



یادگیری ماشین Machine Learning

پیش نیاز: آشنایی با مبانی جبر خطی،
نظریه احتمال و آمار و برنامه نویسی

تعداد واحد: ۳

نوع درس: نظری

مخاطب: دانشجویان کارشناسی ارشد مهندسی صنایع گرایش مدل سازی سیستم و تحلیل داده (درس اصلی)
دانشجویان کارشناسی ارشد و دکتری تخصصی سایر گرایش های مهندسی صنایع (درس اختیاری)

Email : nikbakhsh@modares.ac.ir

مدرس: احسان نیک بخش

زمان و مکان برگزاری کلاس درس (زمستان ۱۴۰۲): سه شنبه ها و چهارشنبه ها (کلاس ۶/۰۶، ۹:۲۰-۸:۰۰)

هدف درس: آشنایی با مبانی نظری و کاربردی یادگیری ماشینی و استفاده از آن در تحلیل داده های کسب و کار

معرفی درس

در دنیای امروزی، یادگیری ماشینی به عنوان یک تکنیک قدرتمند در حوزه تحلیل داده و هوش مصنوعی به شدت مورد توجه قرار گرفته است. مدل ها و روش های این حوزه به دلیل اینکه به ما امکان می دهند با استفاده از داده ها، پیش بینی های دقیق تری انجام دهیم و با کشف الگوها و روابط پنهان در داده ها اطلاعات مفیدی را استخراج کنیم، از اهمیت بسیاری برخوردارند. این مدل ها در بسیاری از صنایع و حوزه ها مورد استفاده قرار می گیرند، که از آن جمله می توان به کسب و کار، مهندسی، مدیریت، پزشکی، تحقیقات علمی و هوش مصنوعی اشاره کرد. از مثال های کاربردهای این حوزه می توان به تشخیص الگوهای تقلب در تراکنش های مالی، تحلیل رفتار مشتریان، پشتیبانی مشتری، تحلیل سیستم های تولیدی، ارزیابی عملکرد تأمین کنندگان، تقاضای انرژی و تحلیل داده ها و تصاویر پزشکی اشاره کرد. در این درس، در ابتدا مباحث پایه ای بر تمرکز بر مبانی تحلیل داده و یادگیری ماشینی، مدل های یادگیری نظارتی و غیرنظارتی، چرخه فرایندی تحلیل داده، پیش پردازش داده ها و استخراج ویژگی ها بررسی خواهند شد. در ادامه نحوه ارزیابی عملکرد مدل های یادگیری در فرایند تحلیل داده مورد بحث قرار می گیرد. سپس مدل های متنوعی در دو حوزه یادگیری نظارتی و غیرنظارتی معرفی خواهند شد. سرانجام، پیاده سازی برخی از این روش ها در زبان برنامه نویسی پایتون به دلیل قدرت و انعطاف پذیری آن پیگیری خواهد شد.

مهم ترین مباحث درسی

- ❖ تحلیل داده، یادگیری ماشینی و کاربردهای آن
- ❖ یادگیری نظارتی (دسته بندی) و غیرنظارتی (خوشه بندی)
- ❖ چرخه فرایندی تحلیل داده
- ❖ داده و پیش پردازش داده ها
- ❖ مدل های استخراج ویژگی و کاهش بعد

- ❖ مدل‌های یادگیری نظارتی (رگرسیون، درخت تصمیم، جنگل تصادفی، شبکه‌های عصبی مصنوعی، ماشین‌های بردار پشتیبان، معیارهای ارزیابی یادگیری نظارتی)
- ❖ مدل‌های یادگیری غیرنظارتی (k-Means، مدل‌های سلسله‌مراتبی، DBSCAN، ارزیابی خوشه‌ها)
- ❖ (در صورت وجود زمان) کشف الگو با استفاده از مدل‌های کاوش اقواعد انجمنی (الگوهای تقریبی، محدودیت‌دار، متوالی و زیرگراف‌ها)
- ❖ (در صورت وجود زمان) یادگیری تقویتی
- ❖ (در صورت وجود زمان) تحلیل ناهنجاری داده‌ها
- ❖ پیاده‌سازی بخشی از روش‌های مطرح در درس در زبان برنامه‌نویسی پایتون
- ❖ آشنایی با آخرین تحقیقات و کاربردها (ارائه دانشجویان)

مراجع درس

- ❖ Bishop, C.M. (2006) *Pattern Recognition and Machine Learning*. Springer-Verlag, New York.
- ❖ James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023) *An Introduction to Statistical Learning with Applications in Python*. Springer, Switzerland.
- ❖ Han, J. Pei, J. & Tong, H. (2022) *Data Mining: Concepts and Techniques* (4th ed.). Morgan Kaufmann, Cambridge, MA.
- ❖ Alpaydin, E. (2020) *Introduction to Machine Learning* (4th ed.). The MIT Press, Cambridge, MA.
- ❖ Marsland, S. (2015) *Machine Learning: An Algorithmic Perspective* (2nd ed.). CRC Press, Boca Raton, FL.
- ❖ Mitchell, T. (1997) *Machine Learning*. McGraw-Hill.
- ❖ مقالات منتخب در حوزه‌های داده کاوی، یادگیری ماشین، و تحلیل داده

ارزشیابی

- ❖ تکالیف: ۲ نمره (سه مجموعه)
- ❖ سمینار: ۴ نمره (ارائه یک مقاله کاربردی در زمینه استفاده از ابزارهای یادگیری ماشین در تحلیل داده‌های مهندسی و کسب و کار)
- ❖ پروژه: ۴ نمره (پیاده‌سازی یک الگوریتم دسته‌بندی بر روی یک مجموعه داده به انتخاب دانشجویان در گروه‌های دو نفره)
- ❖ امتحان نهایی: ۱۰ نمره

سایت‌های مرتبط با درس

- ❖ وب‌سایت دوره: قابل دسترسی از طریق سامانه الکترونیکی دروس به آدرس lms.modares.ac.ir
- ❖ Kaggle - Machine Learning and Data Science Community: <https://www.kaggle.com/>
- ❖ UC Irvine Machine Learning Repository: <https://archive.ics.uci.edu/>
- ❖ TensorFlow – An Open-Source Library for ML Models: <https://www.tensorflow.org/>
- ❖ Machine Learning Open-Source Software: <https://mloss.org/software/>

نرم‌افزارهای مرتبط با آموزش درس

- ❖ Python: <https://www.python.org>
- ❖ R: <https://www.r-project.org>
 - For a better experience, install RStudio too: <https://www.rstudio.com>
- ❖ با اینکه تمرکز درس روی زبان برنامه‌نویسی پایتون خواهد بود، می‌توانید از زبان آماری R هم برای مدل‌سازی و تحلیل داده بهره بگیرید.
- ❖ Instead of installing Python and R separately, I recommend to install Anaconda, a Python/R distribution dedicated to scientific computation: <https://www.anaconda.com/distribution/>
- ❖ چند نکته در خصوص یادگیری نرم‌افزارها و زبان‌های برنامه‌نویسی جدید:

- یادگیری هر موضوع جدیدی مستلزم دقت، تلاش و صبر است. بهترین همراه شما در مسیر یادگیری نرم‌افزارها و زبان‌های برنامه‌نویسی جدید جستجو در اینترنت است. مطمئن باشید شما اولین نفری نیستید که به یک سؤال یا مشکل خاص در پایتون بر می‌خورید و قبلاً افرادی به سؤال شما در قالب یک انجمن یا پست روی وبسایت‌ها و وبلاگ‌ها پاسخ داده‌اند.
- وبسایت‌های متعددی در زمینه آموزش مبانی پایتون و R و کاربرد آنها در تحلیل داده وجود دارند. از جمله بهترین وبسایت‌های موجود می‌توان به *edX*، *Coursera*، *Udemy* و *DataCamp* اشاره کرد. علاوه بر این وبسایت‌ها، می‌توانید از نرم‌افزار نسخه تلفن همراه آنها و یا سایر شرکت‌ها برای یادگیری در زمان‌های آزاد خودتان بهره ببرید.

فهرستی مختصر و نه الزاماً کامل از مجله‌های معتبر حوزه‌های تحلیل داده و یادگیری ماشینی

- ❖ Annals of Statistics
- ❖ Engineering applications of artificial intelligence
- ❖ Expert systems with applications
- ❖ IEEE Transactions on Neural Networks
- ❖ IEEE Transactions on Pattern Analysis and Machine
- ❖ Information sciences
- ❖ Intelligence
- ❖ Journal of Machine Learning Research www.jmlr.org
- ❖ Journal of the American Statistical Association
- ❖ Knowledge based systems
- ❖ Machine Learning
- ❖ Neural Computation
- ❖ Neural Networks
- ❖ Pattern Recognition
- ❖ Pattern Recognition letters

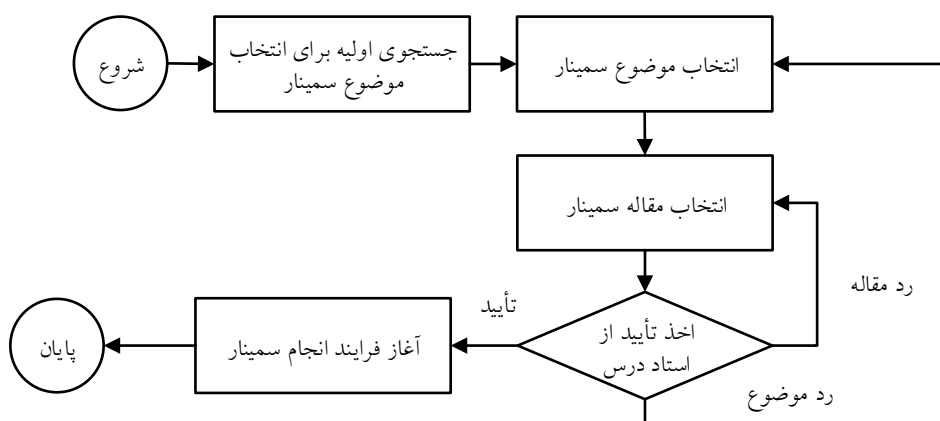
تمرین

این درس شامل چندین مجموعه تمرین است که انجام آنها به صورت انفرادی خواهد بود. موعد تحویل پاسخ هر مجموعه تمرین ابتدای جلسه درسی روز موعد تحویل الکترونیکی از طریق وبسایت دوره است.

❖ در صورتیکه بخشی از تمرین‌ها به صورت دست‌نویس قابل تحویل بودند، حتماً از اسکنر حرفه‌ای یا نرم‌افزارهای اسکنر موبایل (مانند Adobe Scan، Office Lens یا CamScanner) استفاده کنید تا حجم فایل ارسالی کاهش و کیفیت آن افزایش یابد.

سمینار

این درس دارای یک سمینار خواهد بود که در آن دانشجویان باید در یکی از زمینه‌های مسایل مطرح در یادگیری ماشینی بر اساس آخرین تحقیقات چاپ شده در مجلات علمی معتبر بین‌المللی یک مقاله را نقد و بررسی نموده و نتیجه را در کلاس درس ارائه دهند. هدف از این سمینار آشنایی دانشجویان با آخرین مباحث نظری و عملی مطرح در سطح مجامع آکادمیک بر اساس آخرین مقالات انتشار یافته در این حوزه است. به این منظور، دانشجویان باید یک مقاله کاربردی در حوزه مسایل مهندسی صنایع را انتخاب نمایند که تاریخ انتشار آن از سال ۲۰۲۰ به بعد باشد. ارائه گزارش سمینار به صورت یک ارائه عمومی در کلاس درس و در آخرین هفته کلاس‌های درس طبق تقویم آموزشی دانشگاه خواهد بود. دانشجویان باید حداکثر تا انتهای هفته ششم نیم‌سال تحصیلی، یک مقاله را برای انجام سمینار درس به تائید استاد درس برسانند. شایان ذکر است که مقالات مروری، کنفرانسی، و فارسی قابل انتخاب به عنوان مقاله اصلی نیستند و در صورت نیاز صرفاً می‌توان از آنها به عنوان مقالات کمکی استفاده نمود.



فرایند نگارش و اخذ تأیید مقاله سمینار درس

پروژه

این درس دارای یک پروژه است که شامل طراحی و پیاده‌سازی یکی از روش‌های دسته‌بندی در درس برای یک مجموعه داده به انتخاب دانشجو است. هدف از پروژه طی نمودن چرخه کامل تحلیل داده با استفاده از روش‌های یادگیری ماشینی نظارتی در محیط زبان برنامه‌نویسی پایتون است. دانشجویان در گروه‌های حداکثر دو نفره باید تا انتهای هفته ششم نیم‌سال تحصیلی مجموعه داده مورد استفاده در پروژه را به تأیید استاد درس برسانند. با هدف یادگیری بهتر دانشجویان، انجام پروژه به صورت تدریجی در طول نیم‌سال تحصیلی و همزمان با پیشرفت مباحث درسی و کلاس‌های حل تمرین خواهد بود. حداکثر مهلت تحویل گزارش نهایی پروژه هفته آخر کلاس‌های درس در نیم‌سال تحصیلی جاری (طبق تقویم آموزشی دانشگاه) خواهد بود. علاوه بر تحویل گزارش نهایی در وب‌سایت دوره، اعضای گروه باید به صورت حضوری نتیجه پروژه خود را برای استاد درس ارائه کنند.

تاریخ‌های مهم

- ❖ موعد تصویب مقاله سمینار و داده‌های پروژه: ۲۱ فروردین ۱۴۰۳
- ❖ موعد ارائه سمینار: دو هفته آخر درس (جلسات مورخ ۲۲، ۲۳، ۲۹ و ۳۰ خرداد ۱۴۰۳)
- ❖ امتحان نهایی: ۹ تیر ۱۴۰۳ ساعت ۱۱:۰۰
- ❖ موعد ارائه گزارش نهایی پروژه (به صورت الکترونیکی همراه با ارائه حضوری): هفته آخر کلاس‌ها (هفته منتهی به ۳۰ خرداد ۱۴۰۳)
- ❖ جلسات تعطیل: چهارشنبه ۲۲ فروردین و سه‌شنبه ۱۵ خرداد

تذکرات مهم

۱. علی‌رغم آنکه حضور در کلاس مستقیم در ارزیابی دانشجویان منظور نمی‌گردد، ولی حضور فعال و مشارکت در مباحث مطرح شده در کلاس درس نقش مهمی در یادگیری و کسب نمره نهایی بالاتر دارد. لذا به دانشجویان قویاً توصیه می‌شود که بدون عذر موجه از تأخیر یا غیبت خودداری نمایند.
۲. مسئولیت عدم تداخل زمان‌بندی امور مختلف درس (مانند تمرین‌ها، سمینار و به صورت خاص پروژه) یا دیگر تعهدات دانشجو در سایر دروس در طول نیم‌سال تحصیلی بر عهده دانشجو است. لذا با توجه به رویکرد معرفی نمودن تدریجی انتظارات درس در طول نیم‌سال تحصیلی، توصیه اکید می‌شود انجام تعهدات خود را به آخرین لحظه واگذار نکنید و به تاریخ‌های ذکر شده به عنوان مهلت تحویل و نه زمان انجام نگاه کنید.
۳. انجام هر گونه عمل مغایر با اصول اخلاق علمی در هر یک از مراحل دوره (تمرین، پروژه، سمینار، و امتحان نهایی) موجب از دست دادن کامل نمره مربوطه و نیز برخورد با دانشجو طبق قوانین انضباطی دانشکده/دانشگاه می‌گردد.
لطفاً در هنگام ارسال ایمیل به استاد درس، در ابتدای قسمت موضوع ایمیل خود عبارت ML02 را وارد نمایید.