



دانشکده بهداشت

دفتر توسعه آموزش

طرح دوره (Course Plan)

نام درس: روشهای داده کاوی و یادگیری آماری	کد درس: ۱۹۱۰۵۸
مقطع تحصیلی: دکتری تخصصی	رشته تحصیلی: آمارزیستی
گروه آموزشی: آمارزیستی	پیش نیاز: اصول و روشهای داده کاوی
تعداد واحد: ۳ واحد تئوری	محل برگزاری: کلاس ... دانشکده بهداشت
روز و ساعت برگزاری: شنبه و سه شنبه ۸ تا ۱۰	مسئول درس: دکتر جمشید جمالی
طراحی اولیه بازنگری	
هدف کلی دوره: این درس با هدف کمک به دانشجویان تحصیلات تکمیلی آمارزیستی جهت فراگیری روشهای داده کاوی و یادگیری آماری تدوین گردیده است.	
اهداف اختصاصی دوره: ✓ آشنایی با نظریه یادگیری ماشین، روشهای پیش بینی، مدلسازی با راهنما و بدون راهنما ✓ آشنایی با مفهوم یادگیری در داده های با بعد بالا ✓ معرفی مقدماتی مفهوم انبار داده و استخراج داده ✓ معرفی روشهای کاهش بعد و کاهش داده ها ✓ آشنایی با مفاهیم و تکنیکهای ارتباط همبستگی و تجزیه و تحلیل الگوهای مکرر، استخراج انواع مختلف الگوهای مکرر ✓ معرفی مفاهیم و تکنیک های تجزیه و تحلیل طبقه بندی و ارزیابی طبقه ها ✓ آشنایی با الگوهای تصمیم گیری (مدلهای خطی، غیر خطی، درخت تصادفی، ماشین بردار پشتیبان، شبکه عصبی) و قاعده تصمیم گیری (درست نمایی، بیزی، آنتروپی، نزدیکترین همسایه، جینی، Sequential Covering Ruled-Based LAZY (Algorithm) ✓ آشنایی با مفاهیم و خوشه بندی (مبتنی بر پارتیشن، سلسه مراتبی، مبتنی بر تراکم، مبتنی بر الگوریتم EM، براساس مدل شبکه عصبی، طیفی، مقید، نیمه نظارتی، فازی، فازی بر پایه هسته kernel-based، فازی خود سازمانده، روش Gustafson-Kessel)، تجزیه و تحلیل خوشه ها و تحلیل داده های پرت ✓ آشنایی با خوشه بندی مبتنی بر شبکه، خوشه بندی داده ها با بعد بالا ✓ آشنایی با روش های استخراج مجموعه های زمانی، سری های زمانی و داده های توالی زمانی ✓ آشنایی با داده کاوی در داده های جاری (Stream)، تشخیص رانش (drift) در داده های جاری، شناسایی الگوهای پرتکرار در داده های جاری، طبقه بندی و خوشه بندی داده های جاری، تشخیص داده دور افتاده در داده های جاری ✓ آشنایی با داده کاوی در داده های فضایی، فضایی-زمان ✓ آشنایی با گراف کاوی، روشهای کاوش زیر گراف های پرتکرار، طبقه بندی و خوشه بندی داده های گرافی	
جدول ترتیب ارائه محتوای آموزشی در طول دوره	
جلسه	رئوس مطالب



دانشکده بهداشت

دفتر توسعه آموزش

طرح دوره (Course Plan)

اول	نظریه یادگیری ماشین و ارائه مثالهای از آن
دوم	روشهای پیش بینی و ارائه مثالهای از آن
سوم	مدلسازی با راهنما و بدون راهنما و ارائه مثالهای از آن
چهارم	مفهوم یادگیری در داده های با بعد بالا و ارائه مثالهای از آن
پنجم	معرفی مقدماتی مفهوم انبار داده و استخر داده و ارائه مثالهای از آن
ششم	معرفی روشهای کاهش بعد و ارائه مثالهای از آن
هفتم	معرفی روشهای کاهش داده ها و ارائه مثالهای از آن
هشتم	آشنایی با مفاهیم و تکنیکهای ارتباط همبستگی و تجزیه و تحلیل الگوهای مکرر و ارائه مثالهای از آن (یا ارائه پروژه توسط دانشجویان)
نهم	استخراج انواع مختلف الگوهی مکرر و ارائه مثالهای از آن
دهم	معرفی مفاهیم و تکنیک های تجزیه و تحلیل طبقه بندی و ارائه مثالهای از آن
یازدهم	معرفی روشهای ارزیابی طبقه ها
دوازدهم	الگوهای تصمیم گیری (مدلهای خطی، غیرخطی، درخت تصادفی، ماشین بردار پشتیبان، شبکه عصبی) و ارائه مثالهای از آن
سیزدهم	الگوهای تصمیم گیری (مدلهای خطی، غیرخطی، درخت تصادفی، ماشین بردار پشتیبان، شبکه عصبی) و ارائه مثالهای از آن (یا ارائه پروژه توسط دانشجویان)
چهاردهم	قاعده تصمیم گیری (درست نمایی، بیزی، آنتروپی، نزدیکترین همسایه، جینی، Sequential Covering Ruled-Based LAZY Algorithm)
پانزدهم	آشنایی با مفاهیم و خوشه بندی (مبتنی بر پارتیشن، سلسه مراتبی، مبتنی بر تراکم، مبتنی بر الگوریتم EM، براساس مدل شبکه عصبی، طیفی، مقید، نیمه نظارتی، فازی، فازی بر پایه هسته kernel-based، فازی خود سازمانده، روش Gustafson-Kessel) و ارائه مثالهای از آن
شانزدهم	تجزیه و تحلیل خوشه ها و تحلیل داده های پرت و ارائه مثالهای از آن
هفدهم	خوشه بندی مبتنی بر شبکه و ارائه مثالهای از آن
هیجدهم	خوشه بندی داده ها با بعد بالا و ارائه مثالهای از آن
نوزدهم	آشنایی با روش های استخراج مجموعه های زمانی، سری های زمانی و داده های توالی زمانی و ارائه مثالهای از آن (یا ارائه پروژه توسط دانشجویان)
بیستم	داده کاوی در داده های جاری (Stream)، تشخیص رانش (drift) در داده های جاری
بیست و یکم	شناسایی الگوهای پرتکرار در داده های جاری، طبقه بندی و خوشه بندی داده های جاری
بیست و دوم	تشخیص داده دور افتاده در داده های جاری
بیست و سوم	داده کاوی در داده های فضایی، فضایی-زمان و ارائه مثالهای از آن (یا ارائه پروژه توسط دانشجویان)



دانشکده بهداشت

دفتر توسعه آموزش

طرح دوره (Course Plan)

بیست و چهارم	گراف کاوی، روشهای کاوش زیرگراف های پرتکرار، طبقه بندی و خوشه بندی داده های گرافی
روش های آموزشی (نحوه ارائه درس): سخنرانی، ارائه مطالب توسط دانشجویان و نقد آن، پرسش و پاسخ	
وسایل آموزشی: کامپیوتر به تعداد دانشجویان و مدرس، ویدیو پروژکتور، وایت بورد و مژیک	
وظایف و تکالیف دانشجویان: مرور مطالب بیان شده در جلسات قبلی، آماده سازی و ارائه مطلب در کلاس در صورت نیاز، حضور فعالانه در فعالیتهای کلاسی همگام با مدرس	
روش های ارزیابی دانشجویان (لطفا سهم هر روش بر حسب درصد نوشته شود):	
✓ روش های ارزیابی در طول دوره: ارائه دانشجویان/پرسش و پاسخ (۵۰٪ نمره)	
✓ روش های ارزیابی در انتهای دوره: امتحان تئوری و نظری انتهای دوره (۵۰٪ نمره)	
منابع اصلی درس: مطالب آموزشی موجود در کتب زیر	
✓ Hastie T, Tibshirani R, and Friedman J. The Elements of Statistical Learning: data mining, inference and prediction. Springer, Last Edition. ✓ Goodfellow I, Bengio Y, Courville A. Deep Learning (Adaptive Computation and Machine Learning series). MIT press, Last Edition. ✓ J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, Elsevier Inc., Last Edition. ✓ C. C. Aggarwal, Data Mining: The Textbook, Springer, Last Edition. ✓ M. J. Zaki, W. Meira. The Handbook of Data Mining. Cambridge University Press, Last Edition. ✓ C. C. Aggarwal. Data Streams: Models and Algorithms. Springer. Last Edition. ✓ Bifet, Adaptive Stream Mining: Pattern Learning and Mining from Evolving Data Streams, IOS Press. Last Edition ✓ Canlas RD. Data Mining in Healthcare: Current Applications and Issues. Last Edition. ✓ Llatas, C.Fernandez, G.Gomez, J.Miguel. Data Mining in Clinical Medicine, Springer. Last Edition.	