



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 الرحمن الرحيم

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

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

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

۳- این برنامه درسی از شروع سال تحصیلی ۱۳۹۷-۱۳۹۸ به مدت ۵ سال قابل اجراست و پس از آن نیازمند بازنگری می باشد.

مجتبی شریعتی نیاسر

نایب رئیس شورای عالی برنامه ریزی آموزشی

عبدالرحیم نوده ابراهیم

دبیر شورای عالی برنامه ریزی آموزشی



برنامه و سرفصل درس‌های کارشناسی ارشد ریاضی کاربردی

گرایش آنالیز عددی (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 دانشگاه هایی که قبلاً مجوز اجرای رشته به صورت کلی و یا در برخی از گرایش های خاص این رشته را اخذ کرده اند می توانند در همان گرایش هایی که دانشجو گرفته اند کماکان اقدام به پذیرش دانشجو نمایند ولی در سایر گرایش های جدید و یا گرایش هایی که قبلاً مجوز اجرای آن را نداشته اند لازم است نسبت به اخذ مجوز با کد رشته محل مجزا اقدام نموده 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 کنند.

طول دوره و شکل نظام

دوره کارشناسی ارشد ریاضی کاربردی مطابق با آیین نامه جاری دوره ی کارشناسی ارشد وزارت عتف است. تعداد

واحدهای دوره

تعداد واحدهای درسی دوره کارشناسی ارشد ریاضی کاربردی ۳۲ واحد و به قرار زیر است:

درس های الزامی: ۱۲ 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

درس های 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 درس های اختیاری یکی از دوره های کارشناسی ارشد مرتبط.

سمینار: ۲ واحد

پایان نامه: 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 پایان نامه گذراندن دست کم ۱۲ واحد (که شامل درس های الزامی می باشد) و اجازه گروه ضروری است.

با توجه به پایه ای بودن دروس الزامی و تنوع ورودی های کارشناسی ارشد در این رشته، دروس الزامی به صورت ۴ واحدی ارائه خواهند شد. بدیهی است گروه های آموزشی اختیار دارند پس از تصویب دانشگاه به میزان ۱ واحد به محتوی سرفصل این دروس که به صورت ۳ واحدی تنظیم شده است اضافه نمایند.

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



فصل اول

مشخصات دوره کارشناسی ارشد ریاضی کاربردی - گرایش علوم داده



«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

دوره 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 تأکید بر جنبه‌های اصیل ریاضی و ابعاد نظری علوم داده امکان تربیت نیروهای متخصص در این زمینه را فراهم می‌آورد.

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 مهندسين 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 اجتناب‌ناپذیر است. از طرف دیگر پردازش و تحلیل مه‌داده‌هایی نظیر داده‌های هواشناسی، داده‌های زیستی یا داده‌های مالی به دلیل اندازه و معمولاً بعد بالا امر ساده‌ای نیست. چنانکه روش‌های موجود بر مبنای 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 دانش‌آموختگان 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 آیین‌نامه 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 جبر خطی، احتمال، آمار و احتمال مهندسی



فصل دوم

جدول دروس دوره کارشناسی ارشد ریاضی کاربردی - گرایش علوم داده



جدول شماره ۱: درس های الزامی (گرایش علوم داده)

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

* یک درس از دروس اصلی گرایش ها یا زیر گرایش های دیگر علوم ریاضی با نظر گروه یا دانشکده.

جدول شماره ۲- درس های تخصصی - انتخابی (گرایش علوم داده)

شماره درس	نام درس	تعداد واحد	پیش نیاز یا هم نیاز (ها)	
			نام درس	شماره درس
۲۰۱	بهینه سازی در علوم داده	۳		
۲۰۲	روش های عددی در علوم داده	۳		
۲۰۳	روش های توپولوژی در علوم داده	۳		
۲۰۴	روش های هندسی در علوم داده	۳		
۲۰۵	نظریه یادگیری	۳		
۲۰۶	یادگیری آماری	۳		
۲۰۷	مدل سازی و تحلیل داده های با بُعد بالا	۳		
۲۰۸	سری های زمانی	۳		
۲۰۹	مدل های گراف های احتمالاتی	۳		
۲۱۰	داده کاوی	۳		
۲۱۱	دیداری سازی داده ها	۳		
۲۱۲	تشخیص داده های دورافتاده	۳		
۲۱۳	مدل سازی و پردازش مه داده ها	۳		
۲۱۴	گراف کاوی	۳		
۲۱۵	متن کاوی و وب کاوی	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 ساعات
پایه		اختیاری	دروس پیش‌نیاز
نظری		نظری	۴۸
عملی		عملی	۳
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 مارکوفی (توزیع مانا، مونت کارلو با زنجیرهای مارکوفی: روش متروپولیس-هستینگز، نمونه‌برداری گیبس، قدم زدن تصادفی بر گراف‌های بدون جهت، قدم زدن تصادفی در فضای اقلیدسی، شبکه اینترنت به عنوان زنجیر مارکوفی)
- یادگیری ماشین (یادگیری دسته‌ای، تیغ آکام و همگرایی یکنواخت، الگوریتم پرسپترون، الگوریتم تصادفی حرکت در جهت گرادینان، یادگیری برخط، ماشین بردار پشتیبان، جداکننده‌های غیر خطی و توابع هسته، یادگیری قوی و ضعیف، تقویت، بُعد VC)
- خوشه‌بندی (k-میانه‌ها، خوشه‌بندی طیفی، خوشه‌بندی بازگشتی به کمک برش‌های تُنک، زیرماتریس‌های چگال، روش‌های مبتنی بر جریان، یافتن خوشه موضعی، اصول موضوع خوشه‌بندی)
- مدل‌های مارکوفی نهان، مدل‌های گراف‌ی و انتشار باور (مدل‌های مارکوفی نهان، مدل‌های گراف‌ی و انتشار باور، مدل‌های بیزی یا مدل‌های باور، میدان‌های مارکوفی، الگوریتم درختی، انتقال پیام)
- موضوعات دیگر (رتبه‌بندی، رای‌گیری، بردارهای تُنک 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	انگلیسی
الگوریتم‌های علوم داده		Algorithms for Data Science	
نوع واحد		تعداد واحد	تعداد ساعات
پایه		اختیاری	دروس پیش‌نیاز
اصلی		تخصصی	۴۸
نظری		عملی	۳
نظری		عملی	نیاز به اجرای پروژه
حل تمرین: ندارد		عملی: ندارد	

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 کنند، 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	۴۸	۳	ندارد
نظری	عملی	نظری	عملی	نظری	عملی	ندارد
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 سرعت 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)
- دوگانیت (تابع دوگان لژاندر، مساله دوگان لژاندر، تعبیر هندسی، تعبیر نقطه زینی، شرایط بهینگی، اختلال و تحلیل حساسیت، قضایای دگرین)
- تقریب و برازش (تقریب نرم، مسائل کمترین نرم، 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، مرکزبایی، 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 شدنی، 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:

[1] S. Boyd, L. Vanderberghe, *Convex Optimization*, Cambridge University Press, ۲۰۰۴.

[2] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning From Theory to Algorithms*, Cambridge University Press, ۲۰۱۴.



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

هدف:

تحلیل داده از نگاه توپولوژی، روشی نوین جهت بررسی و تحلیل داده ها یا استفاده از ابزار توپولوژیک است. در حالت کلی استخراج اطلاعات از داده های 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 ابزارهای بسیار ژرف و پیشرفته مطالعه اشکال در ریاضیات، به بررسی دقیق و عددی اشکال و استخراج اطلاعات از آنها می پردازد. در این درس در ابتدا روشهای مختلف نسبت دادن یک شکل به داده بیان می شود و در ادامه به بیان دو ابزار مهم مطالعه این اشکال، همولوژی پایدار و نظریه مورس، می پردازیم.

سرفصل درس:

- فلسفه تحلیل داده از نگاه توپولوژی و ترسیم شکل کلی و بیان کاربردهای مختلف.
- داده و مجتمع های سادگی (مقدمه ای بر مجتمع های سادگی، معرفی و بررسی مجتمع های سادگی از جمله، ریب، چک و ویتنس)
- گروه های همولوژی سادگی (معرفی گروه های همولوژی سادگی بر روی میدان دلخواه، معرفی رسته گروه های همولوژی سادگی و ساخت فانکتور همولوژی از رسته فضاهای توپولوژیک به رسته گروه های همولوژی، ساخت رسته همولوژی سادگی موضعی، بیان ویژگی های اصلی گروه های همولوژی، محاسبه گروه های همولوژی کره، چنبره، بطری کلاین)
- همولوژی پایدار (فیلتریشن و معرفی همولوژی پایدار، بیان الگوریتم پایدار، بیان دیاگرام پایدار، بیان کاربردها)
- ساخت و استنتاج شکل های 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، ترانسورسالیته و توابع تکه ای خطی، معرفی گراف ریب، مجتمع سادگی مورس - اسمیل، الگوریتم و کاربردها)

منابع

- [1] Edelsbrunner, Herbert, and John Harer. Computational topology: an introduction. American Mathematical Soc., ۲۰۱۰.
- [2] Dey, Tamal K., Herbert Edelsbrunner, and Sumanta Guha. "Computational topology." Contemporary mathematics ۲۲۳ (۱۹۹۹): ۱۰۹-۱۴۴.
- [3] Zomorodian, Afra J. Topology for computing. Vol. ۱۶. Cambridge university press, ۲۰۰۵.



[^o] Munkres, James R. Elements of algebraic topology. Vol. 7. Menlo Park: Addison-Wesley, 1984.

هدف:

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 لاپلاسیان)
- PCA غیر خطی
- هسته گرمایی (Heat Kernel) ارتباط با مقادیر ویژه لاپلاسیان و خواص داده
- مباحث انتخابی (کاربردهای یادگیری خمینه در فشرده‌سازی داده، کاهش بُعد، حذف نویز، نمایش و مصورسازی داده، یادگیری خمینه نظارت شده، ...)

منابع:

[1] Ma, Y. and Fu, Y. Manifold Learning, Theory and Applications. CRS Press, 2011.



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

هدف:

هدف این درس آشنایی دانشجویان 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 زیربنای نظری آن‌ها است. همچنین در این درس دانشجویان با برخی روش‌های معمول در یادگیری آشنا می‌شوند و می‌آموزند 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 قابلیت‌های 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 شیخ (Ridge Regression)، یادگیری ناپذیر بودن رگرسیون و یادگیری پذیر بودن رگرسیون شیخ
- 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 گرادینان (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, ۲۰۱۴.
- [۲] M. Kearns, U. Vazirani, An Introduction to Computational Learning Theory, The MIT Press, ۱۹۹۴.



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

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 اربیبی-وارینانس، 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	مدل سازی و تحلیل داده‌های با بُعد بالا			
		انگلیسی	Analysis and Modeling High Dimensional Data			
نوع واحد		تعداد واحد	تعداد ساعات	دروس پیش نیاز		
پایه	نظری	عملی	اصولی	نظری	عملی	اختیاری
۴۸	۳					
حل تمرین: ندارد		نیاز به اجرای پروژه				
عملی: ندارد						

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، فضاهاى 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 پواسون 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	مدل های گرافی احتمالاتی							
		انگلیسی	Probabilistic Graphical Models							
نوع واحد				تعداد واحد	تعداد ساعات	دروس پیش نیاز				
پایه	نظری	عملی	اصولی	نظری	عملی	اختیاری	۳	۴۸	حل تمرین: ندارد	عملی: ندارد
نیاز به اجرای پروژه										

هدف:

مدل های گرافی احتمالاتی چارچوبی غنی و انعطاف پذیر برای مدل سازی خانواده های بزرگ از متغیرهای تصادفی با وابستگی های پیچیده هستند. هدف این درس مرور مفاهیم اساسی مدل های گرافی احتمالاتی با ارتباطات علی یا وابستگی و روش های استنباط دقیق یا تقریبی و نیز آشنایی با نحوه استفاده از این مدل ها در عمل است.

سرفصل درس:

- مدل های جهت دار: شبکه های بیزی (استقلال شرطی، استقلال در شبکه های بیزی، روابط علی، ...)
- مدل های بدون جهت: شبکه های مارکوفی (توزیع گیبس، شرطی کردن و استقلال در میدان های تصادفی، ...)
- حذف متغیر (روش حذف متغیر و چالش های آن، پیچیدگی بر حسب ساختار گراف، ترتیب حذف متغیرها، ...)
- الگوریتم های انتشار باور (الگوریتم و ویژگی ها، انتقال پیام در گراف خوشه ها، مطالعه الگوریتم در درخت ها، ...)
- روش های نمونه گیری (روش های مونت کارلو با زنجیر مارکوفی، روش گیبس، روش متروپولیس-هستینگز، ...)
- تخمین پارامتر (روش بیشترین درست نمایی و روش بیزی در مدل های گرافی جهت دار و بدون جهت)
- یادگیری ساختار شبکه (بیشترین درست نمایی و سازگاری مجانبی، یادگیری ساختار درختی، روش های 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.



- روش‌های خوشه‌بندی داده‌های ابعاد بالا: روش‌های PROCLUS, CLIQUE, الگوهای مکرری
- 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 بُعد: 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] 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, ۲۰۱۴.
- [۴] M. Kantardzic, Data Mining and Analysis: Foundations and Algorithms, ۲nd Edition, Wiley-IEEE Press, ۲۰۱۱.
- [۵] M. J. Zaki, W. Meira, The Handbook of Data Mining, Cambridge University Press, ۲۰۱۴.
- [۶] 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, ۲۰۱۴.
- [۸] M. Sato-Ilic, L. C. Jain, "Innovations in Fuzzy Clustering, Theory and Applications", Springer, ۲۰۰۶.
- [۹] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, ۲nd edition, Springer, ۲۰۰۹.



عنوان درس		فارسی		دیداری سازی داده ها	
		انگلیسی		Data Visualization	
نوع واحد		تعداد واحد	تعداد ساعت	دروس پیش نیاز	
پایه	اصلی	تخصصی	اختیاری	۴۸	۳
نظری	عملی	نظری	عملی		
حل تمرین: حداکثر ۲۴ ساعت		نیاز به اجرای پروژه عملی: دارد			

هدف:

این درس روش ها و اصول کلیدی دیداری سازی داده ها را معرفی می کند و به یادگیرنده نشان می دهد که چگونه بازنمایی های دیداری می توانند در درک داده های پیچیده مفید باشند و در پی آن است که توانایی یادگیرنده را برای فعالیت و پژوهش در این حوزه ارتقا دهد.

سرفصل درس:

- معرفی دیداری سازی داده ها و اهمیت آن.
- انواع داده و دادگان (datasets)، روش های تجزید داده ها، روش های تجزید وظایف.
- نشانه ها و کانال ها، کدگذاری داده ها به کمک نشانه ها و کانال ها، معیارهای کارایی کانال: دقت (Accuracy)، افتراق پذیری (Discriminability)، جدایی پذیری (Separability)، آشکارسازی تنظیم پذیر (Popout) و تمامیت (Integrity).
- قواعدی که همواره در دیداری سازی داده ها باید مد نظر قرار داد.
- اعتبارسنجی و چهار سطح آن: وضعیت دامن، تجزید داده ها و وظایف، کدگذاری دیداری و تعامل، پیچیدگی الگوریتم از نظر زمانی و حافظه.
- طراحی چیدمان جدول های داده، طراحی چیدمان داده های مکانی، طراحی چیدمان داده های شبکه ای و ساختارهای درختی.
- نگاشت رنگ و سایر کانال های غیر فضایی در کدگذاری دیداری، روش های دستکاری نما و دید (view) برای کاهش پیچیدگی نمایش داده ها، روش های نمایش داده های پیچیده با تقسیم نمایش به چندین نما و دید یا لایه.
- روش های کاهش اقلام و کاهش ویژگی برای مواجهه با پیچیدگی های دیداری سازی.
- روش های درج اطلاعات بر روی مجموعه انتخاب شده در یک نما و دید.
- معرفی سیستم های دیداری سازی داده ها و قابلیت های آن ها، همچون: PivotGraph, VisDB, Scagnostics, Hierarchical Clustering Explorer و Constellation, InterRing.

منابع:

- [۱] T. Munzner, Visualization Analysis and Design, CRC Press, ۲۰۱۴.
- [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.
- [۴] I. Meirelles, Design for Information: An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations, Rockport Publishers, ۲۰۱۳.
- [۵] C. Ware, Visual Thinking for Design, Morgan Kaufman, ۲۰۰۸.
- [۶] Alberto Cairo, The Functional Art: An introduction to information graphics and visualization-, New Riders, ۲۰۱۲.



عنوان درس		فارسی	مدل سازی و پردازش مه داده ها
		انگلیسی	Modeling and Processing Big Data
نوع واحد		تعداد واحد	تعداد ساعات
پایه		۳	۴۸
اصلی		اختیاری	دروس پیش نیاز
نظری		عملی	نظری
عملی		نظری	عملی
حل تمرین: ندارد		نیاز به اجرای پروژه	
عملی: ندارد			

هدف:

این درس مفاهیم مربوط به مه داده ها و نحوه مدیریت و تحلیل آنها را معرفی می کند و دانش و بینش لازم را برای ورود به مباحث پیشرفته در این حوزه فراهم می آورد.

سرفصل درس:

- مقدمات (معرفی مه داده ها و مثال های آن، منشا تولید مه داده، ...)
- مشخصه های مه داده ها (حجم، سرعت، تنوع، صحت، ظرفیت و ارزش، تاثیر هر مشخصه بر جمع آوری، پایش، ذخیره سازی، تحلیل و گزارش دهی مه داده ها)
- مولفه های ساختاری و مدل های برنامه نویسی برای تحلیل مه داده های مقیاس پذیر
- روش های دسترسی و دستکاری داده های روان (تفاوت پایگاه داده های سنتی با سامانه های مدیریت مه داده ها)
- Hadoop (HDF) به عنوان نظام فایل توزیع شده در Hadoop، معرفی YARN مدیریت منابع در Hadoop، معرفی MapReduce به عنوان مدل برنامه نویسی)
- بازیابی، یکپارچه سازی و تحلیل مه داده ها
- تحلیل مه داده ها (دسته بندی، رگرسیون، خوشه بندی و تحلیل هم پاشی مه داده ها)
- تحلیل مه داده های گرافیکی
- پیاده سازی (پیاده سازی نمونه ای از سامانه مدیریت مه داده ها 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		گراف کاوی		
		انگلیسی		Graph Mining		
نوع واحد				تعداد واحد	تعداد ساعت	دروس پیش نیاز
پایه		اصلی شاخه		تخصصی		اختیاری
نظری	عملی	نظری	عملی	نظری	عملی	داده کاوی یا اجازه استاد
حل تمرین: حداکثر ۲۴ ساعت						

هدف:

این درس روش های استخراج الگو از داده های گرافی را مورد توجه قرار می دهد. برای این منظور، ابتدا دانش پیش زمینه مورد نیاز از جبری خطی و نظریه گراف مرور می شود و سپس مسائل مطرح در زمینه های گراف کاوی و الگوریتم های مربوط معرفی می شوند تا توانایی لازم برای پژوهش در این حوزه کسب شود.

سرفصل درس:

- گراف ها، پایگاه داده های گرافی، تحلیل جبری گراف ها (ماتریس مجاورت، ماتریس لاپلاسی، ماتریس شباهت)، محاسبه فاصله در گراف ها، تطبیق گراف ها،
- گراف های ایستا: قوانین و الگوها،
- گراف های پویا: قوانین و الگوها،
- گراف های تصادفی،
- گام برداری تصادفی، الگوریتم های PageRank و HITS،
- روش های کاهش بعد در داده های گرافی،
- خوشه بندی و دسته بندی گراف ها،
- یادگیری نیمه نظارت شده در گراف ها،
- خلاصه سازی گراف ها،
- کشف ناپهنجاری (Anomaly Detection) در گراف ها،
- تحلیل پیوند (Link Analysis)،
- زیرگراف های پرتکرار و روش های کاوش آن ها،
- کاوش زیرگراف های چگال،
- استخراج الگو از داده های گرافی روان (Stream Graph Mining)،
- تمایز کاوی (Contrast Mining) در گراف ها،
- روش های هسته برای گراف ها،
- یادگیری ژرف در گراف ها،
- ارائه مثال هایی از کاربرد گراف کاوی در داده های زیستی، داده های شیمیایی و شبکه های اجتماعی.



- [^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, ۲۰۱۴.
- [^۴] C.C. Aggarwal, H. Wang, Managing and Mining Graph Data, Springer, ۲۰۱۰.
- [^۵] G. Dong, J. Bailey, Contrast Data Mining: Concepts, Algorithms, and Applications, CRC Press, ۲۰۱۳.
- [^۶] D.J. Cook, L.B. Holder, Mining Graph Data, Wiley, ۲۰۰۷.
- [^۷] A. Schenker, Graph-Theoretic Techniques for Web Content Mining, World Scientific, ۲۰۰۵.



عنوان درس		فارسی		متن کاوی و وب کاوی	
		انگلیسی		Text Mining and Web Mining	
نوع واحد		تعداد واحد	تعداد ساعت	دروس پیش نیاز	
پایه	اصلی شاخه	تخصصی		اختیاری	
		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 متن کاوی و وب کاوی و الگوریتم‌های مربوط می‌پردازد تا توانایی لازم برای پژوهش در این حوزه‌ها کسب شود.

سرفصل درس:

- معرفی متن کاوی، معرفی وب کاوی و شاخه‌های آن (محتواکاوی، پیوند کاوی و کاربرد کاوی)، تفاوت‌های داده کاوی با متن کاوی و وب کاوی، مرزهای مشترک بازاریابی اطلاعات با متن کاوی و وب کاوی.
- پیش‌زمینه‌های جبری و آماری:
 - مدل سازی فضای برداری (VSM)،
 - توابع شباهت: همبستگی، شباهت کسینوسی،
 - فایله معکوس و فشرده سازی آن،
 - تجزیه SVD و روش‌های شاخص گذاری معنایی LSA (یا LSI) و pLSA.
 - مدل سازی موضوعی (Topic Modeling) و تخصیص پنهان دیریکله (LDA) و گونه پرخط آن (OLDA).
 - پربار کردن ویژگی‌ها در متون کوتاه به کمک تحلیل شاخص گذاری معنایی و بازخورد از کاربران.
- متن کاوی:
 - پیش پردازش و استخراج ویژگی در اسناد متنی،
 - خوشه بندی و دسته بندی اسناد متنی،
 - روش‌های یادگیری فعال و یادگیری نیمه نظارت شده در دسته بندی اسناد متنی،
 - خلاصه سازی متن،
- وب کاوی:
 - محتواکاوی:
 - استخراج موضوع صفحات وب به طور خودکار به کمک مدل سازی موضوعی،
 - خوشه بندی و دسته بندی صفحات وب، استفاده از یادگیری نیمه نظارت شده در دسته بندی صفحات وب،



- 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 خوشه‌بندی.
 - 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 پیشنهادده: روش‌های محتوا-پایه، روش‌های پالایش همکارانه، 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.
- [^۴] S. M. Weiss, N. Indurkha, T. Zhang, Fundamentals of Predictive Text Mining, Springer, ۲۰۱۰.
- [^۵] M. W. Berry, J. Kogan, Text Mining: Applications and Theory, John Wiley & Sons, ۲۰۱۰.
- [^۶] 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 ساعات	دروس پیش نیاز
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 شبکه‌های عصبی طبیعی، مفاهیم، تعاریف، و بخش‌های سازنده شبکه‌های عصبی.
- معرفی پرسپترون، شبکه تک لایه پرسپترون، 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 تداعی گر پیش‌خور، شبکه‌های تداعی گر بازگشتی تک لایه، شبکه‌های تداعی گر 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 سیگنوییدی، 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		مباحث ویژه در علوم داده							
		انگلیسی		Special Topics in Data Science							
نوع واحد		تعداد واحد		تعداد ساعات		دروس پیش نیاز					
پایه	نظری	عملی	نظری	عملی	نظری	عملی	اختیاری	۳	۴۸	اجازه گروه	
											نظری
حل تمرین: ندارد		نیاز به اجرای پروژه									اجازه گروه
عملی: ندارد											

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

