit



- مفهوم زمان در محاسبات ابری، ابر بی‌درنگ^{۳۴}، ساعت‌های منطقی و ساعت‌های برداری^{۳۵}، بررسی موضوع ثبات^{۳۶}
- مجازی‌سازی^{۳۷} در محاسبات ابری
- مدل‌های نقل و انتقال^{۳۸} و پیاده‌سازی آنها، خدمات واقع‌نگاری^{۳۹} در سامانه‌های ابری
- بررسی ساختار فضای ابری در شرکت‌های بزرگ بین‌المللی
- امنیت در زیرساخت‌های ابری، معرفی مفهوم زنجیره‌ی بلوک‌ها، مفهوم گمنامی^{۴۰}، اثبات کار^{۴۱}
- مدل‌های تجاری در کاربردهای ابری
- فضای ابری آینده و تعامل با اینترنت اشیاء

ارزیابی

- فعالیت‌های کوتاه کلاسی و منزل (۲ نمره)
 - ✓ آزمون کوتاه
 - ✓ تحقیق و تمرین
- آزمون میان‌ترم (۷ نمره)
- آزمون پایانی (۷ نمره)
- پروژه نهایی (۴ نمره)

مراجع

- C. Davis. Cloud Native Patterns: Designing change-tolerant software. Simon and Schuster, 2019.
- I. Foster, and D. B. Gannon. Cloud Computing for Science and Engineering (Scientific and Engineering Computation). The MIT Press, 2017.
- M. Kleppmann. Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems. O'Reilly Media Inc, 2017.
- K. P. Birman. Guide to Reliable Distributed Systems: Building High-Assurance Applications and Cloud-Hosted Services. Springer Science & Business Media, 2012. Papers in the literature

اساتید

- دکتر صفایی
- دکتر حبیبی

³⁴ Real-Time
³⁵ Vector
³⁶ Consistency
³⁷ Virtualization
³⁸ Transaction
³⁹ Logging
⁴⁰ Anonymity
⁴¹ Proof of Work



الگوریتم‌های پیشرفته

Advanced Algorithms

اهداف درس

هدف اصلی این درس، ارائه چند موضوع مهم و تکمیل بخش‌هایی از درس کارشناسی «طراحی و تحلیل الگوریتم‌ها» است که برای گرفتن درس‌های ارشد نظری بعدی مورد نیاز است. این درس برای دانشجویان سایر رشته‌های ارشد مهندسی کامپیوتر چون هوش مصنوعی و سیستم‌های نرم‌افزاری توصیه می‌شود.

ریز مواد

- درهم‌سازی
 - درهم‌سازی کامل (hashing perfect)
 - درهم‌سازی سراسری (hashing universal)
 - جدول‌های پویا (tables dynamic)
- رده‌های پیچیدگی
 - تاریخچه کوتاه و تعریف‌ها
 - مفهوم کاهش
 - مسئله‌های کلاسیک و ان‌پی-کامل
 - رده‌های دیگر چون $\#P$ و NC و $complete-P$
- آشنایی با الگوریتم‌های تقریبی
 - پوشش گره‌ای (cover vertex)
 - پوشش مجموعه کمینه (cover set min)
 - کوتاه‌ترین ابررشته (superstring shortest)
 - فروشنده دوره‌گرد (TSP)
- شبکه شار پیشرفته و الگوریتم‌های شبکه
 - تاریخچه کوتاه و تعریف‌ها
 - الگوریتم‌های relabel-push, front-to-relabel
 - flow cost-min و برخی از الگوریتم‌های شبکه
- برنامه‌ریزی خطی
 - معرفی
 - محدودیت تفاضل
 - با ابعاد محدود
 - الگوریتم سیمپلکس
 - اثبات و تحلیل
 - دوگانی
- داده‌ساختارهای پیشرفته
 - هرم فیبوناچی
 - tree Boas Emde van



- الگوریتم‌های برخط
- آشنایی با هندسه محاسباتی
- آشنایی با الگوریتم‌های موازی
- الگوریتم‌های چندریشه‌ای
- آشنایی با الگوریتم‌های کوانتومی

ارزیابی

- تمرین: ۶ یا ۷ تمرین (۳۰ درصد نمره)
- آزمون: آزمون‌های میان‌نیمسال (۲۰ درصد نمره)، و پایان‌نیمسال (۲۵ درصد نمره)
- ارائه: ارائه در کلاس (۵ درصد نمره)
- مطالعه: فهم یک مقاله پژوهشی، تهیه خلاصه و ارائه (۲۰ درصد نمره)

مراجع

- T. Cormen, C. Leiserson, R. Rivest, and C. Stein. Introduction to Algorithms. 3rd ed. MIT Press, 2009.
- J. Kleinberg and E. Tardos. Algorithm Design. Addison-Wesley, 2005 (recommended).
- V. Vazirani. Approximation Algorithms. Springer-Verlag, 2001.
- M. Bazaraa, J. Jarvis, and H. Sherali. Linear Programming and Network Flows. 4th ed. Wiley, 2010 (recommended).
- F.T. Leighton. Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes. Morgan Kaufmann, 1992.

اساتید

- دکتر قدسی
- دکتر آبام
- دکتر ضرابی‌زاده



بهینه‌سازی ترکیبیاتی

Combinatorial Optimization

اهداف درس

هدف این درس آشنایی دانشجو با بهینه‌سازی ترکیبیاتی از طریق بررسی مسائل بهینه‌سازی و الگوریتم‌های خاص این حوزه از جمله الگوریتم‌های Simplex و Prime-Dual، دوگانی و برنامه‌ریزی خطی صحیح است.

ریز مواد

- مسئله‌های بهینه‌سازی
 - معرفی، همسایگی، بهینه‌سازی محلی و سراسری، مجموعه و توابع محدب، مسئله‌های برنامه‌ریزی محدب
- الگوریتم Simplex
 - فرم‌های مسئله‌ی برنامه‌ریزی خطی، راه‌حل امکان‌پذیر، هندسه‌ی برنامه‌ریزی خطی، الگوریتم و پیچیدگی سیمپلکس، جنبه‌های هندسی انتخاب محور
- دوگانی
 - دوگان برنامه‌ریزی خطی در حالت کلی، Slackness، مسئله‌ی کوتاه‌ترین مسیر و دوگان آن، مسئله‌ی شبکه‌ی شاره، دوگان الگوریتم سیمپلکس، جنبه‌های محاسباتی الگوریتم سیمپلکس
- الگوریتم Dual-Primal
 - الگوریتم و کاربرد آن در شبکه‌ی شاره و مسئله‌های کوتاه‌ترین مسیر دایکسترا، الگوریتم‌های کارا برای مسئله‌های شبکه‌ی شاره، تطابق (عادی و وزن‌دار)، درخت‌های پوشای کمینه
- برنامه‌ریزی خطی صحیح
 - ان‌پی-تمام بودن مسئله و راه‌های تقریبی، استفاده از بهینه‌سازی در حل تقریبی مسئله‌های ان‌پی-سخت

ارزیابی

- تمرین‌های نظری: ۳ نمره
- آزمون‌های میان‌ترم و پایانی: ۱۵ نمره
- گزارش پژوهشی: ۲ نمره

مراجع

- C.H. Papadimitriou and K. Steiglitz. Combinatorial Optimization: Algorithms and Complexity. Dover Publications, 1998.

اساتید

- دکتر ضرابی‌زاده
- دکتر قدسی



نظریه الگوریتمی بازی‌ها

Algorithmic Game Theory

اهداف درس

نظریه بازی‌ها کاربردهای وسیعی در بسیاری از حوزه‌ها دارد که مهم‌ترین آن‌ها حوزه‌های اقتصادی، کسب‌وکار، و علوم اجتماعی است. به طور کلی در نظریه بازی‌ها با سیستم‌هایی شامل عامل‌های هوشمند و خودخواه سروکار داریم که هر کدام از آن‌ها بنا به مصالح خویش وضعیت سیستم را تغییر می‌دهند. نظریه بازی‌ها ابزار تحلیل این‌گونه از سیستم‌ها را در اختیار ما قرار می‌دهد و کمک می‌کند تا بتوانیم آن‌ها را به شیوه‌ای درست و منطقی کنترل کنیم.

با توجه به رشد بازارهای دیجیتال و گسترده‌تر شدن و بزرگ‌شدن سیستم‌های کاربردی چندعامله با عامل‌های هوشمند، نیاز به توسعه روش‌ها و الگوهای محاسباتی و الگوریتمی بیشتر و بیشتر شده است. در بسیاری از موارد بدون این روش‌ها امکان تحلیل و طراحی استراتژی‌های کارا وجود ندارد.

هدف از این درس آشنایی با این روش‌ها و الگوریتم‌هاست که با توجه به ماهیت آن، در اصل کاربرد برخی تئوری‌های اصیل علوم کامپیوتر در این حوزه است.

ریز مواد

- آشنایی با مقدمات نظریه بازی‌ها - یادآوری (۸ جلسه)
- محاسبه نقاط تعادل و مسائل مربوطه (۶ جلسه):
 - بازی‌های صفرجمع دونفره (Zero-sum Games) و قضیه MinMax
 - بازی‌های صفرجمع چندنفره
 - قضیه نش (Nash Theorem)، لم اسپرنر (Sperner's Lemma) و قضیه بروور (Brouwer's Theorem)
 - الگوریتم لمکه هاوسون (Lemke Howson Algorithm)
 - مسائل جستجوی تام (Total Search Problems) و کلاس‌های پیچیدگی PPA, PPP, PPAD و PLS
 - کلاس پیچیدگی مسائل یافتن نقاط تعادل نش
- منطق، اتوماتا و بازی‌های بی‌نهایت (۷ جلسه):
 - گراف بازی (Game Graph) و شرایط برد
 - شرایط برد در حالت غیرقطعی (شرایط بوخی، مولر، رابین و ...) و تبدیلات آن‌ها
 - بازی‌های بی‌نهایت و تشخیص (Determinacy) و تشخیص بی‌حافظه (Memoryless Determinacy)
 - شرایط برد منطقی (Logical Winning Conditions)
 - اتوماتای درختی
 - بازی‌های زوجیت (Parity Games) و بازی‌های نیمه بازپرداخت (Mean Payoff) و حل آن‌ها
 - بازی‌های قابلیت رسیدن (Reachability Games)
 - بازی‌های تکرارشونده (Repeated Games)
 - فرآیند تصمیم‌گیری مارکف و بازی‌های تصادفی
 - شبیه‌سازی، دوتشابهی (Bisimulation) و بازی‌های ارنفوش-فریز (Ehrenfeucht-Fraïssé)
- طراحی مکانیزم الگوریتمی
 - آشنایی با مقدمات طراحی مکانیزم و لم مایرسون (Myerson's Lemma)
 - مزایده‌های پیشینه‌کننده سود (Revenue Maximizing Auctions) و قیمت رزرو شده (Reserved Price)
 - مزایده‌های نزدیک بهینه ساده (Near Optimal Auctions)، نامساوی پیامبر (Prophet Inequality) و قضیه بولو کلمپرر (Bulow Klemperer Theorem)
 - طراحی مکانیزم چند پارامتره و مکانیزم‌های VCG



- مزایده‌های ترکیبیاتی (Combinatorial Auctions) و مزایده طیف‌های بی‌سیم (Wireless Spectrum)
- طراحی مکانیزم برای حالت‌های غیرخطی - حالت‌های بودجه محدود و مزایده‌های پرچی (Clinching Auctions) و مکانیزم‌های بدون پول
- بازارهای تطابق و تطابق‌های پایدار، بازارهای تبادل کلیه (Kidney Exchange Markets)
- دینامیک بازی‌ها و مسائل یادگیری:
 - هزینه آشوب (Price of Anarchy) و هزینه ثبات (Price of Stability)
 - بازی‌های پتانسیلی (Potential Games)
 - بازی‌های نرم (Smooth Games) و هزینه آشوب مستحکم
 - تعادل‌های نش قوی (Strong Nash Equilibria)
 - دینامیک‌های بهترین پاسخ (Best Response Dynamics) و نقاط تعادل تقریبی
 - یادگیری، دینامیک‌های بازی ساختگی (Fictitious Play) و دینامیک‌های بدون حسرت (No-Regret)
- Dynamics و یادگیری تقویتی

ارزیابی

- آزمون (۶۰ درصد نمره)
- تمرین (۲۰ درصد نمره)
- سمینار (۲۰ درصد نمره)

مراجع اصلی

- T. Roughgarden. Twenty Lectures on Algorithmic Game Theory. Cambridge University Press, 2016.
- K.R. Apt and E. Grädel (Eds.), Lectures in game theory for computer scientists. Cambridge University Press, 2011.
- Y. Shoham and K. Leyton-Brown. Multiagent systems: Algorithmic, game-theoretic, and logical foundations. Cambridge University Press, 2008.
- N. Nisan, T. Roughgarden, E. Tardos, and V.V. Vazirani. Algorithmic game theory. Cambridge University Press, 2007.

اساتید

- دکتر صدیقین
- دکتر دهقانی
- دکتر فضلی



روش‌های تصادفی و احتمالاتی

Randomized and Probabilistic Methods

اهداف

هدف این درس، آشنا کردن دانشجویان کارشناسی ارشد و دکترا با روش‌های احتمالی و تصادفی در حوزه علوم داده می‌باشد.

ریز مواد

- مفاهیم پایه احتمال
 - تعاریف و قضایای پایه‌ای
 - امید ریاضی و نامساوی مارکوف
 - انحراف معیار و نامساوی چبیشف
 - نامساوی چرنف و نامساوی هوفدینگ و کاربردها: درجه گراف تصادفی
 - توزیع‌های احتمال: پواسون، هندسی، گاوسی، نمایی، ...
- الگوریتم‌های تصادفی پایه‌ای
 - الگوریتم‌های لاس و گاس و مونتوکارلو: مسئله میانه تقریبی
 - جایگشت تصادفی و کاربردهای آن: انتخاب، مرتب‌سازی، مسئله استخدام
 - مسئله برش کمینه
 - روش اثرانگشت: تساوی چندجمله‌ای‌ها و تطابق رشته‌ها
- قدم‌زدن تصادفی و فرایند مارکوف
 - قدم‌زدن تصادفی در گراف‌های بدون جهت و کاربرد آن در مسئله‌های 2-SAT و 3-SAT
 - فرایند مارکوف (قدم زدن تصادفی در گراف‌های جهت‌دار با توزیع داده‌شده) و محاسبه احتمال بودن در یک گره در بی‌نهایت
- روش مونتوکارلو
 - معرفی روش مونتوکارلو: تخمین عدد پی
 - معرفی الگوریتم‌های تقریبی تصادفی: مسئله شمارش جواب‌های DNF
 - نمونه‌برداری تقریبی و شمارش تقریبی: مسئله شمارش مجموعه‌های مستقل در گراف
 - روش نمونه‌برداری مونتوکارلو با استفاده از فرایند مارکوف
- مدل جویباری و روش‌های احتمالی
 - معرفی مدل و انواع آن: مسئله پیدا کردن اعداد با تکرار زیاد
 - توابع درهم‌ساز universal و 2-universal
 - پیدا کردن تعداد اعداد متمایز
 - تخمین Frequency Moments و کاربردهای آن: محاسبه تعداد مثلث‌ها در گراف
- بردارهای تصادفی در ابعاد بالا
 - تمرکز نرم‌ها
 - ماتریس کوواریانس بردارهای تصادفی
 - مثال‌هایی از توابع توزیع احتمالی در ابعاد بالا: برش بیشینه در گراف‌ها
 - لم Johnson-Lindenstrauss
- ماتریس‌های تصادفی
 - نت‌ها، عدد پوششی و عدد بسته‌بندی
 - کران بالا برای ماتریس‌های گاوسی: تشخیص اجتماع‌ها در شبکه‌ها



- کران دو طرفه برای ماتریس‌های گاوسی: تخمین کوواریانس و خوشه‌بندی
- نامساوی برنشتاین برای ماتریس‌ها: تشخیص اجتماع‌ها در شبکه‌های خلوت

ارزیابی

- آزمون: ۶۰٪ کل نمره
- تمرین تئوری: ۲۰٪ کل نمره
- تمرین عملی: ۲۰٪ کل نمره

مراجع اصلی

- M. Mitzenmacher and E. Upfal. Probability and Computing: Randomized Algorithms and Probabilistic Analysis. Cambridge University Press, 2005.
- R. Vershynin. High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge University Press, to appear.

اساتید

- دکتر آبام
- دکتر قدسی
- دکتر موقر



سامانه‌های پیشرفته ذخیره‌سازی داده

Advanced Storage System

اهداف

هدف از این درس، آشنایی دانش‌جویان با سیستم‌های پیشرفته ذخیره‌سازی داده و چالش‌ها و فناوری‌های مربوطه است

ریز مواد

- معرفی و لزوم استفاده از سامانه‌های ذخیره‌سازی
 - تاریخچه روش‌های ذخیره‌سازی داده
 - مقایسه سیر کارایی دیسک‌ها و پردازنده‌ها
 - بررسی قانون Amdahl در سامانه‌های ذخیره‌ساز داده
 - معماری روش ذخیره‌سازی مبتنی بر کارگزار
- معماری روش ذخیره‌سازی مبتنی بر سامانه ذخیره‌ساز داده
- معماری و پیکربندی ورودی/خروجی در زیرسامانه دیسک
 - محفظه‌ی JBOD
 - محفظه‌ی JBOD به همراه کنترلر RAID
 - زیرسامانه دیسک هوشمند
- معیارهای کیفی و کمی سامانه‌های ذخیره‌سازی داده
 - پهنای باند، زمان پاسخ، دسترس‌پذیری، قابلیت سرویس، قابلیت توسعه‌پذیری
 - ویژگی‌شناسی بارهای کاری ورودی/خروجی
- انواع پیکربندی دیسک‌ها در سامانه‌های ذخیره‌سازی داده
 - RAID ۱, RAID ۵, RAID ۱۰, RAID ۶
 - کدهای محوکننده‌ی خطا در سامانه‌های ذخیره‌سازی داده
 - بررسی کارایی خواندن، کارایی نوشتن و دسترس‌پذیری
- طراحی یک سامانه پیشرفته ذخیره‌ساز داده
 - طراحی منطق پسین
 - طراحی منطق پیشین
 - طراحی سامانه حافظه
- روند جریان داده در سامانه‌های ذخیره‌ساز داده
 - خواندن، نوشتن و کپی داده
- بررسی ویژگی‌های پیشرفته سامانه‌های ذخیره‌ساز داده
 - Remote Mirroring
 - Instant Copies
 - Data Migration
 - LUN Masking
- حافظه نهان در سامانه‌های ذخیره‌سازی داده
 - بررسی ساختار حافظه نهان در سامانه‌های ذخیره‌سازی داده
 - مقایسه ساختار حافظه نهان در سامانه‌های ذخیره‌سازی داده با ساختار حافظه نهان در ریزپردازنده‌ها
 - الگوریتم‌های حافظه نهان در سامانه‌های ذخیره‌سازی داده
- بررسی معماری‌های متداول سامانه‌های ذخیره‌ساز داده (IBM, HP, EMC)



- تکنیک‌های ورودی/خروجی در سامانه‌های ذخیره‌سازی داده
- حافظه‌های غیرفرار و دیسک‌های حالت جامد (Solid-State Drives)
 - انواع و ویژگی‌های حافظه‌های فلش
 - انواع و ویژگی‌های حافظه‌های غیرفرار
 - فناوری‌های ذخیره‌سازی نوین مبتنی بر حافظه‌های غیرفرار
 - معماری دیسک‌های حالت جامد
 - معماری لایه انتقال دیسک حالت جامد (Flash Translation Layer)
 - الگوریتم‌های Wear Leveling در دیسک‌های حالت جامد
 - روش‌های افزایش طول عمر دیسک حالت جامد
 - روش‌های زمان‌بندی ورودی/خروجی در دیسک‌های حالت جامد

ارزیابی

- تمرین‌های نظری: ۳ نمره
- آزمون‌های میان‌ترم و پایانی: ۱۵ نمره
- آزمونک‌ها: ۲ نمره

مراجع

- Information Storage and Management, by G. Somasundaram and A. Shrivastava, Wiley Publishing Inc., EMC Education Services, 2009.
- Storage Networks Explained: Basics and Application of Fibre Channel SAN, NAS, iSCSI, InfiniBand and FCoE, U. Troppens, R. Erkens, W. Mueller-Friedt, and R. Wolafka, 2nd Edition, John Wiley & Sons Inc., 2009.
- Storage Area Networks Essentials, R. Barker and P. Massiglia, John Wiley & Sons Inc., 2002.

اساتید

- دکتر اسدی



یادگیری ژرف

Deep Learning

اهداف درس

این درس به حوزه‌ای از یادگیری ماشین تحت عنوان یادگیری ژرف که در سال‌های اخیر بسیار مورد توجه قرار گرفته و به عملکرد چشم‌گیری در بسیاری از کاربردها دست یافته است خواهد پرداخت. در طول این درس ابتدا مفاهیم اولیه، نظیر شبکه‌های عصبی چندلایه، قدرت مدل‌سازی این شبکه‌ها و نحوه‌ی آموزش آن‌ها بحث می‌شود. سپس آشنایی با معماری‌های اصلی نظیر شبکه‌های CNN و RNN صورت خواهد گرفت. همچنین پیشرفت‌هایی که در طراحی، بهینه‌سازی، بهبود تعمیم‌پذیری و نحوه‌ی آموزش شبکه‌ها در حوزه یادگیری ژرف صورت گرفته است معرفی می‌شود. مدل‌های مولد نیز به عنوان یکی از شاخه‌های مهم مورد بررسی قرار خواهند گرفت. به علاوه به تعدادی از شبکه‌های ژرف معروف که طی سال‌های اخیر معرفی شده‌اند، اشاره خواهد شد. در طول درس به کاربردهای مهم شبکه‌های معرفی شده به خصوص در زمینه‌های بینایی ماشین و پردازش زبان طبیعی اشاره خواهد شد.

ریز مواد

- مقدمه و معرفی شبکه‌های عصبی مصنوعی
- پرسپترون چند لایه (Multi-layer Perceptron)
 - MLP به عنوان تقریب‌زننده عمومی (Universal approximator)
- الگوریتم انتشار رو به عقب خطا (Error back propagation)
- بهینه‌سازی در شبکه‌های ژرف
 - مروری بر بهینه‌سازی محدب
 - معرفی انواع روشهای SGD, Momentum, RMS Prop, Adams و ...
- تکنیک‌هایی در آموزش، طراحی و تعمیم‌پذیری شبکه‌های ژرف
 - معرفی تکنیک‌های بهبود تعمیم‌پذیری نظیر regularization, dropout, data augmentation
 - هنجارسازی بسته‌ای (Batch Normalization)
 - انتخاب توابع فعالیت (activation function)، مقداردهی اولیه و وزن‌ها، هنجارسازی ورودی و ...
- شبکه‌های عصبی کانولوشنی (Convolutional Neural Networks)
 - لایه‌های convolution و pooling
 - معماریهای معروف شبکه‌های CNN
 - کاربردهای مختلف شبکه‌های CNN
- شبکه‌های عصبی بازگردنده (Recurrent Neural Networks)
 - مدل‌سازی دنباله‌ها
 - حافظه‌های بلند کوتاه مدت (Long Short Term Memories)
 - شبکه‌های توجه (Attention Networks)
 - مدل‌سازی زبانی (Language Modeling) با استفاده از شبکه‌های RNN
 - کاربردهای دیگر شبکه‌های RNN در زمینه‌های مختلف نظیر پردازش زبان طبیعی (Natural Language Processing)
- معماری تبدیل کننده (Transformer)
- شبکه‌های جمع-ضرب (Product-Sum)
- مدل‌های مولد (Generative Models)
 - مدل‌های Autoregressive
 - خودکدگذار وردشی (Variational)



- شبکه‌های مولد حریفانه (Generative Adversarial Networks)
- مدل‌های مولد مبتنی بر جریان (Flow based)
- یادگیری تقویتی ژرف (Deep Reinforcement Learning)
 - یادگیری تقویتی ژرف با استفاده از توابع Q
 - رویکرد گرادین سیاست (Policy Gradient)
 - رویکرد بازیگر-نقاد (Actor Critic)
- نمونه‌های خصمانه (Adversarial) و مقاومت شبکه‌های ژرف به نمونه‌های خصمانه
- مباحث پیشرفته
 - شبکه‌های دوگان و یادگیری دوگان (Dual Learning)
 - شبکه‌های کانولوشن گرافی
 - یادگیری خودنظارتی (Self-supervised)

ارزیابی

- تمرین: ۳۰٪
- میان‌ترم: ۲۰٪
- پایان‌ترم: ۳۰٪
- آزمون‌های کوتاه: ۱۰٪
- پروژه یا کار تحقیقاتی: ۱۰٪

مراجع

- Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, Book in preparation for MIT Press, 2016.
- Michael Nielsen, Neural networks and deep learning, Preprint, 2016

اساتید

- دکتر سلیمانی
- دکتر بیگی



تحلیل آماری داده‌ها

Statistical Data Analysis

اهداف

این درس سعی دارد تا پایه‌های تئوری و عملی آماری لازم برای تحلیل داده‌ها را فراهم نماید.

ریزموارد

- آشنایی با احتمالات و آمار
 - تخمین
 - تخمین نقطه‌ای
 - تخمین بازه‌ای
 - تخمینگر بیشینه درست‌نمایی
 - تست فرضیه
 - خطای نوع ۱ و نوع ۲
 - P-value
 - ارتباط با بازه اطمینان
 - برخی از تست‌های مهم
 - رگرسیون خطی
 - روش کمترین مربعات
 - خواص تخمین‌گر کمترین مربعات
 - استنتاج درباره ضرایب رگرسیون
 - پیش‌بینی
 - روش تحلیل واریانس
 - رگرسیون چند متغیره
 - انتخاب مدل و بررسی صحت آن
 - تحلیل واریانس
 - مدل تحلیل واریانس یک طرفه
 - مدل تحلیل واریانس دوطرفه
 - روش بدون پارامتر
 - تست Wilcoxin signed-ranked
 - تست Kruskal-Wallis
 - ضرایب همبستگی Spearman rank
 - آمار بیزی
 - استنتاج بیزی
 - تخمینگر نقطه‌ای بیزی
 - تخمینگر بازه‌ای بیزی

مراجع



- Ronald Walpole, Raymond Myers, Sharon Myers, Keying Ye, Probability & Statistics for Engineers & Scientists, Pearson; 9th edition (March 7, 2016)
- Ajit Tamhane, Dorothy Dunlop, Statistics and Data Analysis: From Elementary to Intermediate, Pearson; 1st edition (October 18, 1999)
- George Casella, Roger L. Berger, Statistical Inference, Cengage Learning; 2nd edition (June 18, 2001).

اساتید

- دکتر موقر
- دکتر جعفری
- دکتر ربیعی
- دکتر مطهری



تحلیل سری‌های زمانی

Time Series Analysis

اهداف

بسیاری از نیازمندی‌های دنیای واقعی را می‌توان به صورت فرآیندهای تصادفی مدل‌سازی کرد. در این مسائل، انواع تحلیل، مانند توصیف، پیش‌بینی کمک شایانی به تصمیم‌گیری‌های هوشمند می‌کند. در این درس، با انواع سری‌های زمانی و کاربردهای آن و همین‌طور تحلیل‌های مرتبط آشنا می‌شویم.

ریز موارد

- مرور آمار و احتمال و فرآیندهای تصادفی
- مقدمات سری‌های زمانی
 - انواع و ویژگی‌های سری‌های زمانی
 - مدل‌سازی‌های اولیه
 - مدل‌های مرتبه اول خود رگرسیون
 - تابع خودهمبستگی (ACF)
 - تابع خودهمبستگی پاره‌ای (PACF)
- آشنایی با مدل‌های متوسط متحرک خود رگرسیون (ARMA)
- مدل Box-Jenkins یا ARIMA
- مدل‌های فصلی (seasonal)
 - تفاوت بین فصلی بودن و روند داشتن
 - مدل ARIMA فصلی
- روش‌های نرم کردن و تجزیه (Decomposition)
- تشخیص پریودهای اصلی سری زمانی به کمک Periodogram
- رگرسیون به کمک خطای مدل ARIMA و تابع Cross Correlation یا CCF
- پیش سفید کردن (Pre-whitening)
- تحلیل مداخله (Intervention Analysis)
- تحلیل داده‌های Longitudinal
- مدل‌های خود رگرسیون برداری و مدل ARCH
- روش‌های تحلیل غیرپارامتری و تخمین طیفی توزیع
- مدل‌های حافظه‌دار ARIMA
- روش‌های یادگیری ماشین
 - روش CNN
 - روش LSTM
 - روش‌های مبتنی بر یادگیری تقویتی

مراجع اصلی

- R.H. Shumway, D.S. Stoffer, "Time Series Analysis and Its Applications With R Examples," 4th Edition, Springer, 2012



اساتید

- دکتر رهبان
- دکتر موقر



بهینه‌سازی محدب

Convex Optimization

اهداف درس

هدف از این درس، آشنایی دانش‌جویان با مفاهیم بهینه‌سازی محدب، الگوریتم‌ها، و کاربردهای آن است.

ریز مواد

- مجموعه‌های محدب (Convex Set)
 - مجموعه‌های محدب و آفین (Affine and Convex Sets)
 - عملیاتی که تحدب را حفظ می‌کند
- توابع محدب
 - خواص مقدماتی
 - عملیاتی که تحدب را حفظ می‌کنند
 - توابع مزدوج (Conjugate Functions)
 - توابع شبه‌محدب (Quasiconvex Functions)
- مسایل بهینه‌سازی محدب
 - مسایل بهینه‌سازی در حالت کلی
 - بهینه‌سازی محدب
 - مسایل بهینه‌سازی خطی
 - مسایل بهینه‌سازی درجه دوم
- دوگانگی (Duality)
 - تابع دوگان لاگرانژ
 - مسئله دوگان لاگرانژ
 - تعبیر هندسی و تعبیر نقطه زینی (Geometric and Saddle-point Interpretations)
 - شرایط بهینگی (Optimality Conditions and KKT conditions)
 - اختلال و تحلیل حساسیت (Perturbation and Sensitivity Analysis)
- الگوریتم‌های بهینه‌سازی نامقید
 - مسایل بهینه‌سازی نامقید
 - الگوریتم‌های کاهشی (Descent Algorithms)
 - الگوریتم کاهشی گرادیان (Gradient Descent Algorithm)
 - الگوریتم کاهش با تندترین شیب (Steepest Descent Algorithm)
 - الگوریتم نیوتون (Newton's Method)
- الگوریتم‌های بهینه‌سازی با قیود تساوی
 - مسایل بهینه‌سازی با قیود تساوی
 - روش نیوتون با قیود تساوی
- الگوریتم‌های بهینه‌سازی با قیود نامساوی (روش‌های نقطه درونی، Interior-point Methods)



- مسایل بهینه‌سازی با قیود نامساوی
- تابع سد لگاریتمی و مسیر مرکزی (Logarithmic Barrier Function and Central Path)
- روش سد (Barrier Method)
- امکان‌پذیری و روش‌های فاز یک (Feasibility and Phase I Methods)
- روش‌های نقطه درونی اولیه-دوگان (Primal-Dual Interior-point Methods)
- کاربردها (Applications)
 - رگرسیون خطی، رگرسیون لجستیکی
 - ماشین بردار پشتیبان
 - مسایل بهینه‌سازی رنک-پایین (low-rank optimization problems (e.g., Netflix, video security))
 - اینترنت به عنوان یک مسئله بهینه‌سازی محدب
- بهینه‌سازی تصادفی (Stochastic Optimization)
 - تعریف کلی یک مسالهی بهینه‌سازی تصادفی
 - انواع قیود در بهینه‌سازی تصادفی
 - مسایل بهینه‌سازی تصادفی نوع یک و دو (Type I & Type II Stochastic Optimization Problems)
 - الگوریتم کاهشی گرادیان تصادفی (Stochastic Gradient Descent Algorithm)
 - الگوریتم‌های بهینه‌سازی تصادفی موازی و توزیع‌شده (Parallel and Distributed Stochastic Optimization)
- (Algorithms)

مراجع

- Stephen Boyd and Lieven Vandenberghe, "Convex optimization," Cambridge university press, 2004.
- Jorge Nocedal and Stephen Wright, "Numerical optimization," Springer Series in Operations Research and Financial Engineering, 2nd edition, 2006.
- "Convex Optimization" taught by Ryan Tibshirani at CMU from 2013 to 2019

اساتید

- دکتر جعفری
- دکتر رهبان



مدل‌های احتمالاتی گرافی

Probabilistic Graphical Models

اهداف درس

مدل‌های گرافی احتمالی، چارچوب محاسباتی کلی برای استنتاج و یادگیری در شرایط نایقینی (uncertainty) فراهم می‌آورند. در این مدل‌ها برای نمایش وابستگی‌های شرطی بین متغیرهای تصادفی از یک گراف استفاده شده و توزیع توام مجموعه‌ی متغیرهای تصادفی از این طریق مشخص می‌شود. در این درس، برای معرفی مدل‌های گرافی احتمالی سه جنبه موردتوجه قرار می‌گیرد: بازنمایی در این مدل‌ها (شبکه‌های بی‌زین و میدان‌های تصادفی مارکوف)؛ یادگیری (پارامترها و ساختار) این مدل‌ها از روی داده‌ها؛ انجام استنتاج (با روش‌های دقیق و تقریبی) جهت استفاده از مدل‌های گرافی احتمالی برای تصمیم‌گیری در شرایط نایقینی. در شروع این درس لازم است دانشجویان آشنایی با مباحث آمار و احتمال و همچنین مقدمات یادگیری ماشین داشته باشند.

ریز مواد

- معرفی مدل‌های گرافی (برای نمایش دانش احتمالی)
 - مدل‌های گرافی جهت‌دار: شبکه‌ی بی‌زین (Bayesian Networks)
 - مدل‌های گرافی بدون جهت: میدان‌های تصادفی مارکوف (Markov Random Fields)
 - دیدگاهی واحد برای مدل‌های گرافی جهت‌دار و بدون جهت
- استنتاج دقیق در مدل‌های گرافی
 - الگوریتم حذف متغیر (Variable Elimination)
 - انتشار اعتقاد (Belief Propagation) یا انتقال پیام (Message Passing)
 - گراف‌های عامل (Factor Graphs) و الگوریتم جمع-ضرب (Product-Sum)
 - تخمین MAP: الگوریتم بیشینه-ضرب (Product-Max)
 - الگوریتم درخت اتصال (Junction Tree)
- یادگیری مدل‌های گرافی
 - یادگیری مدل‌های جهت‌دار کاملاً مشاهده شده
 - یادگیری مدل‌های بدون جهت کاملاً مشاهده شده
 - الگوریتم EM برای یادگیری مدل‌های گرافی نیمه مشاهده شده
 - یادگیری ساختار مدل‌های گرافی
- مدل‌های گرافی مشهور
 - خانواده نمایی (Exponential Family)
 - مدل‌های گرافی گاوسی
 - مدل‌های Ising (یا MRF دوبه‌دو)
 - میدان تصادفی شرطی (CRF)
 - مدل‌های زمانی و مدل‌های حالت-فضا
 - مدل مخفی مارکوف (HMM)
 - سامانه خطی پویا (LDS)
 - فیلتر کالمن (Kalman Filter)



● استنتاج تقریبی رویکرد قطعی

○ انتشار اعتقاد حلقه‌ای (Loopy Belief Propagation)

○ استنتاج وردشی (Variational Inference)

○ تقریب میدان میانگین (Field-Mean)

○ تخمین چگالی مفروض

○ روش‌های وردشی ساختاردار (structured)

● استنتاج تقریبی رویکرد تصادفی

○ نمونه‌برداری رد (Rejection Sampling)

○ نمونه‌برداری اهمیت (Importance Sampling)

○ زنجیره‌ی مارکوف مونت کارلو (MCMC)

○ الگوریتم متروپلیس-هیستینگز (Metropolis Hastings)

○ نمونه‌برداری گیبس (Gibbs)

ارزیابی

● تمرین: ۱۵٪

● میان‌ترم: ۳۰٪

● پایان‌ترم: ۴۰٪

● پروژه یا کار تحقیقاتی: ۱۵٪

مراجع

- D. Koller and N. Friedman, “Probabilistic Graphical Models: Principles and Techniques”, MIT Press, 2009.
- M. Wainwright and M.I. Jordan, “Graphical Models, Exponential Families, and Variational Inference”, Foundations and Trends in Machine Learning, vol. 1, pp. 1-305, 2008.
- M.I. Jordan, “An Introduction to Probabilistic Graphical Models”, In preparation.
- C.M. Bishop, “Pattern Recognition and Machine Learning”, Springer, 2006.
- K.P. Murphy, “Machine Learning: A Probabilistic Perspective”, MIT Press, 2012

اساتید

● دکتر سلیمانی



آمار در ابعاد بالا

High-Dimensional Statistics

اهداف

آمار سنتی با رژیم‌هایی روبرو است که ابعاد داده مشخص و کم است و در آن هدف یافتن پارامترهای ناشناخته با استفاده از داده‌ها است. در این رژیم‌های همینطور تعداد داده انبوه فرض می‌شود. برخلاف این رویکرد، در آمار با ابعاد بالا فرض می‌کنیم ابعاد داده الزاماً مشخص و محدود نیست و ممکن است با تعداد داده‌ها تغییر کند. در این درس چارچوب نظری آماری برای تحلیل مسائل ابعاد بالا ارائه می‌کنیم.

ریز مواد

- مقدمه و مرور کلی مباحث
- کران‌های تمرکز (Concentration Inequalities)
 - کران‌های کلاسیک
 - متغیرهای تصادفی زیرگوسی و باندهای هافدینگ
 - متغیرهای تصادفی زیرنمایی و باندهای Bernstein
 - روش‌های مبتنی بر Martingale
 - توابع Lipschitz از متغیرهای تصادفی گوسی
- تمرکز معیار (Concentration Measure)
- قانون یکنواخت اعداد بزرگ
- انتروپی متریک
- ماتریس‌های رندم و تخمین ماتریس کوواریانس
- مدل‌های خطی تنک در ابعاد بالا
 - فرض‌های NSP و Restricted NSP
 - کران خطای بازیابی Support و همینطور خطای نرم ۲
 - باندهای Oracle
- مفهوم Reproducing Kernel Hilbert Space
- مدل‌های غیرپارامتری
 - کران‌های مبتنی بر انتروپی متریک
- کران‌های minimax
 - نامساوی Le Cam
 - روش Fano

مراجع اصلی

- Martin J. Wainwright, "High-dimensional statistics: A non-asymptotic view point," Cambridge University Press, 2019.
- Roman Vershynin, "High-dimensional probability: An introduction with applications in data science," Cambridge University Press, 2018.

اساتید

- دکتر رهبان
- دکتر مطهری



• دکتر موقر



پردازش زبان طبیعی

Natural Language Processing

اهداف درس

پردازش زبان طبیعی یکی از شاخه‌های بسیار مهم در پردازش زبان طبیعی به حساب می‌آید. هدف این علم ایجاد یک راه تعامل میان انسان و ماشین از طریق زبان طبیعی انسانی است. در این درس بنا داریم با مفاهیم پردازش زبان طبیعی آشنا شده و روش‌های پایه در حل مسایل موجود در آن را معرفی کنیم. همچنین تا حدی در مورد روشهای به روز در حل مسائل نیز صحبت خواهیم کرد.

ریز مواد

- معرفی درس و اهداف آن
 - معرفی با کلیات درس، سیلابس آن و آنچه در درس پوشش داده میشود
- آشنایی با پردازش زبان طبیعی
 - معرفی پردازش زبان طبیعی، تاریخچه مختصر، مسائلی مختلف موجود در آن و چالش‌های آن
- روش‌های پیش پردازش متن
 - عبارات منظم، توکنایزر، نرمالسازی، تشخیص ریشه و لمّا، تشخیص مرز جملات، نحوه محاسبه فاصله با MED
- مدل‌های زبانی
 - روش‌های پایه در مدل‌سازی زبانی، ان گرام‌ها، smoothing, perplexity
- روش‌های پایه در دسته‌بندی مستندات متنی
 - مفهوم دسته‌بندی، نحوه استخراج ویژگی از متن، چند دسته‌بند ساده، معرفی مدل لاجستیک رگرسیون برای دسته‌بندی، تعمیم مدل لاجستیک رگرسیون به شبکه عصبی
- روش‌های پایه در خوشه‌بندی مستندات متنی
 - مفهوم خوشه‌بندی و روش‌های پایه kmeans و mixture models
- بازنمایی کلمات
 - آشنایی با روش‌های مختلف بازنمایی کلمات: روش‌های پایه‌ای، روش‌های مبتنی بر جبر خطی، روش‌های مبتنی بر شبکه عصبی، چالش‌های مختلف در بازنمایی کلمات و راه‌های حل آنها، بازنمایی‌های مبتنی بر بافت (معرفی ساده)
- ترجمه ماشینی
 - معرفی مدل‌های سنتی ترجمه ماشینی، مدل‌های IBM، مدل‌های مبتنی بر عبارات
- شبکه‌های عصبی بازگشتی و مسدل مبتنی بر توجه
 - معرفی ساختار شبکه‌های عصبی بازگشتی ساده و ساختارهای معروف مثل LSTM و GRU، معرفی مدل‌های روز
- ترجمه ماشینی، معرفی مدل توجه
- مساله تجزیه در پردازش زبان
 - آشنایی با انواع تجزیه‌های معنایی و نحوی
 - معرفی مدل‌های پایه برای حل مساله تجزیه
- مسایل کاربردی دیگر در پردازش زبان طبیعی
 - معرفی مسایل دیگر در حوزه پردازش زبان طبیعی مانند استخراج اطلاعات، خلاصه‌سازی، برچسب‌زنی ادات سخن و ...

ارزیابی



- تمرین‌های عملی و سمینار پایانی: ۳۰٪
- آزمون میان‌ترم: ۲۰٪
- آزمون پایانی: ۳۰٪
- پروژه پایانی: ۲۰٪

مراجع

- Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. Draft)
- Manning and Schuetze, Foundations of Statistical Natural Language Processing
- Yoav Goldberg. A Primer on Neural Network Models for Natural Language Processing
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning

اساتید

- دکتر قاسم ثانی
- دکتر عسگری



تحلیل داده‌های چندرسانه‌ای مقیاس بزرگ

Large Scale Multi-Media Data Analysis

اهداف

در این درس دانشجویان با پردازش صدا و گفتار، فایل‌های چند رسانه‌ای، پردازش سیگنال چندوجهی، و بازیابی ویدئو در مقیاس بزرگ آشنا خواهند شد. دانشجویان جنبه‌های یک سیستم جستجوی چندرسانه‌ای پیشرفته را درک کرده و با الگوریتم‌های یادگیری ژرف برای بازیابی اطلاعات، تشخیص گفتار و پردازش صدا، پردازش تصویر و ویدئو آشنا شده و پیچیدگی‌های مدیریت حجم زیادی از داده‌های چندرسانه‌ای ناهمگن را تجربه خواهند کرد. آنها تجربه عمیق و عملی با برخی از الگوریتم‌های درگیر در پردازش (تشخیص و یا سنتز) و همجوشی چندوجهی خواهند داشت.

ریز مواد

- مقدمه و کاربردها؛ مشکلات چند رسانه‌ای
- مقدمه‌ای بر روش تحقیق چندرسانه‌ای
- معرفی ابزارهای مدل‌سازی و پردازش چند رسانه‌ای ها
- ضبط / ذخیره‌سازی چند رسانه‌ای - صدا
- پردازش چند رسانه‌ای - صدا
- پردازش چند رسانه‌ای - الگوریتم‌های عمومی
- ضبط / ذخیره‌سازی چند رسانه‌ای - متن
- ضبط / ذخیره‌سازی چند رسانه‌ای - تصاویر
- ضبط / ذخیره‌سازی چند رسانه‌ای - ویدئو
- پردازش تصویر در مقیاس بزرگ
- پردازش ویدئو در مقیاس بزرگ
- همجوشی داده‌های چند رسانه‌ای
- مقیاس‌پذیری چند رسانه‌ای
- چند رسانه‌ای در حقوق بشر، امنیت عمومی، و اخلاق
- یادگیری عمیق چند رسانه‌ای
- یادگیری خود نظارتی چند رسانه‌ای
- ترانسفورماتورهای بصری
- سیستم‌های سوالات-پاسخ‌های چند رسانه‌ای
- جستجوی چند رسانه‌ای
- استفاده از بینایی و زبان برای تولید تصویر
- شبکه متخاصم تولیدکننده متن به تصویر و ویدئو

ارزیابی

- تمرین‌ها: ۳ نمره
- آزمونک‌ها: ۲ نمره
- پروژه: ۷ نمره
- آزمون‌های میان‌ترم و پایانی: ۸



مراجع

- Stefanos Vrochidis, Benoit Huet, Edward Y. Chang, Ioannis Kompatsiaris, Big Data Analytics For Large-Scale Multimedia Search, Wiley & Sons, 2019.
- Edward Chang, Foundations of Large-Scale Multimedia Information Management and Retrieval, Springer, 2011.
- Sicheng Zhao, Min Xu, Qingming Huang, Björn W. Schuller, IEEE Multimedia, Special Issue: Multimodal Affective Computing of Large-Scale Multimedia Data, 2021.

اساتید

- دکتر ربیعی
- دکتر منظوری
- دکتر جم‌زاد
- دکتر صامتی



یادگیری ماشین مقیاس پذیر

Scalable Machine Learning

اهداف

با توجه به افزایش دادگان حجیم و نیاز بسیاری از مدل‌های یادگیری ماشین به دادگان بزرگ برای یادگیری، نیازمند چارچوب‌هایی هستیم که امکان استفاده از روش‌های یادگیری ماشین برای این نوع دادگان را فراهم کند. در این درس، اصول و مفاهیم یادگیری ماشین توزیع شده و موازی و همچنین ابزارهای موجود برای پیاده‌سازی این مفاهیم را معرفی می‌کنیم.

ریز مواد

- مفاهیم اولیه محاسبات توزیع شده
- آشنایی با Apache Spark
 - اجرای برنامه‌های کاربردی Spark
 - آشنایی با نشست Spark
 - آشنایی با دادگان توزیع شده قابل بازیابی (Resilient Distributed Dataset)
 - منابع داده‌ای Spark
- کتابخانه Spark MLlib (دو جلسه)
 - مروری بر الگوریتم‌های پیاده‌سازی شده
 - خط لوله یادگیری ماشین
 - پایداری مدل (Model Persistence)
- روش‌های یادگیری با ناظر توزیع شده
 - روش‌های رگرسیون
 - روش XGBoost
 - روش ماشین بردار پشتیبان
- روش‌های بدون ناظر توزیع شده
 - خوشه‌بندی (k-means)
 - مدل‌سازی موضوع (topic modeling)
 - تشخیص ناهنجاری
 - کاهش بعد
- تحلیل داده‌های مبتنی گراف مقیاس پذیر
 - آشنایی با GraphX
- روش‌های یادگیری متحد (Federated Learning)
- یادگیری ژرف توزیع شده و موازی
 - موازی سازی مدل در مقابل موازی سازی داده
 - ابزارهای موازی سازی
 - روش Data Parallel در Pytorch
 - روش Distributed Data Parallel در Pytorch
 - روش‌های فشرده‌سازی گرادیان و مدل



مراجع اصلی

- Butch Quinto, “Next-Generation Machine Learning with Spark,” Apress, 2020.
- Ron Bekkerman, Mikhail Bilenko, John Langford, “Scaling Up Machine Learning : Parallel and Distributed Approaches,” Cambridge University Press, 2012.

اساتید

- دکتر بیگی
- دکتر رهبان



نظریه یادگیری ماشین

Machine Learning Theory

اهداف درس

هدف از این درس، آشنایی دانش‌جویان با مفاهیم نظری الگوریتم‌های یادگیری ماشین و تحلیل الگوریتم‌ها است. تاکید این درس بیشتر روی روش‌های آماری محاسباتی متمرکز می‌شود. تلاش می‌شود در درس به پرسش‌های زیر پاسخ داده شود. آیا می‌توان الگوریتم یادگیری را طراحی نمود که کران کارایی آن قابل اثبات باشد؟ چگونه می‌توان الگوریتم یادگیری طراحی کرد که خواص ویژه‌ای داشته باشد.

ریز مواد

- مقدمه و مدل رسمی یادگیری
- مدل‌های یادگیری و محاسبه کران خطا
 - یادگیری در مدل سازگاری
 - یادگیری در مدل احتمالا تقریبا درست و یادگیری به کمک همگرایی یکنواخت
 - یادگیری در مدل احتمالا تقریبا درست بدون پیش فرض
 - یادگیرهای عمومی
 - تحلیل الگوریتم‌های کمینه سازی خطای تجربی
- معیارهای محاسبه غنای فضای فرضیه و محاسبه کران خطا
 - تابع رشد، بعد VC و اعداد پوششی
 - پیچیدگی راداماکر
- انتخاب مدل
- یادگیری نایکنواخت و تحلیل الگوریتم‌های کمینه سازی خطای ساختاری مانند توصیف با کمترین طول
- روش‌های منظم سازی و پایداری الگوریتم‌های یادگیری
- پیچیدگی محاسباتی الگوریتم‌های یادگیری
- الگوریتم‌های یادگیری برخط
- الگوریتم‌های یادگیری رتبه بندی
- الگوریتم‌های یادگیری فعال
- نظریه PAC-Bayesian
- مبانی نظری خوشه بندی
- مسایل یادگیری محدب
- تحلیل الگوریتم‌های یادگیری
 - تحلیل الگوریتم ماشین بردار پشتیبان
 - تحلیل الگوریتم‌های مبتنی بر هسته
 - تحلیل الگوریتم‌های بوستینگ
 - تحلیل الگوریتم‌های رگرسیون

ارزیابی

- تمرین‌های نظری: ۴ نمره
- آزمون‌های میان‌ترم و پایانی: ۱۲ نمره



- آزمونک‌ها: ۲ نمره
- ارایه مقاله و پروژه: ۲ نمره

مراجع

- M. Mohri, A. Rostamizadeh, and A. Talwalkar, Foundations of Machine Learning, second edition ed. MIT Press, 2018.
- S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning : From Theory to Algorithms. Cambridge University Press, 2014

اساتید

- دکتر بیگی
- دکتر رهبان
- دکتر اعتصامی



یادگیری ماشین آماری

Statistical Learning Theory

اهداف

- آشنایی با روش‌های بیزین ناپارامتری
- آشنایی با فرآیندهای تصادفی نقطه‌ای
- آشنایی با یادگیری عمیق
- آشنایی با داده‌های حجیم

ریزموارد

- روش‌های بیزین ناپارامتری
 - فرآیند دیریشله
 - فرآیند بوفه‌ی هندی
 - فرآیند گوسین
- فرآیندهای تصادفی نقطه‌ای
 - قضیه‌ی جمع آثار و نگاشت
 - اندازه‌ی تصادفی کامل و وابسته
 - فرآیندهای نقطه‌ای زمانی و هاوکس
- یادگیری عمیق
 - خودرمزگذار عمیق
 - شبکه‌های عمیق پیش‌آموزش شده
 - شبکه‌های عمیق پشت‌هم
- داده‌های حجیم
 - سیستم‌های حجیم‌داده
 - تکنولوژی هداپ
 - اسپارک
 - الگوریتم‌های مقیاس‌پذیر

منابع

- P Orbanz. Lecture Notes on Bayesian Nonparametrics. 2014.
- Carl E. Rasmussen, Christopher K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006.
- Kingman, John Frank Charles. Poisson processes. Oxford university press, 1992.
- Rasmussen, Jakob Gulddahl. Aspects of temporal and spatio-temporal processes. PhD Thesis, Department of Mathematical Science, Aalborg University, 2006.
- Daley, Daryl J., and David Vere-Jones. An introduction to the theory of point processes. Vol. 2. New York: Springer, 1988.
- Li Deng and Dong Yu. Deep Learning: Methods and Applications, Microsoft Technical Report MSR-TR-2014-21.
- J. Leskovec, A. Rajaraman, and J. D. Ullman, Mining of Massive Datasets, Stanford University, 2014.

اساتید

- دکتر ربیعی
- دکتر رهبان





فرآیندهای تصادفی

Stochastic Processes

اهداف درس

هدف از این درس، آشنایی دانش‌جویان با مدل‌های تصادفی اندیس شده با زمان یا مکان است. به صورت دقیق‌تر، توصیف و استنتاج بر اساس این مدل‌ها و همین‌طور تخمین پارامترهای ناشناخته از جمله اهداف این درس است.

ریز مواد

- پیشگفتار
- مروری بر نظریه احتمالات
 - اصول موضوعه احتمالات
 - وقایع و آزمایش‌های مستقل
 - متغیرهای تصادفی
 - امید ریاضی
 - گشتاورهای یک متغیر تصادفی
 - امید ریاضی شرطی
 - تابع مولد گشتاور
 - نامساوی‌های پایه در نظریه احتمال: نامساوی‌های مارکف، چبیشف و چرنف
 - قوانین اعداد بزرگ
 - قضیه حد مرکزی
 - انواع همگرایی‌های متغیرهای تصادفی
- فرایندهای تصادفی (مفاهیم اصلی)
 - مفاهیم پایه و تعاریف مقدماتی فرایندهای تصادفی
 - خواص آماری فرایندهای تصادفی
 - فرایندهای تصادفی ایستا (Stationary Stochastic Processes)
 - فرایندهای تصادفی متناوب با معیار Mean Square
 - چند مثال برای فرایندهای ایستا: فرایندهای iid، فرایند برنولی، نویز سفید
 - Ergodicity
 - بررسی سیستم‌ها با ورودی تصادفی
 - طیف توان (Power Spectrum)
- فرایند پواسن
 - معرفی فرایندهای arrival و renewal
 - خاصیت بدون حافظه بودن یک متغیر تصادفی
 - خواص stationary increment و independent increment فرایند پواسن
 - تابع چگالی احتمال S_n و تابع چگالی احتمال مشترک S_1, \dots, S_n
 - تابع جرمی احتمال متغیر تصادفی $N(t)$



- چند نکته درباره توزیع پواسن
- تعاریف دیگری برای فرایند پواسن
- ترکیب کردن و انشعاب کردن فرایندهای پواسن
- بردارهای گاوسی و فرایندهای گاوسی
 - متغیر تصادفی گاوسی
 - بردارهای تصادفی گاوسی
 - فرایندهای تصادفی گاوسی
 - چند مثال از فرایندهای گاوسی گسسته-زمان
 - قضیه‌ای در مورد ایستایی فرایندهای تصادفی گاوسی
- زنجیره‌های مارکف متناهی-حالت
 - تعریف زنجیره مارکف
 - نمایش‌های مختلف زنجیره مارکف
 - دسته‌بندی حالت‌های مختلف یک زنجیره مارکف
 - نمایش ماتریسی
 - حالت پایدار یک زنجیره مارکف
 - حالت پایدار با فرض $P > 0$
 - حالت پایدار برای زنجیره‌های ارگودیک
 - حالت پایدار برای ergodic unichains
 - حالت پایدار برای زنجیره‌های مارکف متناهی-حالت دلخواه
- تئوری تخمین
 - مقدمه
 - اصل کفایت (The Sufficiency Principle) و آمار کافی (Sufficient Statistic)
 - قضیه factorisation
 - آمار کافی کمین (Minimal Sufficient Statistics)
 - اصل likelihood
- تخمین‌گرهای نقطه‌ای
 - مقدمه
 - روش‌های یافتن تخمین‌گرها
 - روش گشتاورها
 - تخمین‌گرهای بیشینه likelihood
 - تخمین‌گرهای بیز
 - روش‌هایی برای ارزیابی تخمین‌گرها
 - خطای میانگین مربع
 - بهترین تخمین‌گر بدون بایاس
 - کران Cramer-Rao



- اطلاعات فیشر
- کفایت و بدون بایاس بودن
- قضیه Rao-Blackwell

مراجع

- Athanasios Papoulis and S. Unnikrishna Pillai, "Probability, Random Variables and Stochastic Processes," McGraw-Hill Europe, 4th edition, Jan., 2002.
- Robert G. Gallager, "Stochastic Processes: Theory for Applications," Cambridge University Press, 1st edition, Feb., 2014.
- George Casella and Roger L. Berger, "Statistical Inference," Wadsworth Press, 2nd edition, Jun., 2001

اساتید

- دکتر ربیعی
- دکتر رهبان



نظریه اطلاعات و کدینگ

Information Theory and Coding

اهداف درس

هدف از این درس، آشنایی دانش‌جویان با مفاهیم نظریه اطلاعات و کدینگ است.

ریز مواد

- مفهوم اطلاعات و انتروپی
 - آنترپی
 - اطلاعات متقابل
 - نامساوی‌های اطلاعات متقابل
 - AEP
- الگوریتم‌های فشرده‌سازی منبع
 - نامساوی کرفت
 - کد هافمن
 - کد اریتمتیک
 - کد لمپل-زیف
 - کد کردن اعداد طبیعی
- فشرده‌سازی چند منبعی
 - روش اسلپین و ولف
- آمار و تئوری اطلاعات
 - روش نوعی
 - فشرده‌سازی جهانی
 - قضیه سانوف
 - آزمون فرض
- ظرفیت کانال
 - مفهوم ظرفیت کانال
 - اثبات وجود کد برای نرخهای کمتر از ظرفیت
 - اثبات عدم وجود کد برای نرخهای بالاتر از ظرفیت
- کدینگ‌های خطی
 - کد همینگ
 - کد رید و سولمون
 - کد کانولوشنال
 - کد LDPC
- تئوری اطلاعات و یادگیری ماشین

مراجع



- MacKay, David JC. Information theory, inference, and learning algorithms. Vol. 7. Cambridge: Cambridge university press, 2003.
- Cover, Thomas M., and Joy A. Thomas. Elements of information theory. John Wiley & Sons, 2012

اساتید

- دکتر جعفری



مبانی علم داده

Foundations of Data Science

اهداف درس

هدف این درس آشنایی با مفاهیم و مبانی نظری علم داده است. این مبانی از جنبه‌های مختلفی همانند هندسه ابعاد بالا، کاهش بعد داده‌های با ابعاد بالا، قدم زنی تصادفی و زنجیره مارکوف، گراف‌های تصادفی همچنین در خصوص جنبه‌های مهم ارزیابی و تحلیل این سیستم‌ها بحث خواهد شد.

ریزمواد

۱. مقدمه
۲. هندسه با ابعاد بالا
 - هندسه با ابعاد بالا و ویژگی‌های آن
 - تولید داده در ابعاد بالا
 - توزیع گوسی در ابعاد بالا
 - تخمین توزیع برای داده‌های با ابعاد بالا
 - تعیین نزدیک‌ترین همسایه برای داده‌های با ابعاد بالا
۳. کاهش بعد برای داده‌های با ابعاد بالا
 - داده‌های با ابعاد بالا و مصیبت ابعاد
 - کاهش بعد داده‌های با ابعاد بالا با کمک روش تحلیل مؤلفه اصلی
 - کاهش بعد داده‌های با ابعاد بالا با کمک روش تجزیه مقادیر تکین
 - شیوه پیاده‌سازی روش‌های کاهش بعد برای داده‌های با ابعاد بالا
۴. زنجیره مارکوف و قدم زنی تصادفی
 - زنجیره مارکوف
 - قدم زنی تصادفی و همگرایی آن برای گراف‌های مختلف
 - قدم زنی تصادفی در فضاء اقلیدسی
 - شبکه وب و زنجیره مارکوف
 - روش‌های مختلف نمونه برداری
۵. گراف تصادفی و تحلیل شبکه‌های اجتماعی
 - گراف‌های تصادفی و خواص آنها
 - مدل‌های مختلف تولید گراف‌های تصادفی
 - مسایل مختلف در گراف‌های تصادفی و شیوه حل آنها
 - مدل‌های مختلف گراف‌های تصادفی مانند جهان کوچک
۶. مسایل همام سازی و هم تراز
۷. مدل سازی موضوعی
۸. روش‌های مختلف تجزیه ماتریس
۹. روش‌های طیفی در خوشه بندی و کاربردهای آن در شبکه‌های اجتماعی
۱۰. الگوریتم‌های تحلیل داده‌های از قبیل جویبارسازی، نمونه برداری و Sketching
۱۱. موضوعات دیگر



منابع

- Avrim Blum, John Hopcroft, and Ravindran Kannan, Foundations of Data Science, Cambridge University Press; 1st edition (March 12, 2020)

اساتید

- دکتر اعتصامی
- دکتر بیگی
- دکتر موقر



تحلیل داده‌ای متنی

Text Data Analysis

اهداف درس

هدف از این درس فراگیری تکنیک‌های استخراج اطلاعات از متون و توانایی به‌کارگیری ابزارهای مناسب برای این منظور است. برای رسیدن به این هدف مواردی از جمله ساخت مخزن اسناد، آماده‌سازی داده متنی، تحلیل زبانی، بازنمایی متون و شاخه‌های مختلف تحلیل داده متنی نظیر مدل‌سازی موضوعی، دسته‌بندی و خوشه‌بندی متون بررسی خواهند شد. همچنین تشخیص روش‌های تحلیلی مناسب برای یک مساله تحلیل متن از جمله اهداف این درس است. به علاوه ابزارهای تحلیل متن (از جمله Natural Language Toolkit) در حل مسایل مورد استفاده قرار خواهند گرفت.

ریزمواد

- ۱- جمع‌آوری داده
 - web scraping و استفاده از API
 - ساخت مخزن (corpus)
- ۲- تمیزسازی و آماده کردن داده
 - جداسازی واحدها (Tokenization)
 - ریشه‌یابی (Stemming) و بن‌واژه‌سازی (Lemmatization)
 - پیدا کردن عبارات منظم (Regular Expressions)
 - نرمال‌سازی (Normalization)
 - تصحیح املایی (Spelling Correction)
- ۳- شکل‌های زبانی کلمات
 - فرم‌های مختلف "n-ym"، نظیر هم‌معنا (synonym)، هم‌شکل یا هم‌آوا (homonym)، جزء واژه (meronym) و ...
 - محدود یا گسترده کردن از طریق زیرشمول (hyponyms) و فراشمول (hypernyms)
- ۴- دیکشنری
 - هستان‌شناسی (ontology)
 - رده بندی (taxonomy)
- ۵- کشف روابط Syntagmatic
 - آن‌تروپی، آن‌تروپی شرطی و اطلاعات متقابل (mutual information)
- ۶- تحلیل زبانی
 - تجزیه (Parsing)
 - مضمون کلمه (word sense)
 - برچسب‌زنی ادات سخن (Part-of-Speech) یا POS
 - ابهام زدایی مضمون کلمه (Word Sense Disambiguation)
 - تشخیص موجودیت‌های نامدار (Named Entity Recognition)
- ۷- بازنمایی متون
 - نمایش کیسه لغات (Bag of Words)، فرکانس واژه (Term Frequency)، فرکانس واژه-معکوس فرکانس سند (TF-IDF)
 - N-گرام
 - روش‌های مبتنی بر دیکشنری
 - مدل‌سازی زبانی (language modeling)
 - تعبیه کلمات (word embedding)
- ۸- مدل‌سازی موضوعی (Topic Modeling)
 - روش‌های احتمالاتی مدل‌سازی موضوعی
 - تحلیل معنایی نهان احتمالاتی (Probabilistic Latent Semantic Analysis)
 - تخصیص دیریکله نهان (Latent Dirichlet Allocation)



- مدل‌سازی موضوعی ساختاری (Structural Topic Modeling)

۹- خوشه‌بندی متون

- رویکرد احتمالاتی
- روشهای مبتنی بر معیار شباهت
- ارزیابی (evaluation)

۱۰- دسته‌بندی متون

- دسته‌بند احتمالاتی مولد (Generative)
- دسته‌بند تمایزی (Discriminative)
- روشهای دسته‌بندی
- ارزیابی

۱۱- تحلیل احساس (sentiment analysis) و نظرکاوی (opinion mining)

- دسته‌بندی احساس
- رگرسیون لاجستیک ترتیبی (ordinal)

۱۲- تکنیک‌های بصری‌سازی (visualization) در تحلیل متون

۱۳- حکومت داده (Governance) و اخلاق (Ethics) در تحلیل متن

شیوه ارزیابی

- امتحان میان ترم ۲۰٪
- امتحان پایان ترم ۳۰٪
- تمرین ۳۵٪
- پروژه ۱۵٪

منابع

- Grimmer, Justin, and Brandon Stewart. 2013. Text as Data. The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis 21: 267-297.
- Jurafsky, Daniel, and James H. Martin. 2018. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. 3rd edition.

اساتید

- دکتر سلیمانی
- دکتر عسگری



شبکه‌های اجتماعی و اقتصادی

Social and Economic Networks

اهداف

این درس برای کسب یک دید سطح بالا از مدل‌ها و تکنیک‌هایی که برای مطالعه و تحلیل شبکه‌های اجتماعی و اقتصادی به کار برده می‌شوند ارائه شده است. شبکه‌های اجتماعی و اقتصادی چند مشخصه اصلی دارند:

- ۱- شبکه پیچیده (Complex Network) هستند و بسیاری از مشخصه‌های اصلی آن را دارند.
- ۲- در اثر تعامل عوامل باهوش و خودخواه ایجاد می‌شوند.
- ۳- تصادفی هستند و در بسیاری موارد نمی‌توان به صورت قطعی ویژگی‌های آن‌ها را مشخص کرد.

بنابراین مدل‌ها و تکنیک‌هایی که در این درس با آن‌ها سر و کار داریم از هر سه دسته بالا نشئت گرفته و در بسیاری موارد نیز با هم ترکیب می‌شوند. به همین منظور در این درس با مدل‌هایی سر و کار داریم که با استفاده نظریه احتمال، نظریه بازی و نظریه گراف و ترکیبیات ساخته می‌شوند.

ریزموارد

- پیش‌زمینه‌ها و مبانی تحلیل شبکه
 - مقدمه- مثال‌هایی از شبکه‌های اجتماعی و اقتصادی و کاربردهای آن در علوم مختلف
 - نمایش و اندازه‌گیری شبکه‌ها- بعضی مقدمات از تئوری گراف- مشخصات عمومی شبکه‌ها
 - مشاهدات تجربی در مورد شبکه‌های اجتماعی و اقتصادی
- مدل‌های شکل‌گیری شبکه‌ها
 - گراف‌های تصادفی
 - مدل‌های تصادفی رشد شبکه
 - مدل‌های استراتژیک تشکیل شبکه
- انتشار در شبکه‌ها
- یادگیری در شبکه‌ها
- بازی‌های تحت شبکه
- بازارهای شبکه‌ای
- نظریه‌ی بازی‌های تعاونی و قوانین تخصیص منابع در شبکه‌ها
- اندازه‌گیری تعاملات اجتماعی و اقتصادی

ارزیابی

- آزمون: ۶۰ درصد
- تمرین: ۲۰ درصد
- سمینار: ۲۰ درصد

مراجع اصلی

- Jackson, Matthew O. Social and economic networks. Princeton university press, 2010.
- Easley, David, and Jon Kleinberg. Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press, 2010.



اساتید

• دکتر فضلی



تحلیل داده‌های حجیم زیستی

Large-Scale Biological Data Analysis

ریزموارد

- آماده سازی کتابخانه
 - تکنیکهای موجود
 - محاسبه پیچیدگی کتابخانه
- توالی یابی دنا
 - برهم نهی تمام ژنوم
 - برهم نهی تمام اگزون
 - phase کردن
- توالی یابی رنا
 - نرمالیزه کردن بیان
 - یافتن ژنهای تفریقی
 - یافتن شبکه‌های تنظیمی بیان ژن
- تحلیل برهم کنش دنا و پروتئین
 - روشهای نرمالیزه کردن داده
 - ChIP-Seq
 - ATAC-Seq
 - یافتن متیلفها
- روشهای تک سلولی
 - نرمالیزه کردن داده های تک سلولی
 - دسته بندی داده‌های تک سلولی
 - پر کردن داده‌های تک سلولی
 - استنتاج از داده‌های تک سلولی

مراجع اصلی

- Ye, Shui Qing, ed. Big data analysis for bioinformatics and biomedical discoveries. CRC Press, 2016.

اساتید

- دکتر مطهری



علم داده در مهندسی نرم افزار

Data Science in Software Engineering

اهداف درس

مخزن‌های نرم‌افزاری مانند مخازن کد، مخازن خطاها، سایت‌های پرسش و پاسخ نرم‌افزاری، فروشگاه‌های برنامه‌های همراه، و ... حاوی داده‌های ارزشمندی راجع به یک نرم‌افزار و تاریخچه آن می‌باشند. تحلیل این داده‌ها به پژوهشگران حوزه مهندسی نرم‌افزار این امکان را می‌دهد تا به صورت تجربی درک کنند که چه روش‌هایی اغلب در عمل مورد استفاده مهندسين نرم‌افزار قرار گرفته است. به خود توسعه‌دهندگان و مهندسين نرم‌افزار نیز این امکان را می‌دهد تا بهتر بتوانند یک سیستم نرم‌افزاری پیچیده را مدیریت و نگهداری کنند. با گسترش روزافزون استفاده از تکنیک‌های علم داده و هوش مصنوعی مانند یادگیری ماشین و یادگیری عمیق، استفاده از این تکنیک‌ها در تحلیل داده‌های موجود در مخازن نرم‌افزاری و مهندسی نرم‌افزار نیز کاربردهای فراوانی یافته است. به عنوان نمونه، می‌توان به رفع خودکار خطاهای نرم‌افزاری، خلاصه‌سازی مصنوعات نرم‌افزاری، و تحلیل نظرات راجع به برنامه‌های همراه اشاره نمود. هدف از این درس، آشناسازی دانشجویان با استفاده از این تکنیک‌ها در عمل به منظور حل مسائل مهندسی نرم‌افزار می‌باشد.

ریز مواد

- معرفی علم داده و کلان داده‌ها.
- مقدمه‌ای بر کاربردهای علم داده در مهندسی نرم‌افزار.
- آشنایی با کتابخانه‌های مرتبط در زبان برنامه‌نویسی پایتون برای تحلیل داده‌ها.
- آشنایی با تکنیک‌های یادگیری ماشینی
- درخت‌های تصمیم‌گیری.
- شبکه‌های عصبی
- یادگیری ژرف
- خوشه‌بندی.
- پردازش زبان طبیعی.
- کاربردهای تحلیل متن در مهندسی نرم‌افزار.
- کاربردهای LDA در استخراج موضوعات از مصنوعات نرم‌افزاری.
- بازبینی کد.
- تحلیل داده‌های IDE.
- خلاصه‌سازی مصنوعات نرم‌افزاری.
- تحلیل فروشگاه‌های برنامه‌های همراه.
- تحلیل انرژی در برنامه‌های همراه.
- استفاده از زیرساخت BOA برای تحلیل مخازن نرم‌افزاری.

روش ارزیابی



- آزمون میان ترم: ۲۵٪ کل نمره

- آزمون پایان ترم: ۳۰٪ کل نمره

- ارائه مقاله علمی: هر دانشجو موظف به ارائه حداقل یک مقاله علمی می‌باشد که در یکی از بهترین کنفرانس‌ها یا مجلات مرتبط به درس در سال‌های اخیر به چاپ رسیده باشد. (۱۵٪ کل نمره)

- پروژه و گزارش پژوهشی: موضوع پروژه پژوهشی قبل از آزمون نیمسال تعیین می‌شود. دانشجو کار پژوهش را با کمک استاد درس آغاز نموده و پس از انجام پروژه، نتیجه پژوهش را در قالب گزارش ارائه می‌دهد. (۳۰٪ کل نمره)

مراجع

- آخرین مقالات مروری (Survey) ارائه شده در زمینه سرفصل‌های مطالب در برترین مجلات و کنفرانس‌های مهندسی نرم‌افزار.
- Joel Grus, Data Science from Scratch, O'Reilly, 2019.
- Christian Bird, Tim Menzies, and Thomas Zimmermann, The Art and Science of Analyzing Software Data, Morgan Kaufmann, 2015.
- Tim Menzies, Laurie Williams, and Thomas Zimmermann, Perspectives on Data Science for Software Engineering, Morgan Kaufmann, First edition, 2016.

اساتید

- دکتر حیدرنوری



تحلیل داده‌های مالی

Financial Data Analysis

اهداف درس

در این درس دانشجویان با تحلیل داده‌های مالی آشنا می‌شوند. در واقع می‌آموزند که چگونه مدل‌ها و فرآیندهای مالی را پیاده‌سازی کنند و با استفاده از آن‌ها به پیش‌بینی بازارها بپردازند.

ریز مواد

- قیمت و مفهوم آن
- بازدهی فردی اوراق بهادار
- بازدهی سبدها
- ریسک
- مدل‌های فاکتوری
- متریک‌های تعدیل‌شده با ریسک بازدهی سبد
- بهینه‌سازی میانگین-واریانس مارکوویتز
- اوراق درآمد ثابت
- اوراق مشتقه (option)
- مدل‌های یادگیری ماشین در تحلیل داده‌های مالی
 - مدل‌سازی احتمالاتی
 - رگرسیون بیزی و فرآیند گاوسی
 - شبکه‌های عصبی Feed Forward
 - تبیین‌پذیری مدل یادگیری ماشین
 - مدل‌سازی دنباله‌ها
 - مدل‌سازی احتمالاتی دنباله‌ها
 - شبکه‌های عصبی پیشرفته
 - یادگیری ماشین تقویتی
 - یادگیری ماشین تقویتی معکوس و یادگیری تقلیدی

روش ارزیابی

- آزمون میان ترم: ۲۵٪ کل نمره
- آزمون پایان ترم: ۳۰٪ کل نمره
- ارائه مقاله علمی: هر دانشجو موظف به ارائه حداقل یک مقاله علمی می‌باشد که در یکی از بهترین کنفرانس‌ها یا مجلات مرتبط به درس در سال‌های اخیر به چاپ رسیده باشد. (۱۵٪ کل نمره)



- پروژه و گزارش پژوهشی: موضوع پروژه پژوهشی قبل از آزمون نیمسال تعیین می‌شود. دانشجو کار پژوهش را با کمک استاد درس آغاز نموده و پس از انجام پروژه، نتیجه پژوهش را در قالب گزارش ارائه می‌دهد. (۳۰٪ کل نمره)

مراجع

- Ang, Clifford S. Analyzing financial data and implementing financial models using R. Springer, 2015.
- Dixon, Matthew F., Igor Halperin, and Paul Bilokon. Machine Learning in Finance. Springer International Publishing, 2020.

اساتید

- دکتر فضلی
- دکتر رهبان



تحلیل داده‌های سلامت

Health Data Analysis

اهداف درس

آشنایی با مفاهیم، روش‌ها، زمینه‌ها و کاربرد تحلیل داده در حوزه سلامت

ریزموارد

- داده‌های سلامت، ماهیت، رفتار، چالش‌ها و فرصت‌ها
 - آشنایی عمومی با سیستم سلامت (چشم‌انداز، مأموریت، مولفه‌ها و کارکردها، محدوده، فرصت‌ها، تهدیدات، روندها و ...)
 - آشنایی با حوزه‌های داده‌ای سیستم سلامت، ویژگی‌ها و ملاحظات مورد نیاز داده‌های سلامت
 - آشنایی با ساختار اطلاعاتی پزشکی، استانداردهای کدگذاری، ترمینولوژی‌ها و کاربرد آنها
 - آشنایی با پلتفرم‌های ذخیره‌سازی و انتقال داده‌های سلامت در ایران و جهان
 - آشنایی با الگوریتم‌ها، روش‌ها، پلتفرم‌ها، ابزار و فناوری‌های مرسوم تحلیل داده در سیستم سلامت
- آشنایی با تحلیل داده‌های سلامت به تفکیک حوزه‌های اصلی (براساس مرور مطالعات موردی)
 - کاربرد تحلیل داده در تحلیل و بهبود کارایی سیستم سلامت محلی، ملی و جهانی
 - کاربرد تحلیل داده در خصوص تحلیل و بهبود کارایی بیمه‌های سلامت
 - تحلیل داده در مدیریت همه‌گیری
 - تحلیل داده در پیشگیری از بیماری‌ها غیرواگیر
 - تحلیل داده در مطالعات پزشکی (ریشه‌یابی بیماری‌ها و عوامل موثر بر آنها، تولید دارو و ...)
 - تحلیل داده در طراحی و توسعه فناوری‌های پزشکی و پیراپزشکی
 - تحلیل داده در بررسی عوامل اجتماعی سلامت (SDH)
- کارگاه تحلیل داده‌های سلامت (ارائه همزمان با بخش‌های اول و دوم)
 - معرفی متدولوژی برای اجرای پروژه‌های تحلیل داده
 - راهبری پروژه‌های تحلیل داده تیم‌های فراگیران
 - ارائه پروژه‌ها توسط فراگیران و ارزیابی آنها

مراجع

- Trevor L. Strome, Healthcare Analytics for Quality and Performance Improvement, Wiley; 1st edition (October 2, 2013)
- Chandan K. Reddy, et al. Healthcare Data Analytics (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series Book 36) , Chapman and Hall/CRC; 1st edition (June 23, 2015)

اساتید

- دکتر مطهری
- دکتر شریفی



تحلیل داده‌های کسب و کارها

Data Analysis in Business

اهداف درس

داده‌های کسب و کاری و تجزیه تحلیل آن‌ها به عنوان یک شاخه کاربردی و مبتنی بر علم داده و پیشرفت‌های آن است. که با توجه به تاثیرگذاری آن به عنوان یک محرک مهم در علم داده مطرح می‌باشد. مهم ترین دارایی هر کسب و کار داده‌های آن است که با بهره‌گیری از آن و اتخاذ استراتژی‌های درست در این زمینه می‌تواند به دستیابی به اهداف خود با بهره‌وری بالاتری عمل کند. در این درس دانشجویان با بهره‌گیری از مفاهیم حوزه علم داده و کاربردهای آن در کسب کار را با کاربردها و ابزارهای مفید در این زمینه آشنا شده و جایگاه استراتژی مختلف داده محور را در تصمیم‌گیری‌های کسب و کاری فرا می‌گیرند.

ریز مواد

- تفکر تحلیلی داده چیست؟
- مشکلات کسب و کاری و راه حل‌های علم داده
- علم داده و استراتژی‌های کسب و کار
- معرفی داده‌های کسب و کاری و روش‌های شناسایی آن‌ها
- کاوش داده و کاربرد آن در کسب و کارها
- تجزیه و تحلیل پیشبینی کننده و مدلسازی داده‌های دامنه‌های کاری
- ارزیابی مدل‌های تصمیم‌گیری داده محور در کسب کارها
- تکنولوژی‌های نو و استفاده از آنها در مهندسی تحلیلی (IoT, Blockchain,...)
- ابزارهای کاربردی در حوزه تحلیل کسب و کار

مراجع

- Joel Grus, Data Science from Scratch, O'Reilly, 2019.
- Data Science For Business: What You Need To Know About Data Mining And Data-Analytic Thinking" by Foster Provost & Tom Fawcett
- Business Intelligence Roadmap: The Complete Project Lifecycle For Decision-Support Applications" by Larissa T. Moss & Shaku Atre
- Performance Dashboards – Measuring, Monitoring, And Managing Your Business" by Wayne Eckerson

اساتید

- دکتر حبیبی
- دکتر معاون
- دکتر فضلی



سیستم‌های توصیه‌گر

Recommender Systems

اهداف درس

هدف این درس آشنایی با مفاهیم، مسائل و تکنیک‌های زمینه سیستم‌های توصیه‌گر (recommender systems) و نقش و کاربرد عملی آن‌ها است. انواع سیستم‌های توصیه‌گر نظیر پالایش مبتنی بر محتوا (content-based) و پالایش همکارانه (collaborative filtering) و همچنین حوزه‌های کاربردی این سیستم‌ها معرفی خواهد شد. همچنین در خصوص جنبه‌های مهم ارزیابی و تحلیل این سیستم‌ها بحث خواهد شد.

ریزمواد

۱- معرفی سیستم‌های توصیه‌گر و کاربردهای آن

- مقدمه
- ترجیحات کاربر و امتیازدهی
- پیش‌بینی ترجیح یا امتیاز
- ۲- انواع رویکردها و مدل‌های سیستم‌های توصیه‌گر
 - روش‌های تحلیل سبد خرید (Market Basket)
 - رویکرد مبتنی بر محتوا: روش‌های مبتنی بر همسایگی، مبتنی بر شباهت و مبتنی بر مدل
 - رویکرد پالایش همکارانه: روش‌های مبتنی بر همسایگی، روش‌های مبتنی بر مدل، تجزیه ماتریسی (SVD، منظم‌سازی، تجزیه ماتریسی محدودیت‌دار، تجزیه ماتریس با رویکرد احتمالاتی)، روش‌های یادگیری ژرف
 - رویکرد ترکیبی (hybrid): انواع روش‌های ترکیب، ادغام و وارد کردن اطلاعات جانبی (side information)
 - ماشین‌های تجزیه (factorization machines)
 - توصیه آگاه از زمینه (context-aware)
 - مدل‌سازی زمانی (temporal) و پویایی زمانی در سیستم‌های توصیه‌گر

۳- ارزیابی سیستم‌های توصیه‌گر

- متدولوژی ارزیابی
- انواع آزمایش‌ها
- معیارهای ارزیابی

۴- تجارب کاربر (user experiences)

- فاکتورهای انسانی در طراحی سیستم
- درک رفتار کاربر
- ارزیابی کاربر-محور (user-centered)

۵- کاربردها

- وب‌سایتهای تجارت الکترونیک (E-commerce)
- شبکه‌های اجتماعی (social networks)
- دیگر زمینه‌ها نظیر اخبار، سلامت و ...
- مطالعه موردی (Amazon, Netflix, Google News, YouTube)

۶- دیگر جنبه‌های عملی

- توضیح (explanation) توصیه‌ها و اعتماد (Trust)
- مقیاس‌پذیری (scalability)
- اعمال نفوذ (manipulation)، حریم خصوصی (Privacy) و حملات (attacks)



شیوه ارزیابی

٪۲۰	امتحان میان ترم
٪۳۰	امتحان پایان ترم
٪۲۰	تمرین
٪۳۰	پروژه

منابع

- Jannach Dietmar, Recommender Systems: An Introduction, 1st Edition, New York: Cambridge University Press, 2011.

اساتید

- دکتر بیگی
- دکتر فضلی