جمهوری اسلامی ایران

وزارت علوم، تحقیقات و فناوری

شورای عالی برنامه ریزی آموزشی

برنامه درسی

دوره: کارشناسی ارشد

رشته: ریاضی کاربردی

گرایش: علوم داده

گروه برنامه ریزی علوم ریاضی

مصوب جلسه شماره ۸۸۷ مورخ ۱۳۹۶/۰۹/۰۴ شورای عالی برنامه ریزی آموزشی

بسم الله الرّحمن الرّحیم

عنوان برنامه: دوره کارشناسی ارشد رشته ریاضی کاربردی گرایش علوم داده
تدوین شده توسط گروه برنامه ریزی علوم ریاضی

۱ـ عنوان برنامه درسی فوق در جلسه شماره ۸۸۷ مورخ ۱۳۹۶/۰۹/۰۴ شورای عالی برنامه ریزی آموزشی به تصویب رسیده است.

۲ـ برنامه درسی مذکور در سه فصل: مشخصات کلی، جدول واحد های درسی و سرفصل دروس تنظیم شده وبرای تمامی دانشگاه ها ومؤسسه های آموزش عالی و پژوهشی کشور که طبق مقررات مصوب وزارت علوم، تحقیقات و فناوری فعالیت می کنند، برای اجرا ابلاغ می شود.

۳ـ این برنامه درسی از شروع سال تحصیلی ۱۳۹۸ـ۱۳۹۷ به مدت ۵ سال قابل اجراست و پس از آن نیازمند بازنگری می باشد.

عبدالرحیم نوه‌ابراهیم
دبیر شورای عالی برنامه‌ریزی آموزشی

مجتبی شریعتی نیاسر
نایب رئیس شورای عالی برنامه‌ریزی آموزشی

برنامه و سرفصل درس‌های کارشناسی ارشد ریاضی کاربردی

گرایش آنالیزعددی (مصوب شده)

گرایش بهینه سازی (مصوب شده)

گرایش رمز و کد (مصوب شده)

گرایش معادلات دیفرانسیل و سیستم های دینامیکی(مصوب شده)

گرایش ریاضی مالی(مصوب شده)

گرایش علوم داده (مصوب شده)

گرایش زیست ریاضی (مصوب شده)

مقررات عمومی برنامه کارشناسی ارشد ریاضی کاربردی

کلیه دانشگاه‌هایی که قبلاً مجوز اجرای رشته را به صورت کلی یا در گرایش‌های مختلف اخذ کرده و با کد رشته محل‌های مربوطه به پذیرش دانشجو در این رشته می‌پرداخته‌اند کماکان می‌توانند با پذیرش دانشجو در همان کد رشته محل‌ها نسبت به پذیرش دانشجو اقدام کنند. این دانشگاه‌ها می‌توانند با پذیرش دانشجو در کد رشته محل "ریاضیات و کاربردها" به صورت تجمعی اقدام کرده و هر یک از دانشجویان پذیرفته شده را با در نظر گرفتن تخصص اعضای هیأت علمی و امکانات موجود در هر یک از گرایش های اخذ شده این رشته با رعایت مقررات برنامه گرایش مربوطه در برنامه فعلی با قید گرایش دانش‌آموخته کنند.

چنانچه دانشگاهی تمایل داشته باشد در رشته ریاضیات و کاربردها و در یکی از گرایش‌های برنامه که قبلاً مجوز اجرای آن را نداشته است، با کد رشته محل مجزا دانشجو پذیرد، لازم است که قبلاً نسبت به اخذ مجوز اجرا اقدام کرده و فقط در صورت احراز شرایط و پس از اخذ مجوز از وزارت عتف نسبت به پذیرش دانشجو با کد رشته محل مختص گرایش مربوطه اقدام کنندطول دروه و شکل نظام

دوره کارشناسی ارشد ریاضی کاربردی مطابق با آیین‌نامه جاری دوره‌ی کارشناسی ارشد وزارت عتف است.

تعداد واحدهای دوره

تعداد واحدهای درسی دوره کارشناسی ارشد ریاضی کاربردی و ۲۹ به قرار زیر است:

درس‌های الزامی: ۹ واحد، شامل درس(های) اصلی گرایش یا زیر گرایش و یک درس از دروس اصلی گرایش ها یا زیر گرایش های دیگر علوم ریاضی با نظر استاد راهنما یا دانشکده

درس‌های تخصصی- اختیاری: ۱۲ واحد، شامل دست کم سه درس از جدول درس های تخصصی- اختیاری و حداکثر یک درس با نظر استاد راهنما و تأیید گروه از درس‌های اختیاری یکی از دوره‌های کارشناسی ارشد مرتبط.

سمینار: ۲ واحد

پایان‌نامه: ۶ واحد

اخذ درس سمینار و پایان نامه در نیمسال اول تحصیل مجاز نیست. برای اخذ درس سمینار نیاز به گذراندن ۹ واحد درسی و برای اخذ پایان‌نامه گذراندن دست کم ۱۲ واحد (که شامل درس های الزامی می باشد) و اجازه گروه ضروری است.

دانشجویان دوره کارشناسی ارشد ریاضی کاربردی با اخذ دست کم ۶ واحد تمام وقت محسوب می شوند.

با توجه به پایه‌ای بودن دروس الزامی گرایش‌ها وتنوع طیف دانشجویان در دانشگاه های مختلف، دانشکده مجری می‌تواند در صورت نیاز تعداد واحدهای این دروس را از ۳ به ۴ افزایش دهد. سقف واحدهای این دوره با این تغییر از ۲۹ به حداکثر ۳۲ افزایش خواهد یافت.

گروه های مجری می توانند درس های جدیدی را به عنوان درس اختیاری مطابق با روال جاری دانشگاه مصوب و ارایه دهند.

فصل اول

مشخصات دوره کارشناسی ارشد ریاضی کاربردی- گرایش علوم داده

مقدمه

«علوم داده» موضوعی میان‌رشته‌ای است که ریاضیات، آمار و علوم کامپیوتر در آن نقش تعیین‌کننده‌ای ایفا می‌کنند. در حال حاضر علوم داده از فعال‌ترین زمینه‌هایی است که به طور مستقیم با فن‌آوری‌های نوین و پیشرفته ارتباط داشته، و به همین سبب با سرعت شگرفی رو به تکامل و پیشرفت است. اهمیت داشتن متخصصینی در این موضوع که بتوانند به ابعاد نظری این زمینه تسلط داشته و به تبع آن در پیشرفت علمی کشور موثر باشند، در راستای برنامه‌های توسعه‌ای کشور و اهداف سند چشم‌انداز الزامی به نظر می‌رسد.

تعریف

دورهٔ کارشناسی ارشد ریاضی کاربردی- گرایش علوم داده برنامه‌ای در مقطع کارشناسی ارشد است که با جذب دانش‌آموختگان زمینه‌های مختلف علوم ریاضی(آمار، ریاضی و علوم کامپیوتر) و با تاکید بر جنبه‌های اصیل ریاضی و ابعاد نظری علوم داده امکان تربیت نیروهای متخصص در این زمینه را فراهم می آورد.

اهداف

هدف برنامه کارشناسی ارشد ریاضی کاربردی- گرایش علوم داده ایجاد فضای آموزشی مناسب برای تربیت دانش‌آموختگانی است که به ابعاد نظری علوم داده تسلط داشته باشند و بتوانند در کنار مهندسین در کاربردهای مختلف این علم موثر بوده و همچنین توانایی نوآوری و تولید علم در این زمینه تخصصی را دارا باشند.

ضرورت و اهمیت

نیاز به پردازش و تحلیل مه‌داده‌ها از الزامات عصر حاضر و تربیت متخصصینی که بتوانند در این زمینه راهگشا باشند اجتناب ناپذیر است. از طرف دیگر پردازش و تحلیل مه‌داده‌هایی نظیر داده‌های هواشناسی، داده‌های زیستی یا داده‌های مالی به دلیل اندازه و معمولا بعد بالا امر ساده ای نیست. چنانکه روش‌های موجود بر مبانی نظری ژرفی در علوم ریاضی استوار هستند. لذا، پیشبرد این حوزه و حتی فعالیت جدی در ابعاد کاربردی آن به دانش نظری ژرف نیازمند است. تربیت نیروی انسانی متخصص در این موضوع نیاز به برنامه‌های آموزشی و پژوهشی منسجمی دارد.

نقش و توانایی

دانش‌آموختگان برنامه کارشناسی ارشد ریاضی کاربردی- گرایش علوم داده می‌توانند به فعالیت در گروه‌های تحقیقاتی یا حرفه‌ای مرتبط با موضوع علوم داده در صنعت یا مراکز تحقیقاتی بپردازند. همچنین امکان ادامه تحصیل در مقطع دکتری در هر یک از رشته‌های علوم ریاضی برای دانش آموختگان این برنامه وجود دارد.

کلیات برنامه

- گذراندن درس‌های ۱۰۱ و ۱۰۲ الزامی است.
- گذراندن حداقل ۳ درس از جدول ۲ الزامی است.
- مابقی دروس شامل حداقل یک درس ۳ واحدی اختیاری، ۲ واحد سمینار و ۶ واحد پایان‌نامه با نظر استاد راهنما و تایید نهایی گروه مجری تعیین می‌شود.
- مابقی مقررات بر اساس آیین نامه آموزشی دوره‌های کارشناسی ارشد ناپیوسته خواهد بود.

عنوان دوره : کارشناسی ارشد ریاضی کاربردی- گرایش علوم داده

پیش نیاز ورود :

دارا بودن مدرک کارشناسی در یکی از رشته‌های علوم ریاضی (علوم کامپیوتر، ریاضیات و کاربردها یا آمار و کاربردها)، مهندسی برق، مهندسی کامپیوتر یا یکی از رشته‌های مرتبط

مواد آزمون ورودی(کنکور): دروس ریاضیات عمومی، مبانی ماتریس‌ها و جبرخطی، احتمال، آمارو احتمال مهندسی

فصل دوم

جدول دروس دوره کارشناسی ارشد ریاضی کاربردی- گرایش علوم داده

جدول شماره۱:درس‌های الزامی(گرایش علوم داده)

شماره درس	نام درس	تعداد واحد
۱۰۱	مبانی ریاضی علوم داده	۳
۱۰۲	الگوریتم‌های علوم داده	۳
*	*	۳

*یک درس از دروس اصلی گرایش‌ها یا زیر گرایش‌های دیگر علوم ریاضی با نظرگروه یا دانشکده.

جدول شماره ۲- درس‌های تخصصی – انتخابی(گرایش علوم داده)

شماره درس	نام درس	تعداد واحد	پیش‌نیاز یا هم‌نیاز(ها)	
			نام درس	شماره درس
۲۰۱	بهینه‌سازی در علوم داده	۳		
۲۰۲	روش‌های عددی در علوم داده	۳		
۲۰۳	روش‌های توپولوژی در علوم داده	۳		
۲۰۴	روش‌های هندسی در علوم داده	۳		
۲۰۵	نظریه یادگیری	۳		
۲۰۶	یادگیری آماری	۳		
۲۰۷	مدل‌سازی و تحلیل داده‌های با بُعد بالا	۳		
۲۰۸	سری‌های زمانی	۳		
۲۰۹	مدل‌های گرافی احتمالاتی	۳		
۲۱۰	داده‌کاوی	۳		
۲۱۱	دیداری‌سازی داده‌ها	۳		
۲۱۲	تشخیص داده‌های دورافتاده	۳		
۲۱۳	مدل‌سازی و پردازش مه‌داده‌ها	۳		
۲۱۴	گراف‌کاوی	۳		
۲۱۵	متن‌کاوی و وب‌کاوی	۳		
۲۱۶	یادگیری ژرف	۳		
۲۱۷	سنجش فشرده	۳		
۲۱۸	مباحث ویژه در علوم داده	۳		

دانشجو موظف است دست کم ۶ واحد از درس های جدول شماره ۲ را اختیار کند.

دانشجو با نظر گروه حداکثر یک درس از درس های دوره های کارشناسی ارشد دیگر علوم ریاضی را لازم است اختیار کند.

فصل سوم

سرفصل دروس دوره کارشناسی ارشد ریاضی کاربردی- گرایش علوم داده

عنوان درس	فارسی	مبانی ریاضی علوم داده									
	انگلیسی	Mathematical Foundations of Data Science									
نوع واحد								تعداد واحد	تعداد ساعت	دروس پیش‌نیاز	
پایه		اصلی		تخصصی		اختیاری		۳	۴۸		
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: ندارد				نیاز به اجرای پروژه عملی: ندارد							

هدف:

هدف این درس آشنایی اولیه با گستره وسیعی از مباحث ریاضیات، امار و علوم کامپیوتر نظری است که در فهم بنیادین ما از داده‌ها و چگونگی نمایش، پردازش و تحلیل آن‌ها مفید و موثر هستند. در بسیاری از موارد عنوان یک سرفصل خود موضوع درسی مستقل است که با عمق و وسعت بیشتر قابل مطالعه و پی‌گیری است.

سرفصل درس:

- فضاهای با بُعد بالا (قانون اعداد بزرگ، هندسه ابعاد بالا، تجمع اندازه، تولید نقطه تصادفی در گوی واحد، توزیع گاوسی در ابعاد بالا، تصویرهای تصادفی و قضیه مسطح‌سازی)
- تجزیه مقادیر تکین و بهترین زیرفضای تقریب‌زننده (تجزیه مقادیر تکین، بردارهای تکین و بردارهای ویژه، روش توان برای محاسبه تجزیه مقادیر تکین، کاربردها: تحلیل مولفه‌های اصلی، خوشه‌بندی ترکیب توزیع‌های گاوسی، رتبه‌بندی اسناد، بهینه‌سازی گسسته)
- گراف‌های تصادفی (مدل اردوش-رینی، خوشه بزرگ و گذر فاز، فرآیندهای شاخه‌ای، پیدایش دورها و همبندی کامل، مدل‌های رشد غیر همگن، مدل‌های Small World)
- قدم زدن تصادفی و زنجیرهای مارکُفی (توزیع مانا، مونت‌کارلو با زنجیرهای مارکُفی: روش متروپولیس-هستینگز، نمونه‌برداری گیبس، قدم زدن تصادفی بر گراف‌های بدون جهت، قدم زدن تصادفی در فضای اقلیدسی، شبکه اینترنت به عنوان زنجیر مارکُفی)
- یادگیری ماشین (یادگیری دسته‌ای، تیغ اُکام و همگرایی یکنواخت، الگوریتم پرسپترون، الگوریتم تصادفی حرکت در جهت گرادیان، یادگیری برخط، ماشین بردار پشتیبان، جداکننده‌های غیر خطی و توابع هسته، یادگیری قوی و ضعیف، تقویت، بُعد (VC)
- خوشه‌بندی (k-میانگین‌ها، خوشه‌بندی طیفی، خوشه‌بندی بازگشتی به کمک برش‌های تُنُک، زیرماتریس‌های چگال، روش‌های مبتنی بر جریان، یافتن خوشه موضعی، اصول موضوع خوشه‌بندی)
- مدل‌های مارکُفی نهان، مدل‌های گرافی و انتشار باور (مدل‌های مارکُفی نهان، مدل‌های گرافی و انتشار باور، مدل‌های بیزی یا مدل‌های باور، میدان‌های مارکُفی، الگوریتم درختی، انتقال پیام)
- موضوعات دیگر (رتبه‌بندی، رای‌گیری، بردارهای تُنُک و نمونه‌برداری فشرده، ...)

منابع:

[۱] A. Blum, J. Hopcroft, R. Kannan, Foundations of Data Science, Microsoft Research, https://www.cs.cornell.edu/jeh/book.pdf

عنوان درس	فارسی	الگوریتم‌های علوم داده										
	انگلیسی	Algorithms for Data Science										
نوع واحد										تعداد واحد	تعداد ساعت	دروس پیش‌نیاز
پایه		اصلی		تخصصی		اختیاری				۳	۴۸	
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی					
حل تمرین: ندارد عملی: ندارد						نیاز به اجرای پروژه						

هدف:

هدف این درس آشنایی با الگوریتم‌های مهم در علوم داده است که ما را قادر می‌سازد تا اطلاعات مفید را در کمترین زمان و با کمترین خطا از داده‌ها به دست آوریم. این الگوریتم‌ها به چند دسته مهم تقسیم می‌شوند: الگوریتم‌هایی برای کاهش بُعد و پیچیدگی داده‌ها بدون از دست دادن اطلاعات اساسی آن‌ها، الگوریتم‌هایی برای نمایش مناسب داده‌ها، الگوریتم‌هایی برای پیش‌بینی رفتار داده‌ها و الگوریتم‌هایی برای رتبه‌بندی.

سرفصل درس:

- کاهش بُعد داده‌ها (الگوریتم‌هایی برای به دست آوردن نگاشتی از مجموعه داده‌های با بُعد زیاد به مجموعه‌ای با بُعد پایین‌تر، معرفی الگوریتم‌ها و ساختمان داده مربوط به MapReduce)
- استخراج اطلاعات مفید از داده‌ها (معرفی الگوریتم‌هایی برای نمایش مناسب داده‌ها و به دست آوردن اطلاعات پنهان از آن‌ها، روش‌های رگرسیون و برازش تابع، خوشه‌بندی و دسته‌بندی)
- پیش‌بینی رفتار داده‌ها (معرفی الگوریتم‌هایی که با بررسی نمونه‌هایی از داده‌ها رفتار داده‌های بعدی از این نوع را پیش‌بینی کند، الگوریتم‌های نزدیک‌ترین همسایه)
- بررسی ارتباط و رتبه‌بندی داده‌ها (الگوریتم رتبه‌بندی صفحات)

منابع:

[۱] Brian, Steele and John, Chandler and Swarna, Reddy. Algorithms for Data Science, Springer, ۲۰۱۶.

[۲] Charu C. Aggarwal. Data Mining: The text book, Springer, ۲۰۱۵.

[۳] Jure, Leskovec and Anand, Rajaraman and Jeff Ullman. Mining massive datasets, Stanford, ۲۰۱۰.

عنوان درس	فارسی	بهینه‌سازی در علوم داده									
	انگلیسی	Optimization for Data Science									
نوع واحد									تعداد واحد	تعداد ساعت	دروس پیش‌نیاز
پایه		اصلی				تخصصی		اختیاری		۳	۴۸	ندارد
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: ندارد		نیاز به اجرای پروژه عملی: ندارد									

هدف:

بسیاری از مسائل آماری و علوم داده در نهایت منجر به مسائلی در بهینه‌سازی می‌شوند. با آنکه نظریه ریاضی بهینه‌سازی قدمت زیادی دارد، پیشرفت‌های اخیر، این حوزه را به دو دلیل در کانون اهمیت قرار داده است: نخست کشف روش‌های نیرومندی که امکان می‌دهد در پاره‌ای موارد مسائل پیچیده را به سرعتِ مسائل خطی حل کنیم و دوم کاربردهای جدید و فراگیر روش‌های بهینه‌سازی محدب در حوزه‌های مختلف از جمله تحلیل و مدل‌سازی داده‌ها. هدف این درس آشنایی با مبانی نظری و روش‌های نوین در بهینه‌سازی و کاربردهای آن در علوم داده است.

سرفصل درس:

- مجموعه‌ها و توابع محدب (مجموعه‌های محدب و مستوی، ابرصفحه‌های جداساز و تکیه‌گاه، نامساوی‌های تعمیم‌یافته، مخروط دوگان، توابع مزدوج، توابع شبه محدب، توابع لگاریتم‌محدب و لگاریتم‌مقعر، ...)
- مسائل بهینه‌سازی محدب (بهینه‌سازی محدب، مسائل بهینه‌سازی خطی، مسائل بهینه‌سازی درجه دو، برنامه‌ریزی هندسی، بهینه‌سازی برداری)
- دوگانی (تابع دوگان لژاندر، مساله دوگان لژاندر، تعبیر هندسی، تعبیر نقطه زینی، شرایط بهینگی، اختلال و تحلیل حساسیت، قضایای دگرین)
- تقریب و برازش (تقریب نُرم، مسائل کمترین نُرم، تقریب منظم، تقریب استوار، برازش تابع و درون‌یابی)
- تخمین آماری (تخمین توزیع‌های پارامتری، تخمین توزیع‌های ناپارامتری، طراحی آشکارساز بهینه و آزمون فرض، کران‌های چبیشف و چرنف، طراحی آزمایش)
- مسائل هندسی (تصویر روی یک مجموعه، فاصله مجموعه‌ها، فاصله اقلیدسی و مسائل زاویه، بیضی‌گون‌های با حجم بیشینه یا کمینه، مرکزیابی، دسته‌بندی، جایابی)
- کمینه‌سازی نامقید (مسائل کمینه‌سازی نامقید، روش‌های کاهشی، روش کاهشی گرادیان، روش تندترین کاهش، روش نیوتن، پیاده‌سازی)
- کمینه‌سازی با قید تساوی (مسائل کمینه‌سازی با قید تساوی، روش نیوتن با قید تساوی، روش نیوتن با نقطه شروع نشدنی، پیاده‌سازی)
- روش‌های نقطه درونی (مسائل کمینه‌سازی با قید نامساوی، تابع مانعی لگاریتمی و مسیر مرکزی، روش‌های مانعی، شدنی بودن و روش‌های فاز ۱، تحلیل پیچیدگی، روش‌های نقطه درونی اولیه-دوگان، پیاده‌سازی)
- روش گرادیانی (روش کاهش گرادیانی، زیرگرادیان، روش کاهش گرادیان تصادفی، توابع لیپشیتز، توابع قویاً محدب)

منابع:

[۱] S. Boyd, L. Vanderberghe, Convex Optimization, Cambridge University Press, ۲۰۰٤.

[۲] S. Shalev-Shwartz and S. Ben-David, Understanding Machine Learning From Theory to Algorithms, Cambridge University Press, ۲۰۱٤.

عنوان درس	فارسی	روش‌های عددی در علوم داده													
	انگلیسی	Numerical Methods for Data Science													
نوع واحد												تعداد واحد	تعداد ساعت	دروس پیش‌نیاز	
پایه		اصلی		تخصصی		اختیاری						۳	۴۸		
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی								
حل تمرین: ندارد						نیاز به اجرای پروژه عملی:ندارد									

هدف:

در این درس دانشجویان با کاربردهای روش‌های عددی و به ویژه جبرخطی عددی در علوم داده و داده‌کاوی آشنا می‌شوند.

سرفصل درس:

یادآوری مباحثی از جبر خطی و ماتریس‌ها، روش QR، تقریب با کمترین مربعات، تجزیه ماتریسی: شامل تجزیهٔ مقدارهای منفرد یا تجزیهٔ مقدارهای تکین (SVD)، تجزیه ماتریس‌های نامنفی (NMF)، تجزیه ماتریس‌های دودویی (BMF)، روش تجزیه و تحلیل مولفه اصلی (PCA)، روش‌های کاهش رتبه ماتریس، روش زیرفضای کرایلف و تقریب ماتریسی و کاربرد آن در فشرده‌سازی تصاویر، کاربردها شامل: متن‌کاوی و بازیابی اطلاعات، خوشه‌بندی و بازیابی اطلاعات، جستجو در وب و مرتبه صفحات وب، تجزیه تانسور، تجزیه و تحلیل داده‌های ژنتیکی، تجزیه و تحلیل تصاویر و ....

منابع:

[۱] Matrix Methods in Data Mining and Pattern Recognition, by L. Elden, SIAM, ۲۰۰۷.
[۲] Applied Numerical Linear Algebra, by James W. Demmel, ۱۹۹۷.
[۳] Matrix Computations, by Gene H. Golub and Charles F. van Loan, ۱۹۹۶.
[۴] Pattern Recognition and Machine Learning, by Christopher M. Bishop, ۲۰۰۶.

عنوان درس	فارسی	روش‌های توپولوژی در علوم داده											
	انگلیسی	Topological Data Analysis											
نوع واحد										تعداد واحد	تعداد ساعت	دروس پیش‌نیاز	
پایه		اصلی		تخصصی		اختیاری				۳	۴۸	ندارد	
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی						
حل تمرین: ندارد						نیاز به اجرای پروژه							
عملی: ندارد													

هدف:

تحلیل داده از نگاه توپولوژی، روشی نوین جهت بررسی و تحلیل داده‌ها با استفاده از ابزار توپولوژیک است. در حالت کلی استخراج اطلاعات از داده‌های با بعد بالا، ناکامل و دارای نویز کاری دشوار است. اولین ایده و انگیزه تحلیل داده از نگاه توپولوژی، مطالعه شکل داده است. تحلیل داده از نگاه توپولوژی، با بهره گیری از تاریخچه غنی و ابزاهای بسیار ژرف و پیشرفته مطالعه اشکال در ریاضیات، به بررسی دقیق و عددی اشکال و استراج اطلاعات از آنها می‌پردازد. در این درس در ابتدا روشهای مختلف نسبت دادن یک شکل به داده بیان می‌شود و در ادامه به بیان دو ابزار مهم مطالعه این اشکال، همولوژی پایدار و نظریه مورس، می‌پردازیم.

سرفصل درس:

- فلسفه تحلیل داده از نگاه توپولوژی و ترسیم شکل کلی و بیان کاربردهای مختلف.
- داده و مجتمع‌های سادکی (مقدمه‌ای بر مجتمع‌های سادکی، معرفی و بررسی مجتمع‌های سادکی از جمله، ریپ، چک و ویتنس)
- گروه‌های همولوژی سادکی (معرفی گروه‌های همولوژی سادکی بر روی میدان دلخواه، معرفی رسته گروه‌های همولوژی سادکی و ساخت فانکتور همولوژی از رسته فضاهای توپولوژیک به رسته گروه‌های همولوژی، ساخت رسته همولوژی سادکی موضعی، بیان ویژگی‌های اصلی گروه‌های همولوژی، محاسبه گروه‌های همولوژی کره، چنبره، بطری کلاین)
- همولوژی پایدار (فیلتریشن و معرفی همولوژی پایدار، بیان الگوریتم پایدار، بیان دیاگرام پایدار، بیان کاربردها)
- ساخت و استنتاج شکل‌های توپولوژیکی از روی داده (ساخت منحنی از روی داده، ساخت رویه در فضای سه بعدی از روی داده، محاسبه همولوژی از روی داده)
- توابع مورس (بیان مفاهیم اصلی، توابع مورس، ترانسورسالیتی و توابع تکه‌ای خطی، معرفی گراف ریپ، مجتمع سادکی مورس- اسمیل، الگوریتم و کاربردها)

منابع

[۱] Edelsbrunner, Herbert, and John Harer. Computational topology: an introduction. American Mathematical Soc., ۲۰۱۰.

[۲] Dey, Tamal K., Herbert Edelsbrunner, and Sumanta Guha. "Computational topology." Contemporary mathematics ۲۲۳ (۱۹۹۹): ۱۰۹-۱٤٤.

[۳] Zomorodian, Afra J. Topology for computing. Vol. ۱٦. Cambridge university press, ۲۰۰٥.

[۴] Nakahara, Mikio. Geometry, topology and physics. CRC Press, ۲۰۰۳.
[۵] Munkres, James R. Elements of algebraic topology. Vol. ۲. Menlo Park: Addison-Wesley, ۱۹۸۴.

عنوان درس	فارسی	روش‌های هندسی در علوم داده									
	انگلیسی	Geometrical Methods for Data Science									
نوع واحد									تعداد واحد	تعداد ساعت	دروس پیش‌نیاز
پایه		اصلی		تخصصی		اختیاری			۳	۴۸	
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: ندارد						نیاز به اجرای پروژه عملی: ندارد					

هدف:

تمرکز این درس بر آموزش روش‌های هندسی در مدل‌سازی، نمایش و استخراج ساختار داده‌ها است که بخش مهمی از آن به کاهش بُعد غیر خطی استوار است.

سرفصل درس:

- مفاهیم اصلی (فضای اقلیدسی، متر، اندازه، خم، خمینه ریمانی، ژئودزیک، گراف‌های متریک، گراف نزدیک‌ترین فاصله، گراف شباهت داده‌ها، ...)
- یادگیری خمینه (Manifold Learning)
- الگوریتم‌های Isomap و Local Linear Embedding
- روش‌های طیفی (الگوریتم Laplacian Eigenmaps و رابطه شکل و ساختار داده با مقادیر ویژه لاپلاسین)
- PCA غیر خطی
- هسته گرمایی (Heat Kernel) ارتباط با مقادیر ویژه لاپلاسین و خواص داده)
- مباحث انتخابی (کاربردهای یادگیری خمینه در فشرده‌سازی داده، کاهش بُعد، حذف نویز، نمایش و مصورسازی داده، یادگیری خمینه نظارت شده، ...)
-

منابع:

[۱] Ma, Y. and Fu, Y. Manifold Learning, Theory and Applications. CRS Press, ۲۰۱۱.

عنوان درس	فارسی	نظریه یادگیری									
	انگلیسی	Learning Theory									
نوع واحد									تعداد واحد	تعداد ساعت	دروس پیش‌نیاز
پایه		اصلی		تخصصی		اختیاری			۳	۴۸	-
نظری	عملی	**نظری**	عملی	نظری	عملی	نظری	عملی				
حل تمرین: ندارد						نیاز به اجرای پروژه عملی: ندارد					

هدف:

هدف این درس آشنایی دانشجو با مفاهیم نظری و زیربنایی نظریه یادگیری، از جمله تعریف یادگیری و یادگیری پذیری و آشنایی با روش‌های رایج در یادگیری به همراه آشنایی با زیربناهای نظری آن‌ها است. همچنین در این درس دانشجو با برخی روش‌های معمول در یادگیری آشنا می‌شود و می‌آموزد که به جنبه‌های نظری در مسئله‌های یادگیری چگونه می‌توان پرداخت.

سرفصل درس:

- مقدمات احتمالاتی (نامساوی مارکوف، کران چرنوف) و مقدمات آماری (مقدمه‌ای بر برآورد درست‌نمایی بیشینه)
- تعریف مدل یادگیری احتمالا تقریبا درست (PAC-Learning)، یادگیری احتمالا تقریبا درست انکاری (Agnostic PAC-Learning)
- مدل یادگیری غیریکنواخت، مقایسه قابلیت های یادگیری مدل یادگیری غیریکنواخت و مدل یادگیری احتمالا تقریبا درست.
- تیغ اوکام، تعابیر فلسفی و الگوریتم یادگیری بر اساس تیغ اوکام
- بعد VC
- یادگیری پذیری و یادگیری ناپذیری ذاتی مجموعه‌های فرضیه بر اساس بعد VC
- آشنایی مقدماتی با محاسبات و پیچیدگی محاسبات و محدودیت‌های محاسباتی
- محدودیت‌های محاسباتی یادگیری: مسئله‌هایی که می‌توان از لحاظ آماری یادگرفت اما نمی‌توان از لحاظ محاسباتی در زمان چندجمله‌ای یادگرفت، مگر اینکه P=NP. مانند یادگیری DNF.
- رگرسیون، رگرسیون سنیغ (Ridge Regression)، یادگیری‌ناپذیر بودن رگرسیون و یادگیری‌پذیر بودن رگرسیون سنیغ
- توانایی‌ها و محدودیت‌های الگوریتم‌های ژنتیک در یادگیری مفاهیم
- معیارهای انتخاب و تشخیص مدل برای یادگیری
- آشنایی با یادگیری برخط (Online Learning)
- آشنایی مقدماتی با برخی روش‌های و تکنیک‌ها در یادگیری: یادگیری محدب و الگوریتم تصادفی حرکت در جهت گرادیان (Stochastic Gradient Descent)، درخت تصمیم، نزدیک‌ترین همسایه، Support Vector Machine، شبکه‌های عصبی، کاهش بعد، روش‌های هسته (Kernel Methods)، مدل‌های مولد (Generative Models)، خوشه‌بندی (Clustering)

منابع:

[۱] S. Shalev-Shwartz, S. Ben-David, Understanding Machine Learning: From Theory to Algorithms, Cambridge University Press, ۲۰۱٤.

[۲] M. Kearns, U. Vazirani, An Introduction to Computational Learning Theory, The MIT Press, ۱۹۹٤.

[۳] Valiant, Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World, Basic Books, ۲۰۱٤.

عنوان درس	فارسی	یادگیری آماری									
	انگلیسی	Statistical Learning									
نوع واحد									تعداد واحد	تعداد ساعت	دروس پیش‌نیاز
پایه		اصلی		تخصصی		اختیاری			۳	۴۸	
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: ندارد عملی: ندارد								نیاز به اجرای پروژه			

هدف:

یادگیری آماری به توسعه و مطالعه ابزارهایی برای مدل کردن و فهمیدن هرچه بهتر داده‌های پیچیده می‌پردازد. یادگیری از داده یعنی کشف الگوها و روندهای مهم در داده و فهمیدن آن که داده چه می‌گوید! حجم و پیچیدگی روزافزون داده‌ها در بسیاری از حوزه‌های علم و فن‌آوری و چالش‌ها و موفقیت‌های جدید در یادگیری به انقلابی در علوم داده منجر شده است. هدف از این درس، آشنایی با برخی از مهم‌ترین ابزارهای یادگیری و بررسی آن‌ها از منظر آماری است.

سرفصل درس:

- مقدمات (مفاهیم اولیه یادگیری، انواع یادگیری، مدل‌های پارامتری و ناپارامتری، ...)
- یادگیری نظارت شده: روش‌های خطی (رگرسیون خطی، رگرسیون ستیغی و لاسو، روش‌های خطی دسته‌بندی: مدل لجستیک، ...)
- یادگیری نظارت شده: تعمیم‌های روش خطی (مدل‌های خطی تعمیم‌یافته، توابع پایه، روش‌های ناپارامتری، اسپلاین هموار شده، مدل‌های جمعی،...)
- انتخاب و ارزیابی مدل (درجه آزادی، تجزیه اریبی-واریانس، خطای آزمون و آموزش، اعتبارسنجی متقابل، روش‌های خودگردان، ...)
- یادگیری بدون نظارت (روش‌های کاهش بُعد، خوشه‌بندی، مدل‌های گرافی، ...)
- شبکه‌های عصبی (آموزش شبکه: کاهش گرادیان و انتشار پس‌رو، کاهش گرادیان تصادفی، یادگیری ژرف، ...)

منابع:

[۱] Jamea, G., Witten, D., Hastie, T. and Tibshirani, R., Friedman, J. An Introduction to Statistical Learning. ۲nd edition, Springer, ۲۰۱۳

[۲] Hastie, T., Tibshirani, R. and Friedman, J. The Elements of Statistical Learning. ۲nd edition, Springer, ۲۰۰۹

عنوان درس	فارسی	مدل‌سازی و تحلیل داده‌های با بُعد بالا									
	انگلیسی	Analysis and Modeling High Dimensional Data									
نوع واحد								تعداد واحد	تعداد ساعت	دروس پیش‌نیاز	
پایه		اصلی		تخصصی		اختیاری		۳	۴۸		
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: ندارد عملی: ندارد						نیاز به اجرای پروژه					

هدف:

در اثر پیشرفت‌های سریع در روش‌های جمع‌آوری داده در سال‌های اخیر، آماردانان با مسائلی مواجه شده‌اند که در آن‌ها تعداد متغیرهای اندازه‌گیری شده بسیار زیاد و گاه بسیار بیشتر از تعداد مشاهدات است. به چنین داده‌هایی داده‌های با بُعد بالا می‌گویند و تحلیل آن‌ها از جمله چالش‌های اصلی علوم داده در دوره جدید است. هدف این درس آشنایی با مبانی ریاضی و آماری تحلیل داده‌های با بُعد بالا است.

سرفصل درس:

- مفاهیم اولیه (استنباط آماری کلاسیک و بلای بُعد، فضاهای اقلیدسی با ابعاد بالا، تجمع اندازه، ...)
- مدل‌های خطی در ابعاد بالا (رگرسیون ستیغی و لاسو، انتخاب متغیر، درجه آزادی، ...)
- مدل‌های خطی تعمیم‌یافته در ابعاد بالا (رگرسیون لجستیک و پوآسون در ابعاد بالا، ...)
- مدل‌های گرافیکی (گراف‌های وابستگی، خواص مارکوفی، مدل‌های گوسی، ...)
- آزمون‌های فرض همزمان (p- مقدارها، کنترل خطای جمعی، نرخ کشف خطا، روش بنجامینی و هوشبرگ، ...)
- مباحث انتخابی (گذار فاز، انتخاب مدل، جنگل‌های تصادفی، بیز تجربی، ...)

منابع:

[۱] Buhlmann, P. and van de Geer, S. Statistics of High Dimensional Data, Springer, ۲۰۱۱
[۲] Efron, B. and Hastie, T. Computer Age Statistical Inference, Cambridge University Press, ۲۰۱۶
[۳] Giraud, C. Introduction to High-Dimensional Statistics, CRC Press, ۲۰۱۶
[۴] Hastie, T., Tibshirani, R. and Wainwright, M. Statistical Learning with Sparsity, CRC Press, ۲۰۱۵

عنوان درس	فارسی	سری‌های زمانی									
	انگلیسی	Time Series									
نوع واحد								تعداد واحد	تعداد ساعت	دروس پیش‌نیاز	
پایه		اصلی		تخصصی		اختیاری		۳	۴۸		
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: ندارد						نیاز به اجرای پروژه عملی:ندارد					

هدف:

داده‌های دارای روند زمانی یا سری‌های زمانی در کاربردهای متنوعی از جمله داده‌های مالی و اقتصادی، زیستی، آب و هوا، مهندسی، ... به وفور ظاهر می‌شوند. هدف این درس آشنایی با روش‌های تحلیل سرهای زمانی است. مدل‌سازی، تخمین مدل، پیش‌بینی، تشخیص، پالایش از جمله مسائلی است که در تحلیل سری‌های زمانی مورد مطالعه قرار می‌گیرند.

سرفصل درس:

- مفاهیم اولیه (روندها ومولفه‌های فصلی، ایستایی، تابع خودهمبستگی، تابع خودهمبستگی جزئی، تجربه والد، ...)
- رگرسیون و تحلیل اکتشافی (کشف روند، ارتباط با سری‌های زمانی دیگر، هموارسازی، تحلیل اکتشافی ...)
- مدل‌های (ARMA) مدل‌های خودبازگشتی، مدل‌های میانگین متحرک، وجود و یکتایی، استنباط و پیش‌بینی، ...)
- تحلیل طیفی (رفتار تناوبی و دوره‌نگار، تبدیل فوریه و نمایش طیفی، چگالی طیفی، ...)
- پالایش و پیش‌بینی (امید شرطی، فضای هیلبرت متغیرهای تصادفی، پالایه کالمن، روش بیزی، ...)
- مباحث انتخابی (مدل‌های غیرخطی، مدل‌های با حافظه بلندمدت، آزمون‌های ریشه واحد، مدل‌های ARIMA سری‌های زمانی مالی، ...)

منابع:

[۱] Brockwell, P. J. and Davis, R. A. Introduction to Time Series and Forecasting. ۳rd edition, Springer, ۲۰۱۶.

[۲] Cryer, J. D. and Chan, K. S. Time Series Analysis with Applications in R, ۲nd edition, Springer, ۲۰۰۸.

[۳] Shumway, R. H. and Stoffer, D. S. Time Series Analysis and Its Applications, ۲nd edition, Springer, ۲۰۰۶.

[۴] Fristdet, A. Jain, N. and Krilov, N. Filtering and Prediction: A Primer, Student Mathematical Library, Vol. ۳۸, AMS, ۲۰۰۷.

عنوان درس	فارسی	مدل‌های گرافی احتمالاتی										دروس پیش‌نیاز
	انگلیسی	Probabilistic Graphical Models										
نوع واحد								تعداد واحد	تعداد ساعت			
پایه		اصلی		تخصصی		اختیاری		۳	۴۸			
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی					
حل تمرین: ندارد عملی:ندارد						نیاز به اجرای پروژه						

هدف:

مدل‌های گرافی احتمالاتی چارچوبی غنی و انعطاف‌پذیر برای مدل‌سازی خانواده‌های بزرگ از متغیرهای تصادفی با وابستگی‌های پیچیده هستند. هدف این درس مرور مفاهیم اساسی مدل‌های گرافی احتمالاتی با ارتباطات علّی یا وابستگی و روش‌های استنباط دقیق یا تقریبی و نیز آشنایی با نحوه استفاده از این مدل‌ها در عمل است.

سرفصل درس:

- مدل‌های جهت‌دار: شبکه‌های بیزی (استقلال شرطی، استقلال در شبکه‌های بیزی، روابط علّی، ...)
- مدل‌های بدون جهت: شبکه‌های مارکوفی (توزیع گیبس، شرطی کردن و استقلال در میدان‌های تصادفی، ...)
- حذف متغیر (روش حذف متغیر و چالش‌های آن، پیچیدگی بر حسب ساختار گراف، ترتیب حذف متغیرها، ...)
- الگوریتم‌های انتشار باور (الگوریتم و ویژگی‌ها، انتقال پیام در گراف خوشه‌ها، مطالعه الگوریتم در درخت‌ها، ...)
- روش‌های نمونه‌گیری (روش‌های مونت‌کارلو با زنجیر مارکوفی، روش گیبس، روش متروپولیس-هستینگز، ...)
- تخمین پارامتر (روش بیشترین درست‌نمایی و روش بیزی در مدل‌های گرافی جهت‌دار و بدون جهت)
- یادگیری ساختار شبکه (بیشترین درست‌نمایی و سازگاری مجانبی، یادگیری ساختار درختی، روش‌های اکتشافی برای یادگیری ساختار گراف، یادگیری با داده ناکامل، ...)

منابع:

[۱] Koller, D. and Friedman, N. Probabilistic Graphical Models: Principles and Techniques, MIT Press, ۲۰۰۹.

[۲] Wainwright, M. J. and Jordan, M. I. Graphical Models, Exponential Families and Variational Inference, ۲۰۰۸.

عنوان درس	فارسی	داده‌کاوی										
	انگلیسی	Data Mining										
نوع واحد									تعداد واحد	تعداد ساعت	دروس پیش‌نیاز	
پایه		اصلی		تخصصی		اختیاری			۳	۴۸		
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی					
حل تمرین: ندارد نیاز به اجرای پروژه عملی:ندارد												

هدف:

این درس فرایند اکتشاف دانش و گام‌های آن را معرفی می‌کند. دانشجویان در این درس خواهند یافت که چگونه به کمک روش‌ها و الگوریتم‌های محاسباتی داده‌کاوی، از داده‌های خام، الگوها و مدل‌های توصیفی و پیش‌بینی استخراج می‌شود. همچنین، این الگوریتم‌ها از نظر کارایی، پیچیدگی محاسباتی و مصرف حافظه تحلیل خواهند شد.

سرفصل درس:

- فرآیند کاوش دانش در پایگاههای داده (KDD) و بیان جایگاه داده‌کاوی در این فرآیند،
- گام‌های پردازش داده (پاک‌سازی داده، تجمیع و تبدیل داده، کاهش داده، گسسته‌سازی داده)
- معرفی انبار داده (Data Warehouse) و معماری و پیاده‌سازی آن، پردازش تحلیلی بر خط (OLAP)،
- فناوری مکعب داده و تعمیم داده،
- مفهوم ویژگی و روش‌های کاهش بعد (اصول انتخاب ویژگی به کمک روش‌های آماری، رتبه‌بندی ویژگی‌ها براساس آنتروپی)
- تجزیه و تحلیل سبد بازار (معرفی مجموعه اقلام مکرر(Frequent Item Set) و الگوهای مکرر (Frequent patterns)، قواعد همباشی (Association rules) و کاوش آنها و معیارهای جذابیت این قواعد، معرفی الگوریتم‌های Apriori و ECLAT و FPgrowth و مقایسه کارایی و پیچیدگی محاسباتی آن‌ها، مجموعه اقلام پرتکرار بسته و کاوش آنها، استخراج همباشی‌های چند سطحی و چند بعدی، استخراج الگوهای نادر، استخراج الگوهای پرتکرار مقید، استخراج الگوهای فشرده یا تقریبی، استخراج قواعد همباشی منفی)
- روش‌های دسته‌بندی و مقایسه آن‌ها از نظر کارایی و پیچیدگی محاسباتی و حافظه مورد نیاز:
  - اندازه‌های ناسرگی (Impurity Measures) و مثال‌های آن (آنتروپی و شاخص gini)، هرس و مقیاس پذیری درخت، ساخت درخت با داده‌های ناقص
  - دسته‌بندی قاعده‌پایه (Ruled-based)، استخراج قاعده از درخت، الگوریتم پوششی ترتیبی برای استخراج قاعده (Sequential Covering Algorithm)
  - دسته‌بندی به کمک تجزیه و تحلیل قواعد همباشی (کلاسبندی همباشی)
- روش‌های خوشه‌بندی و مقایسه آن‌ها از نظر کارایی و پیچیدگی محاسباتی و حافظه مورد نیاز:
  - روش‌های افراز: روش K-means و روش K-medoids،
  - روش‌های سلسله‌مراتبی: روش جمع‌شونده (Agglomerative) و تقسیم‌شونده (Divisive)، روش سلسله‌مراتبی برای داده‌های رسته‌ای (Rock)
  - روش‌های چگالی‌پایه (Density-Based) روش‌های DBSCAN، OPTICS، DENCUE
  - روشهای توری‌پایه (Grid-Based) روش‌های STING، Wave Cluster

- روش‌های خوشه‌بندی داده‌های ابعاد بالا: روش‌های CLIQUE، PROCLUS، الگوهای مکرر پی
- روش‌های خوشه‌بندی طیفی، روش‌های خوشه‌بندی مقید، روش‌های خوشه‌بندی نیمه نظارت شده
- روش‌های خوشه‌بندی فازی: C-میانگین فازی(FCM)، C-میانگین فازی رابطه‌ای ((RFCM، روش Gustafson-Kessel، روش Gath-Geva، خوشه‌بندی فازی هسته-پایه (Kernel-Based)، خوشه‌بندی فازی خودسازمان‌ده،

- روش‌های تشخیص داده‌های دورافتاده (Outliers):
  - روش‌های آماری (پارامتری و ناپارامتری)، روش‌های مجاورتی (فاصله‌پایه، چگالی‌پایه و توری‌پایه)، روش‌های خوشه‌بندی‌پایه، روش‌های دسته‌بندی‌پایه، مدل‌های نظریه اطلاعات‌پایه.
  - تشخیص داده‌های دورافتاده در داده‌های با ابعاد بالا، تشخیص داده‌های دورافتاده در داده‌های رسته‌ای، روش‌های ترکیبی در تشخیص داده‌های دورافتاده.
- معضل بلای بُعد (Curse of Dimensionality)، استخراج و انتخاب ویژگی.
  - روش‌های کاهش بعد: تحلیل تفکیک خطی (LDA)، تحلیل عاملی، مقیاس‌گذاری چند بعدی MDS، تحلیل مولفه‌های اصلی (PCA)، تحلیل مولفه‌های اصلی مبتنی بر هسته (KPCA)، تحلیل مولفه‌های مستقل (ICA)، ISOMAP، جاسازی محلی خطی (Locally Linear Embedding)
  - روش‌های انتخاب ویژگی: روش‌های پالایشی (Filter)، روش‌های پوشه‌ای (Wrapper)، روش‌های جاسازی (Embedded)

منابع:

[۱] [۱ J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, ۳rd Edition, Elsevier Inc., ۲۰۱۲.

[۲] C. C. Aggarwal, Data Mining: The Textbook, Springer, ۲۰۱۵.

[۳] C. C. Aggarwal, J. Han, Frequent Pattern Mining, Springer, ۲۰۱٤.

[٤] M. Kantardzic, Data Mining and Analysis: Foundations and Algorithms, ۲nd Edition, Wiley-IEEE Press, ۲۰۱۱.

[٥] M. J. Zaki, W. Meira, The Handbook of Data Mining, Cambridge University Press, ۲۰۱٤.

[٦] I. H. Witten, E. Frank, M. A. Hall, Data Mining: Pactical Machine Learning Tools and Techniques, ۳rd Edition, Elsevier Inc., ۲۰۱۱.

[۷] S. Marsland, Machine Learning: An Algorithmic Perspective, ۲nd Edition, Chapman & Hall/CRC, ۲۰۱٤.

[۸] M. Sato-Ilic, L. C. Jain, "Innovations in Fuzzy Clustering, Theory and Applications", Springer, ۲۰۰٦.

[۹] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, ۲nd edition, Springer, ۲۰۰۹.

عنوان درس	فارسی	دیداری‌سازی داده‌ها									
	انگلیسی	Data Visualization									
نوع واحد									تعداد واحد	تعداد ساعت	دروس پیش‌نیاز
پایه		اصلی		تخصصی		اختیاری			۳	۴۸	داده‌کاوی یا اجازه استاد
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: حداکثر ۲۴ ساعت				نیاز به اجرای پروژه عملی: دارد							

هدف:

این درس روش‌ها و اصول کلیدی دیداری‌سازی داده‌ها را معرفی می‌کند و به یادگیرنده نشان می‌دهد که چگونه بازنمایی‌های دیداری می‌توانند در درک داده‌های پیچیده مفید باشند و در پی آن است که توانایی یادگیرنده را برای فعالیت و پژوهش در این حوزه ارتقا دهد.

سرفصل درس:

- معرفی دیداری‌سازی داده‌ها و اهمیت آن،
- انواع داده و دادگان (datasets)، روش‌های تجرید داده‌ها، روش‌های تجرید وظایف،
- نشانه‌ها و کانال‌ها، کدگذاری داده‌ها به کمک نشانه‌ها و کانال‌ها، معیارهای کارایی کانال: دقت (Accuracy)، افتراق‌پذیری (Discriminability)، جدایی‌پذیری (Separability)، آشکارسازی تنظیم‌پذیر (Popout) و تمامیت (Integrality)،
- قواعدی که همواره در دیداری‌سازی داده‌ها باید مد نظر قرار داد،
- اعتبارسنجی و چهار سطح آن: وضعیت دامنه، تجرید داده‌ها و وظایف، کدگذاری دیداری و تعامل، پیچیدگی الگوریتم از نظر زمانی و حافظه،
- طراحی چیدمان جدول‌های داده، طراحی چیدمان داده‌های مکانی، طراحی چیدمان داده‌های شبکه‌ای و ساختارهای درختی،
- نگاشت رنگ و سایر کانال‌های غیر فضایی در کدگذاری دیداری، روش‌های دستکاری نما و دید (view) برای کاهش پیچیدگی نمایش داده‌ها، روش‌های نمایش داده‌های پیچیده با تقسیم نمایش به چندین نما و دید یا لایه،
- روش‌های کاهش اقلام و کاهش ویژگی برای مواجهه با پیچیدگی‌های دیداری‌سازی،
- روش‌های درج اطلاعات بر روی مجموعه انتخاب شده در یک نما و دید،
- معرفی سیستم‌های دیداری‌سازی داده‌ها و قابلیت‌های آن‌ها، همچون: Scagnostics، VisDB، PivotGraph، InterRing، Constellation و Hierarchical Clustering Explorer.

منابع:

[۱] T. Munzner, Visualization Analysis and Design, CRC Press, ۲۰۱٤.

[۲] G. Dzemyda, O. Kurasova, J. Zilinskas, Multidimensional Data Visualization: Methods and Applications, Springer, ۲۰۱۳.

[۳] S. Murray, Interactive Data Visualization for the Web, O'Reilly Media, ۲۰۱۳.

[٤] I. Meirelles, Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations, Rockport Publishers, ۲۰۱۳.

[٥] C. Ware, Visual Thinking for Design, Morgan Kaufman, ۲۰۰۸.

[٦] Alberto Cairo, The Functional Art: An introduction to information graphics and visualization-, New Riders, ۲۰۱۲.

عنوان درس	فارسی	تشخیص داده‌های دورافتاده										دروس پیش‌نیاز
	انگلیسی	Outlier Detection										
نوع واحد										تعداد واحد	تعداد ساعت	
پایه		اصلی		تخصصی		اختیاری				۳	۴۸	
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی					
حل تمرین: ندارد						نیاز به اجرای پروژه عملی:ندارد						

هدف:

این درس مفاهیم مربوط به داده‌های دورافتاده و نحوه تحلیل و شناسایی آن‌ها را معرفی می‌کند و دانشجویان را با نمونه‌هایی از کابردهای مربوط به شناسایی داده‌های دورافتاده آشنا می‌سازد.

سرفصل درس:

- معرفی داده‌های دورافتاده و معضل آن‌ها در مطالعات مختلف داده‌کاوی
- ارزیابی روش‌های شناسایی داده‌های دورافتاده
- شناسایی داده‌های دورافتاده به کمک مدل‌های احتمالی و آماری
- شناسایی داده‌های دورافتاده به کمک مدل‌های خطی
- شناسایی داده‌های دورافتاده به کمک مدل‌های مبتنی بر نزدیکی
- شناسایی داده‌های دورافتاده به کمک مدل‌های مبتنی بر نظریه اطلاعات
- شناسایی داده‌های دورافتاده به کمک روش‌های گروهی (Ensembles)
- شناسایی داده‌های دورافتاده در داده‌های با ابعا بالا
- شناسایی داده‌های دورافتاده به کمک روش‌های نظارت‌شده
- شناسایی داده‌های دورافتاده در داده‌های رسته‌ای، متنی و با ویژگی‌های ترکیبی
- شناسایی داده‌های دورافتاده در سری‌های زمانی و داده‌های روان
- شناسایی داده‌های دورافتاده در دنباله‌های گسسته
- شناسایی داده‌های دورافتاده در داده‌های فضایی
- شناسایی داده‌های دورافتاده در داده‌های گرافی و در شبکه‌ها
- معرفی کاربردهای شناسایی داده‌های دورافتاده در کنترل کیفیت و شناسایی خطا، تشخیص حمله و امنیت،علوم زمین، وب‌کاوی و متن‌کاوی

منابع:

[۱] C. C. Aggarwal, Outlier Analysis, ۲nd Edition, Springer, ۲۰۱۶.
[۲] M. Gupta, J. Gao, C. C. Aggarwal, J. Han, Outlier Detection for Temporal Data, Morgan & Claypool, ۲۰۱٤.
[۳] P. J. Rousseeuw, A. M. Leroy, Robust regression and outlier detection, Wiley, ۱۹۸۷.
[٤] D. M. Hawkins, Identification of Outliers, Springer, ۱۹۸۰.
[٥] V. Barnett, T. Lewis, Outliers in Statistical Data, Wiley, ۱۹۷۸.

عنوان درس	فارسی	مدل‌سازی و پردازش مه‌داده‌ها									
	انگلیسی	Modeling and Processing Big Data									
نوع واحد								تعداد واحد	تعداد ساعت	دروس پیش‌نیاز	
پایه		اصلی		تخصصی		اختیاری		۳	۴۸		
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: ندارد عملی: ندارد						نیاز به اجرای پروژه					

هدف:

این درس مفاهیم مربوط به مه‌داده‌ها و نحوه مدیریت و تحلیل آنها را معرفی می‌کند و دانش و بینش لازم را برای ورود به مباحث پیشرفته در این حوزه فراهم می‌آورد.

سرفصل درس:

- مقدمات (معرفی مه‌داده‌ها و مثال‌های آن، منشا تولید مه‌داده، ...)
- مشخصه‌های مه‌داده‌ها (حجم، سرعت، تنوع، صحت، ظرفیت و ارزش، تاثیر هر مشخصه بر جمع‌آوری، پایش، ذخیره‌سازی، تحلیل و گزارش‌دهی مه‌داده‌ها)
- مولفه‌های ساختاری و مدل‌های برنامه‌نویسی برای تحلیل مه‌داده‌های مقیاس‌پذیر
- روش‌های دسترسی و دستکاری داده‌های روان (تفاوت پایگاه داده‌های سنتی با سامانه‌های مدیریت مه‌داده‌ها)
- Hadoop (HDF) به عنوان نظام فایل توزیع شده در Hadoop، معرفی YARN مدیریت منابع در Hadoop، معرفی MapReduce به عنوان مدل برنامه‌نویسی)
- بازیابی، یکپارچه‌سازی و تحلیل مه‌داده‌ها
- تحلیل مه‌داده‌ها (دسته‌بندی، رگرسیون، خوشه‌بندی و تحلیل هم‌باشی مه‌داده‌ها)
- تحلیل مه‌داده‌های گرافی
- پیاده‌سازی (پیاده‌سازی نمونه‌ای از سامانه مدیریت مه‌داده‌ها و تحلیل آن‌ها)
-

منابع:

[۱] R. Buyya, R. N. Calheiros, A. Vahid Dastjerdi, Big Data. Principles and Paradigms, Morgan Kaufmann, ۲۰۱۶.
[۲] F. Corea, Big Data Analytics: A Management Perspective, Springer, ۲۰۱۶.
[۳] I. Foster, R. Ghani, R. S. Jarmin, F. Kreuter, J. Lane, Big Data and Social Science: A Practical Guide to Methods and Tools, Chapman & Hall/CRC, ۲۰۱۷.
[۴] S. Suthaharan, Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning, ۴rd Edition, Springer, ۲۰۱۵.
[۵] S. Liu, J. McGree, Z. Ge, Y. Xie, Computational and Statistical Methods for Analysing Big Data with Applications, Academic Press, ۲۰۱۶.
[۶] J. Hurwitz, A. Nugent, F. Halper, M. Kaufman, Big Data for Dummies, John Wiley & Sons, ۲۰۱۳.
[۷] S. Perera, T. Gunarathne, Hadoop MapReduce Cookbook, Packt Publishing, ۲۰۱۳.
[۸] J. R. Owens, B. Femiano, J. Lentz, Hadoop Real World Solutions Cookbook, Packt Publishing, ۲۰۱۳.

عنوان درس	فارسی	گراف‌کاوی								دروس پیش‌نیاز	تعداد ساعت	تعداد واحد
	انگلیسی	Graph Mining										
نوع واحد												
پایه		اصلی شاخه		تخصصی		اختیاری				داده‌کاوی یا اجازه استاد	۴۸	۳
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی					
حل تمرین: حداکثر ۲۴ ساعت				نیاز به اجرای پروژه عملی: دارد								

هدف:

این درس روش‌های استخراج الگو از داده‌های گرافی را مورد توجه قرار می‌دهد. برای این منظور، ابتدا دانش پیش‌زمینه مورد نیاز از جبری خطی و نظریه گراف مرور می‌شود و سپس مسائل مطرح در زمینه‌های گراف‌کاوی و الگوریتم‌های مربوط معرفی می‌شوند تا توانایی لازم برای پژوهش در این حوزه کسب شود.

سرفصل درس:

- گراف‌ها، پایگاه داده‌های گرافی، تحلیل جبری گراف‌ها (ماتریس مجاورت، ماتریس لاپلاسی، ماتریس شباهت)،
- محاسبه فاصله در گراف‌ها، تطبیق گراف‌ها،
- گراف‌های ایستا: قوانین و الگوها،
- گراف‌های پویا: قوانین و الگوها،
- گراف‌های تصادفی،
- گام‌برداری تصادفی، الگوریتم‌های PageRank و HITS،
- روش‌های کاهش بعد در داده‌های گرافی،
- خوشه‌بندی و دسته‌بندی گراف‌ها،
- یادگیری نیمه‌نظارت شده در گراف‌ها،
- خلاصه‌سازی گراف‌ها،
- کشف نابهنجاری (Anomaly Detection) در گراف‌ها،
- تحلیل پیوند (Link Analysis)،
- زیرگراف‌های پرتکرار و روش‌های کاوش آن‌ها،
- کاوش زیرگراف‌های چگال،
- استخراج الگو از داده‌های گرافی روان (Stream Graph Mining)،
- تمایزکاوی (Contrast Mining) در گراف‌ها،
- روش‌های هسته برای گراف‌ها،
- یادگیری ژرف در گراف‌ها،
- ارائه مثال‌هایی از کاربرد گراف‌کاوی در داده‌های زیستی، داده‌های شیمیایی و شبکه‌های اجتماعی.

منابع:

[۱] C.C. Aggarwal, Data Mining: The Textbook, Springer, ۲۰۱۵.

[۲] D. Chakrabarti, C. Faloutsos, Graph Mining: Laws, Tools and Case Studies, Morgan Claypool, ۲۰۱۲.

[۳] N.F. Samatova, W. Hendrix, J. Jenkins, K. Padmanabhan, A. Chakraborty, Practical Graph Mining with R, CRC Press, ۲۰۱٤.

[٤] C.C. Aggarwal, H. Wang, Managing and Mining Graph Data, Springer, ۲۰۱۰.

[٥] G. Dong, J. Bailey, Contrast Data Mining: Concepts, Algorithms, and Applications, CRC Press, ۲۰۱۳.

[٦] D.J. Cook, L.B. Holder, Miming Graph Data, Wiley, ۲۰۰۷.

[۷] A. Schenker, Graph-Theoretic Techniques for Web Content Mining, World Scientific, ۲۰۰۵.

عنوان درس	فارسی	متن‌کاوی و وب‌کاوی									
	انگلیسی	Text Mining and Web Mining									
نوع واحد								تعداد واحد	تعداد ساعت	دروس پیش‌نیاز	
پایه		اصلی شاخه		تخصصی		اختیاری		۳	۴۸	داده‌کاوی	
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: حداکثر ۲۴ ساعت				نیاز به اجرای پروژه عملی: دارد							

هدف:

این درس بر روش‌های استخراج الگو از داده‌های متنی و صفحات وب تمرکز دارد. برای این منظور، ابتدا پیش‌زمینه‌های مورد نیاز از جبری خطی و یادگیری ماشین آماری را معرفی می‌کند و سپس به معرفی مسائل مطرح در زمینه‌های متن‌کاوی و وب‌کاوی و الگوریتم‌های مربوط می‌پردازد تا توانایی لازم برای پژوهش در این حوزه‌ها کسب شود.

سرفصل درس:

- معرفی متن‌کاوی، معرفی وب‌کاوی و شاخه‌های آن (محتواکاوی، پیوند[1]کاوی و کاربردکاوی)، تفاوت‌های داده‌کاوی با متن‌کاوی و وب‌کاوی، مرزهای مشترک بازیابی اطلاعات با متن‌کاوی و وب‌کاوی.
- پیش‌زمینه‌های جبری و آماری:
  - مدل‌سازی فضای برداری (VSM)،
  - توابع شباهت: همبستگی، شباهت کسینوسی،
  - فایل معکوس و فشرده‌سازی آن،
  - تجزیه SVD و روش‌های شاخص‌گذاری معنایی LSA (یا LSI) و pLSA،
  - مدل‌سازی موضوعی (Topic Modeling) و تخصیص پنهان دیریکله (LDA) و گونه برخط آن (OLDA)،
  - پربار کردن ویژگی‌ها در متون کوتاه به کمک تحلیل شاخص‌گذاری معنایی و بازخورد از کاربران.
- متن‌کاوی:
  - پیش‌پردازش و استخراج ویژگی در اسناد متنی،
  - خوشه‌بندی و دسته‌بندی اسناد متنی،
  - روش‌های یادگیری فعال و یادگیری نیمه نظارت شده در دسته‌بندی اسناد متنی،
  - خلاصه‌سازی متن،
- وب‌کاوی:
  - محتواکاوی:
    - استخراج موضوع صفحات وب به طور خودکار به کمک مدل‌سازی موضوعی،
    - خوشه‌بندی و دسته‌بندی صفحات وب، استفاده از یادگیری نیمه نظارت شده در دسته‌بندی صفحات وب،

[1] link

- نظرکاوی و تحلیل احساسات (sentiment)
- هرزه‌نگاری در وب (web spamming)
- رتبه‌دهی

- پیوندکاوی:
  - تشخیص پیوند کتابشناختی (bibliographic coupling) و ارجاع به هم (co-citation)،
  - الگوریتم‌های PageRank و HITS،
  - اکتشاف جوامع در وب،
  - دسته‌بندی صفحات وب به کمک پیوندها، هرزه‌نگاری پیوند،
- کاربردکاوی:
  - مدل‌سازی علائق کاربران وب به کمک خوشه بندی،
  - کاربردکاوی به کمک pLSA،
  - مدل‌سازی الگوی وب‌گردی کاربران به کمک تخصیص پنهان دیریکله (LDA)،
  - هم‌خوشه‌بندی (co-clustering) کاربران و صفحات وب،
  - استخراج الگو از تاریخچه پرس و جو کاربران،
- مباحث ترکیبی:
  - تحلیل شبکه‌های اجتماعی: معرفی مفهوم‌های مرکزیت (centrality) و شهرت (prestige)، اکتشاف جامعه در شبکه‌های اجتماع، یادگیری رتبه‌دهی در این شبکه‌ها.
  - سیستم‌های پیشنهاددده: روش‌های محتوا-پایه، روش‌های پالایش‌همکارانه، روش‌های ترکیبی.

منابع:

[۱] G. Xu, Y. Zhang. L. Li, Web Mining and Social Networking: Techniques and Applications, Springer, ۲۰۱۱.

[۲] B. Liu, Web Data: Exploring Hyperlinks, Contents, and Usage Data, ۲nd Edition, Springer, ۲۰۱۱.

[۳] C. C. Aggarwal, Data Mining: The Textbook, Springer, ۲۰۱۵.

[٤] S. M. Weiss, N. Indurkhya, T. Zhang, Fundamentals of Predictive Text Mining, Springer, ۲۰۱۰.

[٥] M. W. Berry, J. Kogan, Text Mining: Applications and Theory, John Wiley & Sons, ۲۰۱۰.

[٦] P. S. Sajja, R. Akerkar, Intelligent Technologies for Web Applications, CRC Press, ۲۰۱۲.

[۷] K.P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, ۲۰۱۲.

عنوان درس	فارسی	یادگیری ژرف									
	انگلیسی	Deep Learning									
نوع واحد									تعداد واحد	تعداد ساعت	دروس پیش‌نیاز
پایه		اصلی شاخه		تخصصی		اختیاری			۳	۴۸	یادگیری ماشین
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی				
حل تمرین: حداکثر ۲۴ ساعت				نیاز به اجرای پروژه عملی: دارد							

هدف:

این درس مفهوم یادگیری ژرف و تحقق آن را به کمک انواع شبکه‌های عصبی مصنوعی معرفی می‌کند و مفاهیم پایه و دید لازم را برای کاربرد این نوع یادگیری و انجام پژوهش در ارتباط با آن، فراهم می‌آورد.

سرفصل درس:

- نورون‌ها و مغز انسان، ساختار نورون‌ها، بررسی اجمالی شبکه های عصبی طبیعی، مفاهیم، تعاریف، و بخش‌های سازنده شبکه‌های عصبی،
- معرفی پرسپترون، شبکه تک لایه پرسپترونی، حل مساله دسته‌بندی به کمک پرسپترون و مشکل آن، مسائل جدایی‌پذیر خطی،
- شبکه چند لایه پیش‌رو و قاعده یادگیری پس انتشار خطا، حل مسائل دسته‌بندی و رگرسیون (تقریب تابع) به کمک این شبکه‌ها، ، بهبود شبکه انتشار خطا به عقب و نسخ مختلف آن، میزان آموزش و قدرت شبکه،
- روش‌های تنظیم (Regularization) در یادگیری ژرف،
- شبکه‌های عصبی کانولوشن، یادگیری ژرف به کمک این شبکه‌ها،
- المان‌های پردازشگر، اتصالات، تداعی الگوها، شبکه‌های تداعی‌گر پیش‌خور، شبکه‌های تداعی‌گر بازگشتی تک لایه، شبکه‌های تداعی‌گر دو طرفه، آموزش شبکه‌های بازگشتی، شبکه‌های بازگشتی ژرف،
- یادگیری بازنمایی به صورت بی‌نظارت،
- خودکدگذارها (Auto-encoders) و یادگیری بازنمایی به کمک آن‌ها،
- ماشین بولتزمن، شبکه‌های باور سیگوئیدی، ماشین قضیه میدان متوسط، ماشین بولتزمن ژرف، شبکه‌های باور ژرف،
- کاربردهای یادگیری ژرف در بینایی کامپیوتر، پردازش گفتار و پردازش زبان طبیعی.

منابع:

[۱] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, ۲۰۱۶.

[۲] N.D. Lewis, Deep Learning Made Easy With R: A Gentle Introduction for Data Science, CreateSpace Independent Publishing Platform, ۲۰۱۶.

[۳] J. Heaton, Artificial Intelligence for Humans, Volume ۳: Deep Learning and Neural Networks, Heaton Research, Inc., ۲۰۱۵.

[۴] J. Patterson, A. Gibson, Deep Learning: A Practitioner's Approach, O'Reilly Media, ۲۰۱۷.

[۵] D. Yu, L. Deng, Automatic Speech Recognition: A Deep Learning Approach, Springer, ۲۰۱۵.

عنوان درس	فارسی	سنجش فشرده											
	انگلیسی	Compressed Sensing											
نوع واحد										تعداد واحد	تعداد ساعت	دروس پیش‌نیاز	
پایه		اصلی		تخصصی		اختیاری				۳	۴۸		
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی						
حل تمرین: ندارد عملی:ندارد								نیاز به اجرای پروژه					

هدف:

در بسیاری از مدل‌سازی‌های مسائل کاربردی دستگاه‌های بزرگ معادلات خطی و غیرخطی به وجود می‌آیند. برای مثال الگوریتم رتبه‌بندی وب‌سایت گوگل منجر به حل دستگاه‌های خطی با میلیاردها مجهول می‌گردد. به دستگاه خطی که تعداد مجهول‌های آن از تعداد معادلاتش بیشتر باشد، دستگاه نامعین (undetermined) گفته می‌شود. در چنین حالتی دستگاه ممکن است بی نهایت پاسخ داشته باشد. معمولا هدف یافتن پاسخی است که تعداد مولفه‌های غیر صفر آن از یک مقدار مشخص کمتر باشد. سنجش فشرده (compressed sensing) روش‌هایی برای یافتن چنین پاسخ‌های از یک دستگاه خطی نامعین است. در این درس دانشجویان با این مبحث و روشهای جدید حل چنین دستگاههایی آشنا میشوند.

سرفصل درس:

مقدمه‌ای بر سنجش فشرده، روشهای حل دستگاههای خطی نامعین، تنکی، آشنایی با الگوریتم‌های مقدماتی برای حل دستگاه‌های خطی نامعین شامل: روشهای بهینه‌سازی، روش‌های مبتنی بر آستانه‌گذاری، روشهای حریصانه. خاصیت فضای پوچ، پایداری، بازیابی بردارها، ماتریس‌های بازیابی رتبه پایین، تعریف چسبندگی (coherence)، ماتریس‌های با چسبندگی کوچک. خاصیت طول پای محدود شده (RIP) و کاربرد آن در سنجش فشرده. مقدمه‌ای بر نظریه قاب‌ها و کابردهای آن در سنجش فشرده. کاربردهایی از سنجش فشرده در پردازش سیگنال و تصویر.

منابع:

[۱] Compressed Sensing: Theory and Applications, by Y. Eldar and G. Kutyniok, Cambridge University Press, ۲۰۱۱.

[۲] A Mathematical Introduction to Compressive Sensing, by S. Foucart and H. Rauhut, Birkhäuser Basel, ۲۰۱۳.

عنوان درس	فارسی	مباحث ویژه در علوم داده								
	انگلیسی	Special Topics in Data Science								
نوع واحد								تعداد واحد	تعداد ساعت	دروس پیش‌نیاز
پایه		اصلی		تخصصی		اختیاری		۳	۴۸	اجازه گروه
نظری	عملی	نظری	عملی	نظری	عملی	نظری	عملی			
حل تمرین: ندارد عملی:ندارد				نیاز به اجرای پروژه						

درسی است در سطح کارشناسی ارشد یا بالاتر در زمینه علوم داده که سرفصل آن بر حسب امکانات و نیاز در نیم‌سال مورد نظر توسط استاد مربوطه پیشنهاد شده و پس از تصویب شورای تحصیلات تکمیلی گروه و دانشکده ارایه می‌شود.