

سار پلیر
۲۰۷۲۲۸۴



مقدمہ اک بر یادگیرے ماشین

با کاربردہایے مبتنے بر Python و R

An Introduction to Machine Learning

With applications in R, Python

تألیف و ترجمہ:

دکتر محمود فتحی (استاد تمام دانشگاه)

علی محمد محمدی (کارشناس ارشد هوش مصنوعی)

ناشر:

کانون نشر علوم



نشر علوم

www.nashreloom.com

پیشگام در نشر آثار برگزیده علوم کاربردی کامپیوتر و الکترونیک

سرشناسه: فتحی، محمود، ۱۳۳۸ - گردآورنده، مترجم
عنوان و نام پدیدآور: مقدمه‌ای بر یادگیری ماشین با کاربردهایی مبتنی بر Python و R = An introduction to machine learning with applications...
مشخصات نشر: تهران: کانون نشر علوم: کاتوزی، ۱۳۹۸.
مشخصات ظاهری: ۳۳۰ ص: مصور، جدول، نمودار.
شابک: ۹۷۸۹۶۴۳۲۷۲۵۰۰

وضعیت ویراست: ویراست ۲.
یادداشت: کتاب حاضر با عنوان "مقدمه‌ای بر یادگیری ماشین با کاربردهایی مبتنی بر R" توسط کانون نشر علوم در سال ۱۳۹۶ منتشر شده است.
عنوان دیگر: مقدمه‌ای بر یادگیری ماشین با کاربردهایی مبتنی بر R

موضوع: فراگیری ماشینی
Machine learning
موضوع: فراگیری ماشینی - ابزار
موضوع: Computer software
Machine learning

موضوع: آر (زبان برنامه‌نویسی کامپیوتر)
موضوع: پایتون (زبان برنامه‌نویسی کامپیوتر)
موضوع: Python (Computer program language)
شناخته افزوده: محمدی، علی محمد، ۱۳۶۰ - گردآورنده، مترجم

رده‌بندی کنگره: ۵ / QC۲۲۵

رده‌بندی دیویی: ۰۰۶/۳۱

شماره کتابشناسی ملی: ۵۹۶۶۶۵۴

نام کتاب: مقدمه‌ای بر یادگیری ماشین با کاربردهایی مبتنی بر Python و R

ناشر: کانون نشر علوم

ناشر همکار: کاتوزی

تألیف و ترجمه: دکتر محمود فتحی و علی محمد محمدی

صفحه‌آرا: فاطمه آدیگوزل‌پور اردبیلی

طرح جلد: زهرا سلطان‌آبادی

چاپ اول: ۱۳۹۸

تیراژ: ۵۰۰

چاپ و صحافی: عطا، کیمیا

قیمت: ۵۰۰۰۰ تومان

شابک: ۹۷۸۹۶۴۳۲۷۲۵۰۰

دفتر مرکزی پخش: خیابان انقلاب، خیابان فخررازی، خیابان وحید نظری شرقی، پلاک ۶۵، واحد ۱

تلفن: ۶۶۹۶۱۵۶۸-۶۶۳۹۴۹۱۱



پیشگفتار

کتاب حاضر مقدمه‌ای بر یادگیری ماشین با استفاده از زبان R و Python می‌باشد. یادگیری ماشین یکی از حوزه‌های علمی مربوط به هوش مصنوعی است و امروزه به صورت گسترده در نرم‌افزارها و وب‌سایت‌ها، بازی‌های کامپیوتری، در سیستم‌های کامپیوتری همه منظوره، در تلفن‌های هوشمند، دستگاه‌های بازی‌های کامپیوتری و غیره، به طریقی مورد استفاده قرار می‌گیرد. یادگیری ماشین بخش اساسی در هوشمندسازی زندگی بشر داشته و در دنیای امروز و آینده وجود نوعی از الگوریتم‌های یادگیری ماشین به طریقی به چشم خورده یا خواهد خورد.

الگوریتم و مدل‌های یادگیری ماشین برخلاف الگوریتم‌های سنتی برنامه‌ریزی کامپیوترها به صورت خودکار الگویی وجود دارد که مجموعه داده‌های متنوع را تشخیص و کشف کرده و با یادگیری از این داده‌ها (معمولاً به داده آموزشی) مدلی می‌سازد که بتواند داده‌هایی که در مجموعه داده‌های آموزشی وجود نداشته (معمولاً به داده آزمون) را تحلیل کرده و خروجی با کارایی بالا برای این داده‌های دیده نشده را تولید کند. مثلاً با یادگیری تغییرات نرخ ارز از مجموعه داده‌های موجود در دوره‌های زمانی گذشته، می‌تواند پیش‌بینی نرخ ارز را در آینده انجام دهد. در حوزه یادگیری ماشین مدل‌های متنوع و گسترده وجود دارد که طراحی و پیاده‌سازی یک مدل مناسب بستگی به نوع کاربرد؛ نوع داده‌ها و پایداری متعددی دارد که در این کتاب به طراحی برخی از این مدل‌ها و نیز کاربردهای مهم این مدل‌ها پرداخته می‌شود. در این کتاب ضمن ورود به مبانی نظری و کاربردی مدل‌های مهم یادگیری ماشین در حد نیاز؛ موضوعات به نحوی ارائه می‌گردند که خواننده علاوه بر درک این مدل‌های یادگیری ماشین؛ تخصص‌های لازم جهت پیاده‌سازی عملی مدل‌های یادگیری ماشین را نیز فرا گیرد. به نحوی که با مطالعه کامل این کتاب خواننده قادر خواهد بود به عنوان یک متخصص یادگیری ماشین وارد بازار کار شده و پروژه‌های جدید در حوزه‌های متنوع یادگیری ماشین؛ که بسیار مورد نیاز جامعه امروزی می‌باشد را طراحی و پیاده‌سازی کند.

این کتاب به زبان ساده مبانی نظری و عملی بسیاری از الگوریتم‌های مهم یادگیری ماشین را پوشش می‌دهد. این کتاب می‌تواند جهت دانشجویان دوره کارشناسی سال آخر گرایش‌های مختلف مهندسی کامپیوتر، فناوری اطلاعات، علوم کامپیوتر، ریاضی کاربردی، و صنایع و آمار مورد استفاده قرار گیرد. همچنین دانشجویان کارشناسی ارشد گرایش هوش مصنوعی و سایر گرایش‌های ذکر



شده می‌توانند به عنوان کتاب درسی؛ در درس یادگیری ماشین از این کتاب بهره ببرند. برای خواندن این کتاب، درس پیش نیاز خاصی مورد نیاز نیست و داشتن دانش ریاضی عمومی دوره کارشناسی رشته‌های مهندسی کفایت می‌کند. بنابراین کلیه علاقه‌مندان، اعم از دانشجویان گرایش‌های مختلف مهندسی، مهندسين و محققینی که به دلیل علاقه شخصی خواهان ورود به این حوزه هستند می‌توانند از این کتاب بهره ببرند و تخصص یادگیری ماشین را کسب کرده و وارد بازار کار این حوزه؛ که بسیار فراگیر هم هست؛ شوند. در نگارش این کتاب سعی شده که در انتهای هر فصل مثال‌ها و عملیاتی با استفاده از زبان R و Python گنجانده شود. کتاب شامل ۱۷ فصل و دو پیوست است که فصل‌های آن اصول طراحی و پیاده‌سازی الگوریتم‌های مهم یادگیری ماشین را به زبانی ساده می‌داند. پیوست ۱ اصول روابط و عملیات ماتریس‌ها که به صورت گسترده در این حوزه استفاده می‌شود را مختصراً توضیح می‌دهد. پیوست ۲ آشنایی مقدماتی ولی کافی در ارتباط با زبان R و بسته‌های نرم‌افزاری مربوط به یادگیری ماشین است و نحوه استفاده از زبان R و این بسته‌ها را آموزش می‌دهد. زبان‌های برنامه‌نویسی R و پایتون در حال حاضر به عنوان محبوب‌ترین زبان‌های برنامه‌نویسی در میان متخصصین یادگیری ماشین شناخته شده است و به صورت گسترده مورد استفاده قرار می‌گیرد. زبان پایتون یک زبان برنامه‌نویسی همه منظوره، سطح بالا و متن باز است که فلسفه این زبان بر تأکید بر دو هدف اصلی است که یکی سادگی و قابل فهم بودن ساختار برنامه‌نویسی، و دیگری توانایی این زبان برای ارائه حداکثر کارایی با حداقل کدهای نوشته شده می‌باشد. این زبان در حوزه داده کاوی یکی از پشتمندترین زبان‌ها است بلکه در حوزه پردازش تصویر و بینایی کامپیوتر یکی از قدرتمندترین زبان‌ها می‌باشد که در سال‌های اخیر مورد استفاده اکثر متخصصین این حوزه قرار گرفته است. همچنین زبان R نیز که یک زبان منبع باز است به صورت گسترده در آمار و ریاضیات مورد استفاده قرار می‌گیرد. در اصل این زبان حاوی محدوده گسترده‌ای از تکنیک‌های آماری از جمله مدل‌سازی خطی و غیر خطی، دسته‌بندی، آزمون‌های کلاسیک آماری و غیره می‌باشد. به صورت کلی زبان R و پایتون دو زبان بسیار شبیه به هم اما دارای خصوصیات کاملاً منحصر به فرد می‌باشند که باعث شده هر دو زبان در حوزه‌های مختلف مورد علاقه افراد بسیاری قرار گیرند. تا کنون چندین هزار جعبه ابزار و بسته‌های نرم‌افزاری آماده به کار و قابل استفاده با زبان R و Python توسعه یافته است که به سادگی کاربر می‌تواند با استفاده از این جعبه ابزار و نرم‌افزارها پروژه‌های نوین و جدید در حوزه‌های متنوع یادگیری ماشین را پیاده‌سازی کند. نحوه استفاده از



فهرست مطالب

پیشگفتار..... ۵

مقدمه..... ۱۹

فصل ۱: رگرسیون..... ۲۵

۱-۱- رگرسیون خطی..... ۲۶

۱-۲- رگرسیون چند متغیره..... ۲۶

۱-۳- تقسیم داده‌ها به داده‌های آموزشی و آزمون..... ۲۷

۱-۴- تابع هزینه..... ۲۹

۱-۴-۱- تشریح روش محاسبه تابع هزینه..... ۳۱

۱-۴-۲- نگاهی عمیق‌تر به تابع هزینه..... ۳۵

۱-۵- گرادیان نزولی..... ۳۷

۱-۵-۱- رابطه گرادیان نزولی..... ۳۸

۱-۵-۲- الگوریتم گرادیان نزولی..... ۳۸

۱-۵-۳- نرخ یادگیری در گرادیان نزولی..... ۳۹

۱-۵-۳-۱- اندازه گام‌های نرخ یادگیری در گرادیان نزولی..... ۴۰

۱-۶- گرادیان نزولی و تابع هزینه در حل رگرسیون خطی..... ۴۲

۱-۷- مثال عملی برای یک رگرسیون خطی تک متغیره با زبان R..... ۴۳

۱-۸- مثال عملی برای یک رگرسیون خطی تک متغیره با زبان پایتون..... ۴۸

فصل ۲: رگرسیون خطی چند متغیره..... ۵۱

۲-۱- رگرسیون خطی چند متغیره..... ۵۲

۲-۲- گرادیان نزولی برای رگرسیون خطی چند متغیره..... ۵۴

۲-۳- به‌روزرسانی پارامترها در گرادیان نزولی..... ۵۵

۲-۴- الگوریتم به‌روزرسانی پارامترهای چند متغیره..... ۵۵

۲-۵- مقیاس‌گذاری ویژگی‌ها..... ۵۶

۲-۶- نرمال کردن متوسط..... ۵۸

۲-۷- نرخ یادگیری در رگرسیون خطی چند متغیره..... ۵۸



- ۵۹-۸-۲- درجه چندجمله‌ای معادله فرضیه رگرسیون خطی.....
۶۱-۹-۲- مثال عملی برای یک رگرسیون خطی چند متغیره با R.....
۶۳-۱۰-۲- مثال عملی برای یک رگرسیون خطی چند متغیره با زبان پایتون.....

فصل ۳: رگرسیون لاجستیک

- ۶۶-۱-۳- رگرسیون لاجستیک.....
۶۷-۲-۳- دسته‌بندی با استفاده از رگرسیون لاجستیک.....
۶۷-۱-۲-۳- معرفی فرضیه در رگرسیون لاجستیک.....
۶۷-۲-۲-۳- تابع سیگموئید.....
۶۸-۳-۳- مرز تصمیم در رگرسیون لاجستیک.....
۶۹-۱-۲-۳- روش محاسبه حل مسئله در فضای نمونه‌ها.....
۷۰-۴-۳- مرز تصمیم غیر خطی.....
۷۱-۵-۳- تابع هزینه.....
۷۲-۱-۵-۳- تابع هزینه در رگرسیون لاجستیک.....
۷۳-۲-۵-۳- نقطه کمینه تابع هزینه.....
۷۳-۳-۵-۳- نقطه بیشینه تابع هزینه.....
۷۴-۶-۳- گرادیان نزولی برای رگرسیون لاجستیک.....
۷۴-۷-۳- رابطه تابع هزینه در رگرسیون لاجستیک.....
۷۵-۸-۳- الگوریتم گرادیان نزولی برای رگرسیون لاجستیک.....
۷۶-۹-۳- رگرسیون لاجستیک چند کلاسه.....
۷۸-۱۰-۳- پیاده‌سازی یک رگرسیون لاجستیک با استفاده از زبان R.....
۸۳-۱۱-۳- پیاده‌سازی یک رگرسیون لاجستیک با استفاده از زبان پایتون.....

فصل ۴: راهکارهایی جهت بهبود طراحی الگوریتم‌های یادگیری ماشین

- ۸۶-۱-۴- راهکارهایی جهت بهبود طراحی الگوریتم‌های یادگیری ماشین.....
۸۶-۲-۴- تعمیم‌پذیری الگوریتم.....
۸۶-۳-۴- عوامل رسیدن الگوریتم به حالت بهینه.....
۸۶-۴-۴- ظرفیت یک فرضیه.....



- ۸۷-۵-۴- ایجاد مدل مناسب برای الگوریتم یادگیری.....
- ۸۸-۶-۴- مفهوم بیش برآزش و کم برآزش، بایاس و واریانس.....
- ۸۸-۶-۴-۱- بیش برآزش و واریانس بالا.....
- ۸۹-۶-۴-۲- کم برآزش و بایاس بالا.....
- ۸۹-۶-۴-۳- حالت مناسب از نظر بایاس و واریانس.....
- ۹۰-۶-۴-۴- نقش انتخاب ویژگی‌ها.....
- ۹۱-۸-۴- اعتبار سنجی.....
- ۹۱-۹-۴- اعتبار سنجی تقاطعی K گروهی.....
- ۹۲-۱۰-۴- تنظیم.....
- ۹۲-۱۱-۴- تنظیم کننده برای رگرسیون خطی.....
- ۹۳-۱۲-۴- رابطه تنظیم کننده برای تابع مزین.....
- ۹۴-۱۳-۴- انتخاب مدل در گرادیان نزولی.....
- ۹۵-۱۴-۴- تنظیم کننده برای رگرسیون لاجستیک.....
- ۹۵-۱۵-۴- الگوریتم گرادیان نزولی با استفاده از تنظیم کنند.....

فصل ۵: شبکه عصبی

- ۹۸-۱-۵- شبکه عصبی.....
- ۹۸-۱-۱-۵- ساختار ترون مغزی.....
- ۹۸-۲-۵- شبکه عصبی مصنوعی.....
- ۹۹-۱-۲-۵- تاریخچه شبکه عصبی مصنوعی.....
- ۹۹-۳-۵- پرسپترون ساده.....
- ۱۰۰-۱-۳-۵- محاسبه پرسپترون ساده.....
- ۱۰۱-۲-۳-۵- شبیه سازی گیت OR با پرسپترون ساده.....
- ۱۰۳-۴-۵- پرسپترون چند لایه.....
- ۱۰۴-۱-۴-۵- الگوریتم آموزش پرسپترون چند لایه.....
- ۱۰۴-۲-۴-۵- پس انتشار خطا و گرادیان نزولی.....
- ۱۰۴-۳-۴-۵- مراحل الگوریتم پرسپترون با روش پس انتشار خطا.....
- ۱۰۵-۴-۴-۵- فرایند اجرای الگوریتم پرسپترون چند لایه با روش پس انتشار خطا.....



۱۶۹

فصل ۹: تشخیص ناهنجاری

- ۱-۹- تشخیص ناهنجاری ۱۷۰
- ۲-۹- کاربردهای مختلف تشخیص ناهنجاری ۱۷۱
- ۱-۲-۹- تشخیص تقلب ۱۷۱
- ۳-۹- مباحث ریاضی برای تعیین ناهنجاری ۱۷۲
- ۳-۶- تابع گوسی ۱۷۲
- ۲-۳-۹- تخمین ارامت های تابع گوسی ۱۷۴
- ۳-۳-۹- گوریت تشخیص ناهنجاری ۱۷۵
- ۴-۳-۹- مراحل الگوریتم تشخیص ناهنجاری ۱۷۶
- ۴-۹- ارزیابی عملکرد الگوریتم تشخیص ناهنجاری ۱۷۶

۱۸۱

فصل ۱۰: یادگیری تقویتی

- ۱-۱۰- یادگیری تقویتی ۱۸۲
- ۲-۱۰- حوزه عملکرد ۱۸۲
- ۳-۱۰- الگوریتم های یادگیری تقویتی ۱۸۳
- ۴-۱۰- روش یادگیری Q ۱۸۳
- ۵-۱۰- آموزش عامل در روش Q ۱۸۵
- ۶-۱۰- مراحل الگوریتم یادگیری Q ۱۸۵
- ۱-۶-۱۰- نرخ یادگیری در روش یادگیری Q ۱۸۶
- ۲-۶-۱۰- عامل تخفیف در روش یادگیری Q ۱۸۶
- ۳-۶-۱۰- ماتریس R یا پاداش در یادگیری Q ۱۸۶
- ۴-۶-۱۰- یک دور آموزشی از عمل عامل در آموزش Q ۱۸۶
- ۷-۱۰- یک مثال مفهومی برای الگوریتم Q ۱۸۶
- ۸-۱۰- برنامه یادگیری Q با زبان R ۱۸۹

۱۹۳

فصل ۱۱: سیستم های توصیه گر

- ۱-۱۱- سیستم های توصیه گر ۱۹۴
- ۲-۱۱- تعریف آیتم در سیستم های توصیه گر ۱۹۴



- ۱۱-۳- تعریف کاربر در سیستم‌های توصیه‌گر..... ۱۹۴
- ۱۱-۴- عملکردها و اهداف سیستم‌های توصیه‌گر..... ۱۹۵
- ۱۱-۵- داده و منبع دانش..... ۱۹۶
- ۱۱-۵-۱- تراکنش..... ۱۹۶
- ۱۱-۶-۱- امتیاز ارزیابی..... ۱۹۶
- ۱۱-۷-۱- روش‌های توصیه‌گری..... ۱۹۷
- ۱۱-۷- انواع رویکردهای اصلی پیاده‌سازی سیستم توصیه‌گر..... ۱۹۷
- ۱۱-۷-۱- مبانی محتوا..... ۱۹۷
- ۱۱-۷-۲- فیلتر مشترک..... ۱۹۷
- ۱۱-۷-۳- جمعیتی..... ۱۹۷
- ۱۱-۷-۴- بر مبنای دانش..... ۱۹۸
- ۱۱-۷-۵- مبتنی بر جامعه..... ۱۹۸
- ۱۱-۷-۶- توصیه‌گر ترکیبی..... ۱۹۸
- ۱۱-۸- کاربردهای سیستم توصیه‌گر..... ۱۹۸
- ۱۱-۹- روابط ریاضی و پیاده‌سازی سیستم توصیه‌گر..... ۱۹۸
- ۱۱-۱۰- محاسبه شباهت با رابطه پیرسون در فیلتر مشترک..... ۱۹۹
- ۱۱-۱۱- پیاده‌سازی فیلتر مشترک با کدهای R..... ۲۰۰

فصل ۱۲: داده‌های جریانی..... ۲۰۱

- ۱۲-۱- داده‌های جریانی..... ۲۰۲
- ۱۲-۲- محدودیت‌های داده‌های جریانی..... ۲۰۲
- ۱۲-۳- تعاریف داده‌های جریانی..... ۲۰۳
- ۱۲-۳-۱- تعریف جریان داده..... ۲۰۳
- ۱۲-۳-۲- تعریف مفهوم..... ۲۰۳
- ۱۲-۳-۳- تعریف رانش مفهوم..... ۲۰۳
- ۱۲-۳-۳-۱- انواع رانش مفهوم..... ۲۰۳
- ۱۲-۳-۴- تعریف پنجره زمانی..... ۲۰۳
- ۱۲-۳-۴-۱- انواع پنجره زمانی..... ۲۰۳



- ۱۵-۴-۲- یادگیری ویژگی با خودمرزکننده ۲۳۷
- ۱۵-۵- ماشین بولتزمن ۲۳۷
- ۱۵-۵-۱- ماشین بولتزمن محدود ۲۳۸
- ۱۵-۵-۲- بازسازی در الگوریتم ماشین بولتزمن ۲۴۰
- ۱۵-۶- شبکه عصبی بازگشتی ۲۴۱
- ۱۵-۷- شبکه باور عمیق ۲۴۲
- ۱۵-۷-۱- شبکه باور عمیق ۲۴۲
- ۱۵-۸- شبکه عصبی کانولوشن ۲۴۳
- ۱۵-۸-۱- ساختار کلی شبکه عصبی کانولوشن ۲۴۵
- ۱۵-۸-۲- ورودی شبکه عصبی کانولوشن ۲۴۵
- ۱۵-۸-۳- لایه کانولوشن ۲۴۵
- ۱۵-۸-۴- تصحیح واحد خطی $(k^2 + 1)$ ۲۴۶
- ۱۵-۸-۵- لایه ادغام ۲۴۶
- ۱۵-۸-۶- لایه تماماً متصل ۲۴۷

فصل ۱۶: روش‌هایی جهت افزایش کارایی الگوریتم‌های یادگیری ماشین ۲۴۹

- ۱۶-۱- روش‌هایی جهت افزایش کارایی الگوریتم‌های یادگیری ماشین ۲۵۰
- ۱۶-۲- بهبود کارایی مدل با روش مجمع ۲۵۱
- ۱۶-۲-۱- بهبود کارایی روی داده‌ای حجیم یا حتی مجموعه داده‌های کوچک ۲۵۳
- ۱۶-۲-۲- توانایی سنتز داده از حوزه‌های مجزا ۲۵۳
- ۱۶-۲-۳- درک بهتر از عملیات یادگیری پیچیده ۲۵۳
- ۱۶-۲-۴- روش Bagging ۲۵۳
- ۱۶-۲-۵- روش Boosting ۲۵۴
- ۱۶-۲-۶- روش Random forest (RF) ۲۵۵
- ۱۶-۳- افزایش کارایی در R از نظر زمان اجرا ۲۵۵
- ۱۶-۴- مدیریت مجموعه داده‌های خیلی بزرگ ۲۵۶
- ۱۶-۵- یادگیری سریع‌تر با پردازش موازی ۲۵۶
- ۱۶-۶- روش اندازه‌گیری زمان اجرا ۲۵۷



۲۵۷ ۷-۱۶ پردازش موازی در محاسبات ابری Hadoop و Mapreduce

۲۵۸ ۸-۱۶ واحد پردازش گرافیکی (GPU)

۲۵۸ ۹-۱۶ پیاده‌سازی مؤثرتر الگوریتم یادگیری ماشین با R

۲۵۹ پیوست ۱: مقدمه‌ای بر جبر خطی (مبانی عملیات ماتریس‌ها)

۲۶۷ پیوست ۱: آموزش زبان برنامه‌نویسی R

۲۹۹ پیوست ۳: آموزش مقدماتی پایتون

۳۱۳ نمایه (انگلیسی و فارسی)

۳۱۴ «انگلیسی به فارسی»

۳۲۱ «فارسی به انگلیسی»

۳۲۸ منابع و مراجع