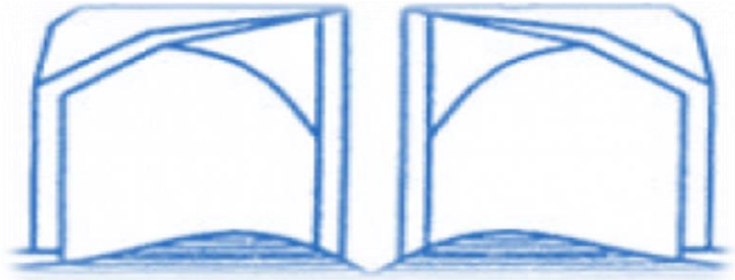


دانشکده علوم و فنون نوین

روش‌های یادگیری ماشین در پردازش زبان طبیعی

Machine Learning Methods in Natural Language Processing

هادی ویسی
h.veisi@ut.ac.ir

نیم‌سال اول ۱۴۰۳-۱۴۰۲

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 سنتی یادگیری ماشین
    - شبکه های عصبی مصنوعی و یادگیری عمیق
    - مرور نمونه کاربردها
  - فعالیت‌های تمرینی با رویکرد کاربردی

منابع

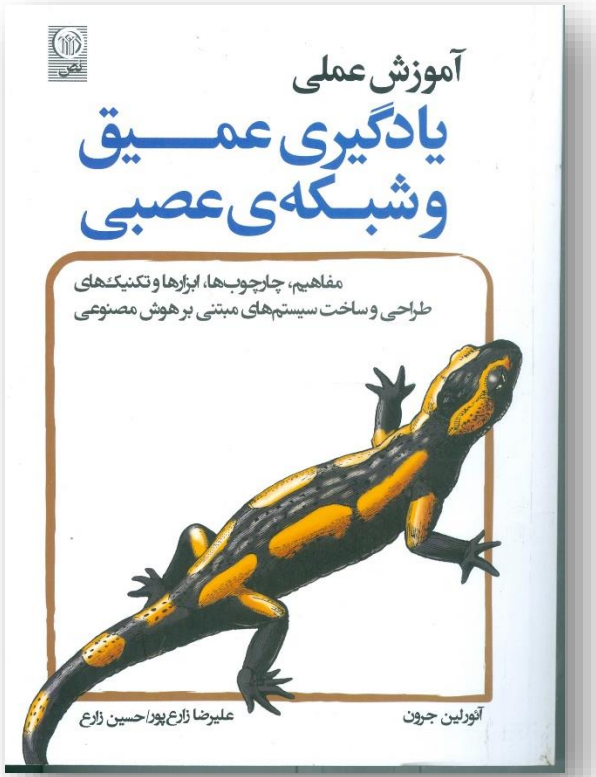
- Christopher Bishop, Pattern Recognition and Machine Learning, Springer, ۲۰۰۶
- Raschka, Sebastian. *Python machine learning*. Packt Publishing Ltd, ۲۰۱۵.
- هادی ویسی، کبری مفاخری، سعید باقری شورکی، مبانی شبکه های عصبی: معماری، الگوریتم‌ها و کاربردها، انتشارات نص، چاپ پنجم، زمستان ۱۳۹۹
- ترجمه Laurene Fausette, Fundamentals of neural networks, architecture, algorithms and application, Prentice Hall, 1994
- Ian Goodfellow, Yoshua Bengio, Aaron Courville, Deep Learning, MIT Press, ۲۰۱۶.
- هادی ویسی، مصطفی صالحی، وحید رنجبر، الما جعفری صدر، فرنار صادقی، محمد بحرانی، پردازش زبان و گفتار، انتشارات نویسه پارسی، پاییز ۱۴۰۰

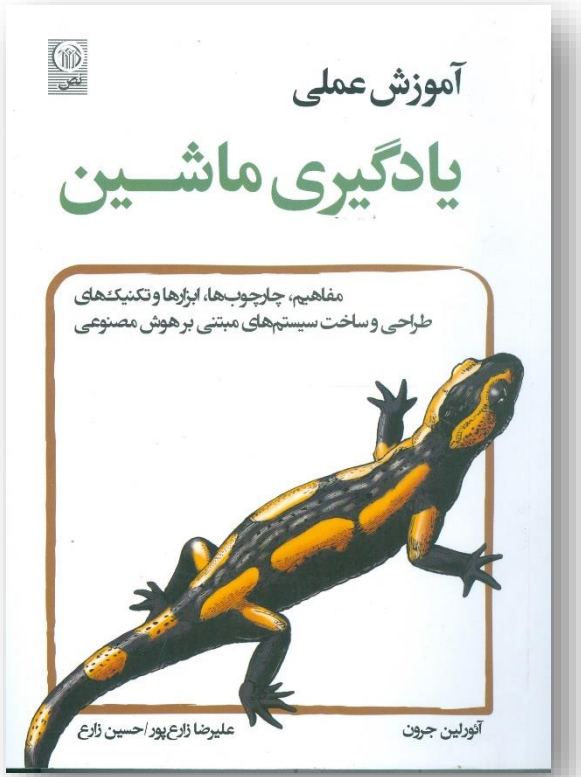
Daniel Jurafsky, James Martin, Speech and Language Processing, 2nd Edition, Prentice Hall, 2009.

Hadi Veisi (h.veisi@ut.ac.ir)

معرفی درس ...

- منابع





◀ Géron, Aurélien. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. " O'Reilly Media, Inc.", ۲۰۱۹

Hadi Veisi (h.veisi@ut.ac.ir)

معرفی درس ...

- ارزیابی . . .
  - تمرین
    - برای هر موضوع: ۴ یا ۵ 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
    - دیرکرد در تحویل
      - ارسال پاسخ حداکثر تا ساعت ۲۳:۵۹ مهلت تعیین شده
      - هر یک ساعت دیرکرد در ارسال پاسخ‌ها (از یک ثانیه تا ۶۰ دقیقه!)، کسر یک درصد نمره آن تمرین به عنوان جریمه دیرکرد
      - امکان بخشودگی یک مورد دیرکرد (برای یک تمرین) حداکثر به اندازه یک روز (۲۴ ساعت) به انتخاب دانشجو
    - 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)
    - یکی یا دو سوال در هر بار
    - ممکن است بدون اطلاع قبلی باشد
  - امتحان میان‌ترم
    - دوشنبه ۲۹/۰۸/۱۴۰۲ ساعت ۸:۰۰ (حضوری)
  - امتحان پایان‌ترم
    - شامل کلیه مطالب تدریس شده: از جمله مطالب میان ترم
    - زمان: طبق برنامه دانشگاه
  - بازنگری نمره‌ها و برگه‌ها
    - در زمان تحویل پروژه درس: اولین هفته بعد از آخرین امتحان پایان‌ترم

معرفی درس ...

- ارزیابی . . .
  - پروژه درس (اختیاری، نمره اضافی)
    - پروژه کاربردی دارای پیاده‌سازی در Python یا سایر زبان‌های برنامه‌نویسی
    - موضوع الزاما مرتبط با مطالب درس باشد
    - آخرین مهلت انتخاب موضوع: ۰۱/۰۹/۱۴۰۲
    - تحویل به صورت حضوری است
    - موارد تحویل دادنی:
      - کلیه کدهای پروژه
      - گزارش مکتوب (به صورت تایپ شده) شامل توضیح روش و جزئیات پیاده‌سازی و نتایج بدست آمده و تحلیل‌های مربوطه
      - داده‌های مورد استفاده در پروژه
      - مقاله‌ها و منابع مورد استفاده
    - بارم‌بندی نمرات:
      - انجام درست پیاده‌سازی و مرتب بودن کدها: ۵۰٪
      - کامل بودن گزارش (شامل نحوه استفاده از کد و مبانی علمی کار) و رعایت اصول نگارشی در آن: ۲۵٪
      - ارائه نتایج و تحلیل آن (در گزارش): ۲۵٪
    - تحویل پروژه: اولین هفته بعد از آخرین امتحان پایان‌ترم

معرفی درس ...

- ارزیابی . . .
  - ارائه شفاهی
    - هدف: آشنایی با مطالب به‌روز در حوزه درس
    - انتخاب یک موضوع مرتبط با مطالب درس
    - مطالعه منابع لازم و انجام یک ارائه کوتاه در کلاس
      - منابع متعلق به سه سال اخیر
    - زمان هر ارائه ۲۰ تا ۲۵ دقیقه
    - انتخاب موضوع با هماهنگی استاد
    - برخی موضوع های پیشنهادی
      - هوش مصنوعی اعتمادپذیر و لزوم آن در پردازش زبان طبیعی
      - مدل‌های زبانی بزرگ (LLM) :روش ساخت و کاربردها
      - مروری بر Bard و ChatGPT و نحوه ساخت آنها
      - مروری بر کتابخانه‌ها و ابزارهای مدرن در پردازش گفتار

معرفی درس ...

- ارزیابی . . .
  - پروژه درس (اختیاری، نمره اضافی)
    - 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
      - درک معنا با استفاده از مدل‌های زبانی بزرگ (LLM)

معرفی درس ...

- ارزیابی

عنوان	وزن	توضیح
تمرین	%۵۰	بعد از هر موضوع (وزن تمرین‌ها برابر نیست)
آزمونک (کویز)	%۵	ممکن است بدون اعلام قبلی باشد.
آزمون میان‌ترم	%۲۰	دوشنبه ۲۹/۰۸/۱۴۰۲ ساعت ۱۰:۰۰ (حضوری)
آزمون پایان‌ترم	%۲۰	از کل مطالب درس، مطابق برنامه دانشگاه
ارائه شفاهی	%۵	ارائه کلاسی از یک موضوع به‌روز
پروژه (نمره اضافی)	%۱۰	موضوع اختیاری، مرتبط با مطالب درس (آخرین مهلت انتخاب موضوع: ۰۱/۰۹/۱۴۰۲) تحویل پروژه: اولین هفته بعد از آخرین امتحان پایان‌ترم

Hadi Veisi (h.veisi@ut.ac.ir)

معرفی درس . . .

دستیار آموزشی

- امیرمحمد کویش‌پور
  - a.m.kouyeshpour@ut.ac.ir
- سیاوش حسین‌پور صفاریان
  - siavash.saffaria@ut.ac.ir
- برگزاری کلاس‌های حل تمرین و رفع اشکال
  - به ویژه راهنمایی پیاده‌سازی و کدنویسی

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
  - روش‌های سنتی یادگیری ماشین
    - بیز ساده
    - نزدیک‌ترین همسایه
    - رگرسیون
    - درخت تصمیم و جنگل تصادفی
    - ماشین بردار پشتیبان (SVM)

معرفی درس ...

- سرفصل‌ها . . .
  - شبکه عصبی مصنوعی و یادگیری عمیق
    - مبانی و مفاهیم
    - شبکه عصبی پرسپترون و آدالاین
    - 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)
      - سازوکار توجه (Attention)
    - مبدل‌ها (Transformer)
      - BERT
      - GPT

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)

Hadi Veisi (h.veisi@ut.ac.ir)

معرفی درس

دانشکده علوم و فنون نوین

زمان‌بندی

هفته	تاریخ	موضوع	توضیح
۱	۱۴۰۲/۰۷/۰۳و۰۱	معرفی درس مروری بر مفاهیم یادگیری ماشین	
۲	۱۴۰۲/۰۷/۱۰و۰۸	مروری بر مبانی آمار و احتمال	
۳	۱۴۰۲/۰۷/۱۷و۱۵	مروری بر نظریه اطلاعات و روش‌های تخمین	
۴	۱۴۰۲/۰۷/۲۴و۲۲	روش‌های سنتی: بیز ساده و KNN و رگرسیون	آزمونک
۵	۱۴۰۲/۰۷/۲۹ ۱۴۰۲/۰۸/۰۱	روش‌های سنتی: درخت تصمیم و SVM	تمرین
۶	۱۴۰۲/۰۸/۰۸و۰۶	روش‌های نمایش کلمات و اسناد: بازیابی اطلاعات و تشابه‌یابی متون	آزمونک
۷	۱۴۰۲/۰۸/۱۵و۱۳	شبکه عصبی مصنوعی مبانی، پرسپترون و آدالاین	
۸	۱۴۰۲/۰۸/۲۲و۲۰	شبکه عصبی پرسپترون چندلایه کاربردها و Wav2Vec	تمرین
۹	۱۴۰۲/۰۸/۲۹و۲۷	شبکه عصبی عمیق: خودرمزگذارها آزمون میان ترم	اعلام موضوع پروژه
۱۰	۱۴۰۲/۰۹/۰۶و۰۴	شبکه عصبی عمیق: CNN و GAN	
۱۱	۱۴۰۲/۰۹/۱۳و۱۱	شبکه عصبی بازگشتی	تمرین
۱۲	۱۴۰۲/۰۹/۲۰و۱۸	LSTM و سازوکار توجه	
۱۳	۱۴۰۲/۰۹/۲۷و۲۵	مبدل‌ها، Bert و GPT	آزمونک
۱۴	۱۴۰۲/۱۰/۰۴و۰۲	ارائه دانشجویان مدل مخفی مارکوف (HMM)	تمرین
۱۵	۱۴۰۲/۱۰/۱۱و۰۹	ارائه دانشجویان مدل مخفی مارکوف (HMM): کاربردها	
۱۶	۱۴۰۲/۱۰/۱۸و۱۶	روش‌های خوشه‌بندی	
۱۷	۱۴۰۲/۱۰/۲۵و۲۳	جبرانی (در صورت نیاز)	

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

Hadi Veisi (h.veisi@ut.ac.ir)