

به نام خدا

درس نامه
یادگیری ماشین

نویسندگان: آندری بورکوف

مترجمان:

دکتر مهدی اسماعیلی

مهندس فائزه اکبری

درس نامه یادگیری ماشین

مترجمان: دکتر مهدی اسماعیلی، مهندس فائزه اکبری

ناشر: انتشارات آتی نگر

ناشر همکار: وینا

تیراژ: ۵۰۰ نسخه

چاپ اول: ۱۳۹۸

قیمت: ۳۲۵,۰۰۰ ریال

شابک: ۹۷۸-۶۲۲-۶۱۰۲-۵۷-۵

ISBN: 978-622-6102-57-5

حق چاپ برای انتشارات آتی نگر محفوظ است.

نشانی دفتر فروش: خیابان جمالزاده جنوبی، روبه روی کوچه رشتچی، پلاک ۱۴۴، واحد ۱

نمابر: ۶۶۵۶۵۳۳۷

تلفن: ۸-۶۶۵۶۵۳۳۶



www.ati-negar.com * info@ati-negar.com

سرشناسه: بورکوف، آندری، Burkov, Andriy

درس نامه یادگیری ماشین / نویسنده اندری بورکوف مترجمان: مهدی اسماعیلی، فائزه اکبری.

تهران: آتی نگر، وینا ۱۳۹۸

۱۹۲ ص.: مصور، جدول، نمودار.

ISBN: 978-622-6102-57-5

وضعیت فهرست نویسی: فیاپا.

یادداشت: عنوان اصلی کتاب: The hundred-page machine learning book, c2019.

موضوع: فراگیری ماشینی - Machine learning

موضوع: مهندسی مواد -- داده پردازی - Materials science -- Data processing

شناسه افزوده: اسماعیلی، مهدی، ۱۳۵۰- مترجم

شناسه افزوده: اکبری، فائزه، ۱۳۶۱- مترجم

رده بندی کنگره:

رده بندی دیویی:

شماره کتابشناسی ملی:

QC۳۲۵/۵

۰۰۶/۳۱

۵۷۷۲۱۱۳

فهرست مطالب

پیشگفتار.....	۹
فصل اول: مقدمه.....	۱۱
۱-۱ یادگیری ماشین چیست.....	۱۱
۲-۱ انواع یادگیری.....	۱۱
۲-۱-۱ یادگیری باناظر.....	۱۲
۲-۲-۱ یادگیری بیناظر.....	۱۳
۳-۲-۱ یادگیری نیمه ناظر.....	۱۳
۴-۲-۱ یادگیری تقویتی.....	۱۳
۳-۱ چگونه یادگیری باناظر کار می کند.....	۱۴
۴-۱ چرا مدل روی داده های جدید کار می کند.....	۱۹
فصل دوم: نشانه گذاری و تعریف ها.....	۲۱
۱-۲ نشانه گذاری.....	۲۱
۱-۱-۲ ساختمان داده ها.....	۲۱
۲-۱-۲ مجموع و علامت Σ	۲۳
۳-۱-۲ ضرب و علامت Π	۲۳
۴-۱-۲ عملیات مجموعه ها.....	۲۴
۵-۱-۲ عملیات روی بردارها.....	۲۴
۶-۱-۲ توابع.....	۲۵
۷-۱-۲ عملگرهای Max و Arg Max.....	۲۷
۸-۱-۲ عملگر انتساب.....	۲۷
۹-۱-۲ مشتق و گرادیان.....	۲۷
۲-۲ متغیرهای تصادفی.....	۲۹
۳-۲ برآوردگرهای نأریب.....	۳۱
۴-۲ قاعده بیز.....	۳۲

- ۳۲ ۵-۲ برآورد پارامتر
 ۳۳ ۶-۲ پارامترها در مقابل آبرپارامترها
 ۳۴ ۷-۲ رده‌بندی در مقابل رگرسیون
 ۳۵ ۸-۲ یادگیری مبتنی بر مدل در مقابل یادگیری مبتنی بر نمونه
 ۳۵ ۹-۲ یادگیری کم‌عمق در مقابل یادگیری ژرف
 ۳۷ فصل سوم: الگوریتم پایه

- ۳۷ ۱-۳ رگرسیون خطی
 ۳۷ ۱-۱-۳ بیان مسئله
 ۳۹ ۲-۱-۳ راه حل
 ۴۲ ۲-۳ رگرسیون لجستیک
 ۴۲ ۱-۲-۳ بیان مسئله
 ۴۴ ۲-۲-۳ راه حل
 ۴۵ ۳-۳ یادگیری درخت تصمیم
 ۴۶ ۱-۳-۳ بیان مسئله
 ۴۶ ۲-۳-۳ راه حل
 ۴۹ ۴-۳ ماشین بردار پشتیبان
 ۵۰ ۱-۴-۳ مواجه شدن با نویز
 ۵۱ ۲-۴-۳ تفکیک‌پذیری غیرخطی
 ۵۴ ۵-۳ الگوریتم k -نزدیکترین همسایه

- ۵۷ فصل چهارم: تشریح یک الگوریتم یادگیری
 ۵۷ ۱-۴ عناصر سازنده یک الگوریتم یادگیری
 ۵۸ ۲-۴ کاهش گرادیان
 ۶۴ ۳-۴ کارشناسان یادگیری ماشین چگونه کار می‌کنند
 ۶۵ ۴-۴ خصوصیات الگوریتم‌های یادگیری

- ۶۷ فصل پنجم: تجارب برتر
 ۶۷ ۱-۵ مهندسی ویژگی

۶۸.....	۱-۱-۵ کُدگذاری one-hot
۶۹.....	۱-۲-۵ بسته‌بندی
۷۰.....	۳-۱-۵ نرمال‌سازی
۷۰.....	۴-۱-۵ استانداردسازی
۷۱.....	۵-۱-۵ ویژگی‌های ناموجود
۷۲.....	۶-۱-۵ تکنیک‌های جان‌هی داده‌ها
۷۳.....	۲-۵ انتخاب الگوریتم یادگیری
۷۵.....	۳-۵ سه مجموعه داده
۷۷.....	۴-۵ کم‌برازش و بیش‌برازش
۷۹.....	۵-۵ تنظیم
۸۱.....	۶-۵ ارزیابی کارایی مدل
۸۲.....	۱-۶-۵ ماتریس درهم‌ریختگی
۸۴.....	۲-۶-۵ دقت و بازخوانی
۸۵.....	۳-۶-۵ درستی (صحت)
۸۵.....	۴-۶-۵ درستی حساس به هزینه
۸۶.....	۵-۶-۵ سطح زیر منحنی ROC (AUC)
۸۸.....	۷-۵ تنظیم اَبَر پارامتر
۹۰.....	۱-۷-۵ اعتبارسنجی متقابل

فصل ششم: شبکه‌های عصبی و یادگیری ژرف ۹۱

۹۱.....	۱-۶ شبکه‌های عصبی
۹۲.....	۱-۱-۶ مثالی از پرسپترون چندلایه
۹۴.....	۲-۱-۶ معماری شبکه عصبی پیشخور
۹۶.....	۲-۶ یادگیری ژرف
۹۷.....	۱-۲-۶ شبکه عصبی کانولوشن
۱۰۵.....	۲-۲-۶ شبکه‌های عصبی برگشتی

فصل هفتم: راه‌حل‌هایی برای چند مسئله ۱۱۱

۱۱۱.....	۱-۷ رگرسیون کرنل
----------	------------------

۱۱۲.....	۲-۷ رده‌بندی چندگانه
۱۱۴.....	۳-۷ رده‌بندی تک‌رده‌ای
۱۱۶.....	۴-۷ رده‌بندی چندبرچسبی
۱۱۹.....	۵-۷ یادگیری تلفیقی
۱۲۰.....	۱-۵-۷ روش‌های تقویتی و استقلالی
۱۲۰.....	۲-۵-۷ جنگل تصادفی
۱۲۲.....	۳-۵-۷ تقویت گرادیان
۱۲۵.....	۶-۷ یادگیری برای برچسب زنی توالی‌ها
۱۲۶.....	۷-۷ یادگیری توالی به توالی
۱۲۹.....	۸-۷ یادگیری فعال
۱۳۰.....	۹-۷ یادگیری نیمه‌ناظر
۱۳۴.....	۱۰-۷ یادگیری تکشات
۱۳۶.....	۱۱-۷ یادگیری بیشات

فصل هشتم: تجارب پیشرفته..... ۱۳۹

۱۳۹.....	۱-۸ کنترل مجموعه داده‌های نامتوازن
۱۴۱.....	۲-۸ مدل‌های ترکیبی
۱۴۲.....	۳-۸ آموزش شبکه‌های عصبی
۱۴۴.....	۴-۸ تنظیمات پیشرفته
۱۴۵.....	۵-۸ کنترل چندین ورودی
۱۴۶.....	۶-۸ کنترل چندین خروجی
۱۴۷.....	۷-۸ یادگیری انتقالی
۱۴۸.....	۸-۸ کارایی الگوریتم

فصل نهم: یادگیری بی‌ناظر..... ۱۵۳

۱۵۳.....	۱-۹ تخمین چگالی
۱۵۶.....	۲-۹ خوشه‌بندی
۱۵۷.....	۱-۲-۹ الگوریتم k-means
۱۵۷.....	۲-۲-۹ الگوریتم‌های DBSCAN و HDBSCAN

۱۵۹	۳-۲-۹ تعیین تعداد خوشه‌ها
۱۶۳	۴-۲-۹ الگوریتم‌های خوشه‌بندی دیگر
۱۶۷	۳-۹ کاهش ابعاد
۱۶۷	۱-۳-۹ تحلیل مؤلفه‌های اصلی
۱۶۹	۲-۳-۹ تکنیک UMAP
۱۷۱	۴-۹ تشخیص داده‌های پرت

فصل دهم: شکل‌های دیگری از یادگیری ۱۷۳

۱۷۳	۱-۱۰ یادگیری متریک
۱۷۵	۲-۱۰ یادگیری برای رتبه‌بندی
۱۷۹	۳-۱۰ یادگیری برای پیشنهاد
۱۸۱	۱-۳-۱۰ ماشین‌های تجزیه
۱۸۳	۲-۳-۱۰ خودرمزگذارهای مخرب
۱۸۴	۴-۱۰ یادگیری خودناظر: جایگزین‌های واژگان

فصل یازدهم: سخن پایانی ۱۸۷

۱۸۸	۱-۱۱ آنچه که در این کتاب دیده نمی‌شود
۱۸۸	۱-۱-۱۱ مدلسازی موضوعی
۱۸۸	۲-۱-۱۱ فرایندهای گاوسی
۱۸۸	۳-۱-۱۱ مدل‌های خطی تعمیم‌یافته
۱۸۹	۴-۱-۱۱ مدل‌های گرافی احتمالاتی
۱۸۹	۵-۱-۱۱ مؤنت کارلوی زنجیره مارکوفی
۱۹۰	۶-۱-۱۱ شبکه‌های تخصصی مولد
۱۹۰	۷-۱-۱۱ الگوریتم‌های ژنتیک
۱۹۱	۸-۱-۱۱ یادگیری تقویتی

آنچنان مهر توام در دل و جان جای گرفت
که اگر سر برود از دل و از جان نرود

پیشگفتار

یادگیری ماشین در واقع مطالعه علمی الگوریتم‌ها و مدل‌های آماری است که در آن سیستم‌های کامپیوتری بدون نیاز به دستورالعمل‌های صریح برنامه‌نویسی، به منظور یک وظیفه خاص استفاده می‌شوند. الگوریتم‌های یادگیری ماشین با استفاده از مجموعه داده‌ها و برای پیش‌بینی و تصمیم‌گیری، به خلق مدل‌های ریاضی می‌پردازند. منابع علمی و کتاب‌های زیادی در حوزه یادگیری ماشین وجود دارند، اما باور کنید با مطالعه کتاب حاضر متوجه خواهید شد که هر چه را نیاز دارید در این کتاب صد و چند صفحه‌ای می‌توانید پیدا کنید.

تکنیک‌های یادگیری ماشین موضوع اصلی این کتاب را تشکیل می‌دهد و مطالب آن در یازده فصل گردآوری و تهیه شده است. در فصل اول با مفهوم یادگیری ماشین و گونه‌های مختلف آن آشنا می‌شوید. در فصل دوم کتاب به ریاضیات مورد استفاده در روش‌های یادگیری ماشین اشاره شده است و به گونه‌ای پیش‌نیاز فصل‌های بعدی محسوب می‌شود. برخی از الگوریتم‌های پایه در حوزه یادگیری ماشین مانند رگرسیون، درخت تصمیم، ماشین بردار پشتیبان و نزدیکترین همسایه در فصل سوم توضیح داده شده است. تشریح یک الگوریتم یادگیری در فصل چهارم آمده است. فصل پنجم کتاب حاوی مطالب مهمی نظیر روش‌های آماده‌سازی داده‌ها، چگونگی انتخاب الگوریتم یادگیری و همچنین ارزیابی آن است. شبکه‌های عصبی و یادگیری ژرف، موضوع فصل ششم است. به جرأت می‌توان گفت که فصل هفتم کتاب، خود یک کتاب کامل است. مطالبی از قبیل کرنل‌ها، رده‌بندی چندگانه، رده‌بندی تک‌رده‌ای، یادگیری تلفیقی، یادگیری توالی‌ها و یادگیری فعال برخی از مطالب این فصل را تشکیل می‌دهند. برخی از چالش‌های موجود در اجرای الگوریتم‌های یادگیری ماشین مانند مجموعه داده‌های نامتوازن، آموزش شبکه‌های عصبی، کنترل ورودی‌ها و خروجی‌های چندگانه در فصل هشتم بحث شده است. فصل نهم کتاب به موضوع یادگیری بی‌ناظر پرداخته است که در آن چند تکنیک خوشه‌بندی و روش‌های کاهش ابعاد توضیح داده شده‌اند. شکل‌های دیگری از یادگیری مانند یادگیری متریک، یادگیری برای رتبه‌بندی و یادگیری برای پیشنهاد در فصل دهم مطرح شده‌اند. در فصل نهم نیز به موضوعاتی اشاره شده است که در این کتاب از آن‌ها صحبتی نشده است. برخی از این موضوعات عبارتند از: یادگیری تقویتی، مدل‌سازی موضوعی، فرایندهای گاوسی، مدل‌های خطی تعمیم‌یافته و الگوریتم‌های ژنتیک.

تمام تلاش خود را انجام داده‌ایم تا مطالب این کتاب به گونه‌ای ترجمه شوند که خوانندگان محترم بتوانند مفاهیم آن را به راحتی درک کنند. بدون شک برای برخی از واژه‌های معادلی که پیشنهاد یا انتخاب شده‌اند، برابری دیگر (و احياناً بهتر) بتوان یافت. آنچه مسلم است این است که اگر چه گاهی تبدیل واژه‌ها آنچنانکه باید و شاید انجام نشده است؛ اما در ادای جملات و بیان موضوع سعی فراوان شده است تا خوانندگان گرامی با متنی مبهم و گیج کننده روبه‌رو نشوند.

در اینجا لازم می‌دانم از همه اساتید و دانشجویان به خاطر راهنمایی‌های ارزشمندشان در حین آماده‌سازی این کتاب سپاسگزاری کنم. بدون شک، همت همکار محترم سرکار خانم فائزه اکبری باعث شد تا این کتاب مهیا شود و بدون لطف ایشان این کار میسر نمی‌شد. همچنین از مدیریت محترم انتشارات آتی‌نگر و دوست عزیزم جناب آقای رامین مولاناپور نیز به خاطر آماده‌سازی، چاپ و پخش این کتاب تشکر می‌کنم. رهین محبت بی‌دریغ خانواده‌ام هستم که با فراهم‌سازی محیطی مناسب مرا یاری کردند. اما با وجود همه سعی و تلاشی که در تمام مراحل آماده‌سازی این کتاب انجام گرفته است، یقین دارم که عاری از اشتباه نیست، چرا که تنها مکتوب بی‌نقص همان معجزه جاوید قرآن کریم است. در آخر ضمن سپاسگزاری از همه کسانی که مرا یاری داده‌اند و با پذیرش مسئولیت هرگونه کاستی احتمالی، امیدوارم که این اندک مفید افتد.

مهدی اسماعیلی

۱۳۹۸

فصل ۱

مقدمه

۱-۱ یادگیری ماشین چیست

یادگیری ماشین شاخه‌ای از علم کامپیوتر محسوب می‌شود که در آن با کمک مجموعه‌ای از نمونه‌های پدیده‌های مختلف، الگوریتم‌های سودمندی ساخته شده است. این نمونه‌ها می‌توانند برگرفته از طبیعت، ساخت دست انسان‌ها و یا تولید شده توسط الگوریتم‌ها باشند.

یادگیری ماشین را می‌توان فرایند حل یک مسئله کاربردی دانست که در گام اول آن یک مجموعه داده جمع‌آوری می‌شود و سپس در گام بعدی بر اساس این داده‌های جمع‌آوری شده و به صورت الگوریتمیک، یک مدل آماری ساخته می‌شود. با کمک این مدل آماری، به گونه‌ای می‌توان مسئله کاربردی مورد نظر را حل کرد. در ادامه، برای سادگی به جای عبارت یادگیری ماشین تنها از واژه یادگیری استفاده خواهیم کرد.

۱-۲ انواع یادگیری

گونه‌های مختلف یادگیری عبارتند از: یادگیری باناظر^۱، یادگیری نیمه‌ناظر^۲، یادگیری بی‌ناظر^۳ و یادگیری تقویتی^۴.

1- Supervised
2- Semi-Supervised
3- Unsupervised
4- Reinforcement

۱-۱-۲ یادگیری باناظر

مجموعه داده‌ها در یادگیری باناظر حاوی یک سری از نمونه‌های برچسب‌دار^۱ است که به صورت $\{(x_i, y_i)\}_{i=1}^N$ نشان می‌دهیم. به هر یک از مجموعه عناصر این مجموعه N تایی مانند x_i ، بردار ویژگی^۲ گفته می‌شود. هریک از ابعاد این بردار شامل مقداری است که به گونه‌ای نمونه مورد نظر را توصیف می‌کند. این مقدار به عنوان یک ویژگی شناخته می‌شود و به صورت $x^{(j)}$ نشان می‌دهیم که در آن j می‌تواند مقداری بین یک تا D را به خود بگیرد. برای مثال اگر هر نمونه x در مجموعه داده‌های ما به یک شخص اشاره داشته باشد، $x^{(1)}$ می‌تواند قد آن شخص به سانتی‌متر، $x^{(2)}$ می‌تواند وزن او به کیلوگرم، $x^{(3)}$ می‌تواند جنسیت شخص و همین‌طور ویژگی‌های دیگر می‌توانند مشخصات دیگر آن را بیان کنند. ویژگی j ام بردار ویژگی در تمامی نمونه‌های موجود در داده‌ها حاوی گونه یکسانی از اطلاعات است. بدین معنی که اگر $x_i^{(2)}$ حاوی وزن نمونه x_i باشد، آنگاه $x_k^{(2)}$ به ازای تمامی مقادیر $k = 1, \dots, N$ نیز شامل وزن نمونه‌های دیگر خواهد بود. برچسب y_i می‌تواند گونه‌های متفاوتی را به خود بگیرد. مقدار y_i را می‌توان از مجموعه محدودی از رده‌ها^۳ مانند $\{1, 2, \dots, C\}$ انتخاب کرد، y_i می‌تواند یک عدد (اعشاری) باشد و حتی y_i می‌تواند یک ساختار پیچیده‌تر مانند بردار، ماتریس، درخت یا یک گراف را به خود بگیرد. در این کتاب، ما کاری به این ساختارهای پیچیده نداریم و برچسب y_i یا یکی از عضوهای مجموعه متناهی رده‌ها است و یا یک عدد (اعشاری) خواهد بود. رده‌ها گونه‌ای طبقه‌بندی نمونه‌های موجود در داده‌ها را نشان می‌دهند. برای مثال اگر ایمیل‌ها نمونه‌های ما را تشکیل می‌دهند و مسئله موردنظر ما نیز شناسایی هرزنامه‌ها^۴ است، بنابراین مجموعه $\{spam, not\}$ به مجموعه رده‌ها اشاره می‌کند که در آن نمونه‌های ما به هرزنامه و غیرهرزنامه دسته‌بندی شده‌اند.

هدف یک الگوریتم یادگیری باناظر، تولید یک مدل با کمک مجموعه داده‌ها است؛ مدلی که بردار ویژگی x را به عنوان ورودی پذیرفته و در خروجی اطلاعاتی را قرار می‌دهد که با کمک آن می‌توان برچسب این بردار ویژگی را استنباط کرد. برای مثال مدلی که با استفاده از مجموعه داده‌های مربوط به برخی از افراد خلق شده است، می‌تواند یک بردار ویژگی را به عنوان ورودی دریافت کند (که یک شخص را توصیف می‌کند) و در خروجی احتمال سرطان داشتن آن شخص را نشان دهد.

1- Labeled Examples
2- Feature Vector
3- Classes
4- Spams