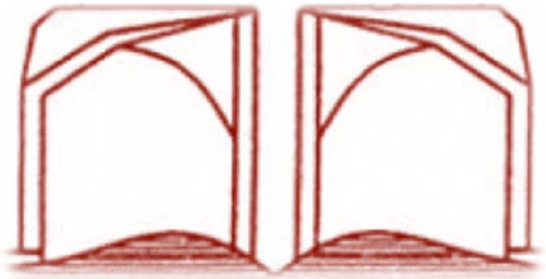


روش‌های یادگیری ماشین در پردازش زبان طبیعی

Machine Learning Methods in Natural Language Processing

هادی ویسی

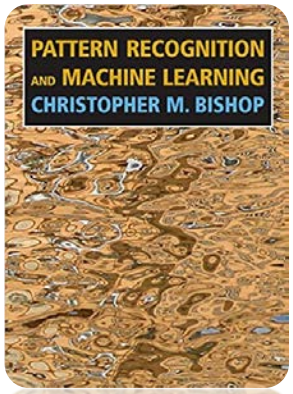
h.veisi@ut.ac.ir

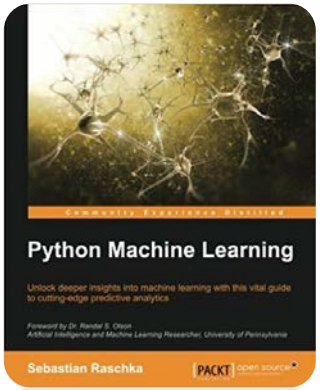
دانشگاه تهران - دانشکده علوم و فنون نوین

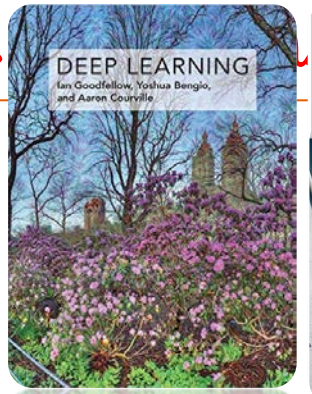
نیم‌سال اول ۱۴۰۱-۱۴۰۰

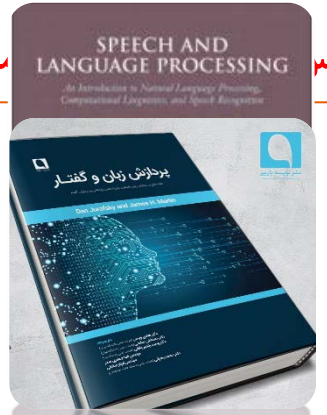
معرفی درس ...

- **زمان و مکان**
  - شنبه و دوشنبه، ساعت ۸:۰۰ الی ۱۰:۰۰، دانشکده علوم و فنون نوین
- **وب سایت**
  - dsp.ut.ac.ir
- **هدف**
  - مرور روش‌های یادگیری ماشین در پردازش زبان طبیعی
    - مفاهیم یادگیری ماشین
    - مفاهیم پایه آمار و احتمال، نظریه اطلاعات و روش‌های تخمین
    - روش‌های تشابه یابی متن و دسته بندی متن
    - شبکه های عصبی مصنوعی و یادگیری عمیق
    - مرور نمونه کاربردها
  - فعالیت‌های تمرینی با رویکرد کاربردی









منابع

- Christopher Bishop, Pattern Recognition and Machine Learning, Springer, 2006
- Raschka, Sebastian. *Python machine learning*. Packt Publishing Ltd, 2015.
- هادی ویسی، کبری مفاخری، سعید باقری شورکی، مبانی شبکه های عصبی: معماری، الگوریتم‌ها و کاربردها، انتشارات نص، چاپ پنجم، زمستان ۱۳۹۹
  - ترجمه Laurene Fausette, Fundamentals of neural networks, architecture, algorithms and application, Prentice Hall, 1994
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, 2016.
- هادی ویسی، مصطفی صالحی، وحید رنجبر، الما جعفری صدر، فرنار صادقی، محمد بحرانی، پردازش زبان و گفتار، انتشارات نویسه پارسی، پاییز ۱۴۰۰

Daniel Jurafsky, James Martin, Speech and Language Processing, 2nd Edition, Prentice Hall, 2009.

معرفی درس ...

- **ارزیابی . . .**
  - **تمرین**
    - برای هر موضوع: ٤ یا ٥ تمرین
    - همفکری و همکاری در یافتن پاسخ سوال‌ها توصیه می‌شود
    - در صورت کپی بودن یکی یا چند مورد از پاسخ‌ها، کل نمره آن تمرین برای طرفین کپی در نظر گرفته نمی‌شود.
    - تمرین‌های دارای پیاده‌سازی، باید هم شامل کدها و هم شامل گزارش مربوطه باشد
    - تاخیر در تحویل
      - ارسال پاسخ حداکثر تا ساعت ۲۳:۵۹ مهلت تعیین شده
      - هر یک ساعت تاخیر در ارسال پاسخ‌ها (از یک ثانیه تا ۶۰ دقیقه!)، کسر یک درصد نمره آن تمرین به عنوان جریمه تاخیر
    - ارسال پاسخ تمرین‌ها
      - تنها به صورت الکترونیکی و به ایمیل استاد درس است.
      - همه فایل‌های مرتبط با یک تمرین را در یک فایل فشرده شده
      - فرمت نام‌گذاری فایل ارسالی: #ML4NLP_Family_StNo_HW
    - وزن تمرین‌های مختلف با هم برابر نیست

معرفی درس ...

- ارزیابی . . .
  - آزمونک (کویز)
    - نداریم (با توجه به غیرحضوری بودن دوره)
  - امتحان میان‌ترم
    - دوشنبه ۰۱/۰۹/۱۴۰۰ ساعت ۸:۰۰
  - امتحان پایان‌ترم
    - شامل کلیه مطالب تدریس شده: از جمله مطالب میان ترم
    - زمان: طبق برنامه دانشگاه

معرفی درس ...

- ارزیابی ...
  - پروژه: یک مورد
    - پروژه کاربردی دارای پیاده‌سازی در Python یا سایر زبان‌های برنامه‌نویسی
    - علاوه‌بر کد برنامه، گزارش مکتوب (به صورت تایپ شده)، داده‌ها و منابع هم تحویل گرفته می‌شود
    - در صورت مساعد شدن شرایط کرونایی، تحویل به صورت حضوری است
    - آخرین زمان تعیین موضوع پروژه: روز شنبه ۰۱/۰۸/۱۴۰۰
    - تحویل پروژه: اولین هفته بعد از آخرین امتحان پایان‌ترم (دوشنبه ۱۱/۱۱/۱۴۰۰)
    - موضوع الزاما مرتبط با یادگیری ماشین باشد
    - برخی موضوع‌های پیشنهادی
      - تشابه‌یابی متن با استفاده از نمایش‌های مبتنی بر یادگیری عمیق (مانند Bert)
      - تشخیص احساس در متن با استفاده از یادگیری عمیق
      - دسته‌بندی/خوشه‌بندی معنایی کلمات در یادگیری عمیق
      - تشخیص گفتار برای تعداد کلمات محدود
      - تبدیل متن به گفتار با استفاده از شبکه‌های عمیق مانند مبدل‌ها یا GAN
      - تولید خودکار متن (مانند متن یا شعر) با شبکه‌های عصبی عمیق

معرفی درس ...

- ارزیابی . . .
  - **حضور و مشارکت در کلاس (نمره اضافی)**
    - مشارکت در بحث‌های کلاس
  - **مقاله (نمره اضافی)**
    - مقاله ارسال شده مورد قبول است
    - به هر قیمتی مقاله ننویسید!
    - همکاری با این درس در نوشتن مقاله را به اطلاع استاد خود برسانید
  - **بازنگری نمره‌ها و برگه‌ها**
    - در زمان تحویل پروژه درس (به صورت حضوری)
      - دوشنبه ۱۱/۱۱/۱۴۰۰

معرفی درس ...

ارزیابی

عنوان	وزن	توضیح
تمرین	٪۵۰	بعد از هر موضوع (وزن تمرین‌ها برابر نیست)
آزمونک (کویز)	-	به دلیل غیرحضوری کلاس‌ها بودن ندارم
امتحان میان‌ترم	٪۱۵	دوشنبه ۰۱/۰۹/۱۴۰۰ ساعت ۸:۰۰
امتحان پایان‌ترم	٪۲۰	از کل مطالب درس، مطابق برنامه دانشگاه
پروژه	٪۱۵	موضوع اختیاری، تعیین موضوع تا شنبه ۰۱/۰۸/۱۴۰۰ تحویل پروژه: اولین هفته بعد از آخربن امتحان پایان‌ترم (دوشنبه ۱۱/۱۱/۱۴۰۰)
حضور و مشارکت کلاس (نمره اضافی)	٪۵	پاسخ دادن به سوالات حین تدریس و مشارکت در بحث‌های کلاس
مقاله (نمره اضافی)	٪۱۵	مقاله ارسال شده به مجله/کنفرانس مورد قبول است

معرفی درس ...

- سرفصل‌ها . . .
  - مروری بر مفاهیم و اصول یادگیری ماشین
  - مروری بر مبانی آمار و احتمال
    - احتمال (توام، شرطی)، امید ریاضی
    - قانون بیز
    - متغیر تصادفی
    - توابع توزیع
  - مروری بر نظریه اطلاعات و آنتروپی
  - مروری بر روش‌های تخمین
    - کمینه میانگین مربعات خطا (MMSE)
    - تخمین بیشینه شباهت (MLE)
    - تخمین بیز (Bayesian)

معرفی درس ...

- سرفصل‌ها . . .
  - **بازیابی اطلاعات و تشابه‌یابی متون**
    - نمایش کلمات و متن: تبدیل متن به بردار ویژگی
  - **دسته‌بندی متون با روش بیز (ساده)**
  - **شبکه عصبی مصنوعی**
    - مبانی و مفاهیم
    - شبکه عصبی پرسپترون
    - شبکه عصبی پرسپترون چندلایه (MLP)
      - نمایش کلمات/جمله/سند با بردار کلمات
    - یادگیری عمیق
      - شبکه خودرمزگذار، پیچشی (CNN) و شبکه مولد مقابله‌ای (GAN)
    - شبکه‌های عصبی بازگشتی (RNN)
      - شبکه حافظه کوتاه مدت ماندگار (LSTM)
      - مفهوم توجه (Attention)
    - مبدل‌ها (Transformer)

معرفی درس ...

- سرفصل‌ها
  - مدل مخفی مارکوف (HMM)
    - کاربرد در برچسپ‌زنی اجزای کلام (POS: Part-of-Speech tagging)
    - کاربرد در تشخیص گفتار (Speech Recognition)
  - روش‌های خوشه‌بندی
    - روش k-میانگین
    - الگوریتم امید-بیشینه (EM)

معرفی درس

زمان‌بندی

هفته	تاریخ	موضوع	توضیحات
۱	۰۳و۱۴۰۰/۰۷/۰۵	معرفی درس مروری بر مفاهیم یادگیری ماشین	
۲	۱۰و۱۴۰۰/۰۷/۱۲	مروری بر مفاهیم یادگیری ماشین	
۳	۱۷و۱۴۰۰/۰۷/۱۹	مروری بر مبانی آمار و احتمال	تمرین
۴	۲۴و۱۴۰۰/۰۷/۲۶	مروری بر نظریه اطلاعات و روش‌های تخمین	
۵	۰۱و۱۴۰۰/۰۸/۰۳	بازیابی اطلاعات و تشابه‌یابی متون	
۶	۰۸و۱۴۰۰/۰۸/۱۰	بازیابی اطلاعات و تشابه‌یابی متون	اعلام موضوع پروژه
۷	۱۵و۱۴۰۰/۰۸/۱۷	دسته‌بندی متون	تمرین
۸	۲۲و۱۴۰۰/۰۸/۲۴	شبکه عصبی مصنوعی	
۹	۱۴۰۰/۰۸/۲۹ ۱۴۰۰/۰۹/۰۱	شبکه عصبی پرسپترون چندلایه	آزمون میان ترم
۱۰	۰۶و۱۴۰۰/۰۹/۰۸	شبکه عصبی عمیق	تمرین
۱۱	۱۳و۱۴۰۰/۰۹/۱۵	شبکه عصبی عمیق: CNN	
۱۲	۲۰و۱۴۰۰/۰۹/۲۲	شبکه عصبی عمیق: GAN	
۱۳	۲۷و۱۴۰۰/۰۹/۲۹	شبکه عصبی بازگشتی و LSTM	
۱۴	۰۴و۱۴۰۰/۱۰/۰۶	مفهوم توجه در شبکه عصبی و مبدل‌ها	تمرین
۱۵	۱۱و۱۴۰۰/۱۰/۱۳	مدل مخفی مارکوف (HMM)	
۱۶	۱۸و۱۴۰۰/۱۰/۲۰	مدل مخفی مارکوف (HMM): کاربردها	
	۱۴۰۰/۰۷/۰۵: اربعین		

متناسب با شرایط و سطح کلاس، و همچنین تغییرات پیش‌بینی نشده در زمان‌بندی، ممکن است سرفصل مطالب و یا زمان‌بندی‌های کلاس مقداری تغییر داشته باشد