

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

یادگیری ماشین در صد صفحه



انتشارات
جهاد دانشگاهی
قزوین

سرشناسه: بورکوف، آندری

Burkov, Andriy

عنوان و نام پدیدآور: یادگیری ماشین در صد صفحه / [آندری بورکوف]؛ مترجمان فاطمه لعل دولت آباد، مریم امیرحسینی؛ ویراستار علمی سیده ربابه میری؛ ویراستار فرحناز بازرگان.

مشخصات نشر: قزوین: جهاد دانشگاهی، سازمان انتشارات، واحد قزوین، ۱۳۹۹.

مشخصات ظاهری: ۲۲۳ ص.؛ مصور، جدول.

شابک: ۹۷۸-۶۲۲-۶۶۴۷-۹۸-۴

وضعیت فهرست نویسی: فیپا

یادداشت: عنوان اصلی: The hundred-page machine learning book, c2019. یادداشت: کتاب حاضر با عنوان «درس‌نامه یادگیری ماشین» با ترجمه مهدی اسماعیلی، فائزه اکبری توسط انتشارات آتی‌نگر: وینادر سال ۱۳۹۸ منتشر شده است.

موضوع: فراگیری ماشینی

موضوع: Machine learning

موضوع: مهندسی مواد -- داده‌پردازی

موضوع: Materials science -- Data processing

شناسه افزوده: لعل دولت آباد، فاطمه، ۱۳۶۰-، مترجم

شناسه افزوده: امیرحسینی، مریم، ۱۳۵۹-، مترجم

شناسه افزوده: میری، سیده ربابه، ۱۳۵۸-، ویراستار

شناسه افزوده: جهاد دانشگاهی. سازمان انتشارات. واحد قزوین

شناسه افزوده: Press Organization Jahade Daneshgahi Ghazvin Branch

رده بندی کنگره: QC۳۲۵/۵

رده بندی دیویی: ۰۰۶/۳۱

شماره کتابشناسی ملی: ۷۳۲۰۶۰۳

عنوان: یادگیری ماشین در صد صفحه

نویسنده: Andriy Burkov

مترجمان: فاطمه لعل دولت آباد، مریم امیرحسینی

ویراستار علمی: سیده ربابه میری

و ویراستار: فرحناز بازرگان

طراح گرافیک و صفحه‌آرا: مرضیه حمیدی زاده

شابک: ۹۷۸-۶۲۲-۶۶۴۷-۹۸-۴

چاپ: نوبت اول - ۱۳۹۹

شمارگان: ۱۰۰۰ نسخه

بهاء: ۴۵۰۰۰۰ ریال

مصوبه شورای شعبه انتشارات جهاد دانشگاهی قزوین

ناشر: انتشارات جهاد دانشگاهی قزوین

کلیه حقوق محفوظ است ©

یادگیری ماشین در صد صفحه

نویسنده: Andriy Burkov

مترجمان:

دکتر فاطمه لعل دولت آباد

عضو هیأت علمی مرکز آموزش عالی فنی و مهندسی بوئین زهرا

دکتر مریم امیرحسینی

عضو هیأت علمی مرکز آموزش عالی فنی و مهندسی بوئین زهرا

ویراستار علمی: دکتر سیده ربابه میری

عضو هیأت علمی مرکز آموزش عالی فنی و مهندسی بوئین زهرا

تقدیم به ساحت مقدّس امام حسن مجتبی (ع)

از نام تو تنهایی می تراود و پرنده‌ها با آسمان آشتی می‌کنند

پیشگفتار

بیا بید با گفتن حقیقت آغاز کنیم: ماشین‌ها یاد نمی‌گیرند. آن‌چه در فرایند «یادگیری ماشین» انجام می‌گیرد، یافتن یک فرمول ریاضی است که وقتی به مجموعه‌ای از ورودی‌ها (با نام «داده‌های آموزش») اعمال می‌شود، خروجی‌های مورد نظر را تولید می‌کند. فرمول ریاضی حاصل، این توانمندی را دارد که برای ورودی‌های متمایز از داده‌های آموزشی، خروجی‌های درست ایجاد کند. البته، به شرطی که این ورودی‌ها از همان (یا مشابه) توزیع آماری که داده‌های آموزشی تهیه شده‌اند، حاصل شوند.

چرا این فرایند، یادگیری نیست؟ زیرا اگر ورودی‌ها را کمی تحریف کنید، احتمالاً خروجی کاملاً اشتباه خواهد شد. این مشابه نحوه یادگیری در حیوانات نیست. اگر یاد گرفته باشیم که با نگاه مستقیم به صفحه نمایش، یک بازی ویدیویی را بازی کنیم؛ اگر کسی صفحه را کمی بچرخاند، باز هم بازیکن خوبی خواهیم بود. در مقابل آن، اگر یک الگوریتم یادگیری ماشین با «نگاه کردن» مستقیم روی صفحه نمایش، آموزش داده شود، قادر به بازی در یک صفحه چرخیده نخواهد بود؛ مگر این‌که آن را برای تشخیص چرخش نیز آموزش داده باشند.

پس چرا این فرایند «یادگیری ماشین» نامیده می‌شود؟ دلیل آن، همان‌طور که اغلب ائتفاق می‌افتد، بازاریابی است: آرتور ساموئل^۱، یکی از پیشگامان آمریکایی در زمینه بازی‌های رایانه‌ای

1 - training data

2 - Arthur Samuel

و هوش مصنوعی، این اصطلاح را در سال ۱۹۵۹ هنگام حضور در IBM ابداع کرد. همان طور که در دهه ۲۰۱۰، IBM تلاش کرد تا اصطلاح «محاسبات شناختی»^۱ را برای باقی ماندن در رقابت، به بازار عرضه کند، در دهه ۱۹۶۰ نیز اصطلاح جالب و جدید «یادگیری ماشین» را برای جذب مشتری و کارمندان با استعداد به کار برد.

همان طور که هوش مصنوعی، دقیقاً مانند هوشمندی، نیست؛ یادگیری ماشین نیز یادگیری نیست. با این حال، یادگیری ماشین یک اصطلاح جهانی شناخته شده است که معمولاً به علوم و مهندسی ماشین آلات ساختمانی اشاره دارد که قادر به انجام کارهای مفید و مختلف هستند؛ بدون این که صریحاً برای انجام این کار برنامه ریزی شده باشند. بنابراین، کلمه «یادگیری» به اصطلاح و در قیاس با یادگیری حیوانات و نه به معنای واقعی کلمه استفاده می‌گردد.

این کتاب برای چه کسانی است؟

این کتاب تنها بخش‌هایی از مطالب گسترده در زمینه یادگیری ماشین را در بردارد که از دهه ۱۹۶۰ تا کنون توسعه یافته‌اند و ارزش عملی قابل توجه آن‌ها به اثبات رسیده است. یک مبتدی در یادگیری ماشین، در این کتاب فقط جزئیات لازم را پیدا می‌کند تا به سطحی از درک موضوع برسد و شروع به پرسیدن سؤالات مناسب کند.

افراد با تجربه می‌توانند از این کتاب به عنوان مجموعه‌ای از دستورالعمل‌ها برای ارتقای بیشتر خود بهره ببرند. این کتاب برای ایده‌پردازی^۲ در ابتدای یک پروژه نیز مفید است. در این هنگام سعی می‌کنیم به این پرسش پاسخ دهیم که آیا مشکل فنی یا تجاری خاصی که با آن سر و کار داریم با بهره‌گیری از یادگیری ماشینی قابل حل است و در ادامه، اگر پاسخ مثبت است، کدام تکنیک‌ها برای حل آن مناسب هستند.

چگونه می‌توان از این کتاب استفاده کرد؟

اگر قصد دارید یادگیری ماشین را بیاموزید، باید این کتاب را از ابتدا تا انتها بخوانید. اگر به یک موضوع خاص مطرح شده در کتاب علاقه‌مندید و می‌خواهید اطلاعات بیشتری کسب کنید، بیشتر بخش‌ها دارای کد پاسخ سریع^۳ هستند.

1 - cognitive computing

2 - brainstorming

3 - Quick Response code

با اسکن یکی از کدهای پاسخ سریع با استفاده از موبایل خود، به پیوند صفحه‌ای در ویکی همراه کتاب (theMLbook.com) دست خواهید یافت که شامل این موارد است: پیشنهاد مطالعه‌کنندگان، ویدئوها، پرسش و پاسخ‌های قطعه‌های رمز، آموزش و پاداش‌های دیگر. ویکی کتاب به‌طور مداوم با کمک نویسنده و هم‌چنین داوطلب‌هایی از سراسر جهان به‌روز می‌شود. بنابراین، کتاب حاضر، بعد از خرید هم‌چنان بهتر می‌شود. برای رسیدن به ویکی کتاب، کد پاسخ سریع سمت چپ را اسکن کنید.

بعضی از بخش‌ها کد پاسخ سریع ندارند، اما به احتمال زیاد هنوز در صفحه ویکی وجود داشته باشند. با ارسال عنوان بخش به موتور جستجوی ویکی، می‌توانید آن‌ها را بیابید.



QR Code

آیا باید این کتاب را بخرید؟

این کتاب بر پایه این اصل توزیع شده است «ابتدا بخوانید، سپس، بخرید». شما می‌توانید یک اتومبیل را قبل از خرید در نمایندگی ببینید و امتحان کنید و هم‌چنین، می‌توانید پیراهن یا لباس را در یک فروشگاه امتحان کنید. من قاطعانه معتقدم که باید بتوانید قبل از پرداخت هزینه، کتاب را بخوانید.

اصل ابتدا بخوانید، بعدا بخرید دلالت بر این دارد که می‌توانید کتاب را آزادانه بازگذاری کنید، آن را بخوانید و با دوستان و همکاران خود به اشتراک بگذارید. در صورتی که کتاب را خواندید و دوست داشتید، یا آن را کمک‌کننده و به هر صورتی مفید یافتید، آن را بخرید.

اکنون همه شما آماده‌اید.
از خواندن خود لذت ببرید!

فهرست

۱	فصل ۱؛ معرفی
۳	۱-۱ یادگیری ماشین چیست؟
۳	۲-۱ انواع یادگیری
۴	۱-۲-۱ یادگیری نظارت شده
۵	۲-۲-۱ یادگیری نظارت نشده
۵	۳-۲-۱ یادگیری نیمه نظارتی
۵	۴-۲-۱ یادگیری تقویتی
۶	۳-۱ یادگیری نظارت شده چگونه عمل می‌کند؟
۱۱	۴-۱ چرا مدل بر روی داده‌های جدید قابل اجرا است؟
۱۳	فرهنگ لغات تخصصی فصل ۱
۱۵	فصل ۲؛ نمادها و تعریف‌ها
۱۷	۱-۲ نماد
۱۷	۱-۱-۲ ساختمان داده‌ها
۱۹	۲-۱-۲ نماد سیگمای بزرگ
۱۹	۳-۱-۲ نماد پای بزرگ

۲۰	۴-۱-۲ عملیات روی مجموعه‌ها
۲۰	۵-۱-۲ عملیات روی بردارها
۲۲	۶-۱-۲ توابع
۲۳	۷-۱-۲ Arg Max و Max
۲۳	۸-۱-۲ عملگر تخصیص
۲۳	۹-۱-۲ مشتق و گرادیان
۲۵	۲-۲ متغیر تصادفی
۲۷	۳-۲ برآوردگر ناریب
۲۸	۴-۲ قاعده بیز
۲۸	۵-۲ برآورد پارامتر
۲۹	۶-۲ مقایسه پارامتر و ابرپارامتر
۳۰	۷-۲ مقایسه دسته‌بندی و رگرسیون
۳۱	۸-۲ مقایسه یادگیری مبتنی بر مدل و یادگیری مبتنی بر نمونه
۳۱	۹-۲ مقایسه یادگیری سطحی و یادگیری عمیق
۳۵	فصل ۳: الگوریتم‌های بنیادی
۳۷	۱-۳ رگرسیون خطی
۳۷	۱-۱-۳ طرح مسأله
۳۹	۲-۱-۳ راه حل
۴۲	۲-۳ رگرسیون منطقی
۴۲	۱-۲-۳ طرح مسأله
۴۴	۲-۲-۳ حل مسأله
۴۵	۳-۳ یادگیری درخت تصمیم‌گیری
۴۵	۱-۳-۳ طرح مسأله
۴۶	۲-۳-۳ راه حل
۵۰	۴-۳ ماشین بردار پشتیبان
۵۱	۱-۴-۳ برخورد با نویز
۵۲	۲-۴-۳ برخورد با غیرخطی بودن ذاتی
۵۴	۵-۳ کی-نزدیک‌ترین همسایه

فرهنگ لغات تخصصی فصل ۳	۵۶
فصل ۴؛ آناتومی یک الگوریتم یادگیری	۵۷
۱-۴ بلوک‌های ساختمان یک الگوریتم یادگیری	۵۹
۲-۴ گرادیان کاهشی	۶۰
۳-۴ چگونه مهندسان یادگیری ماشین، کار می‌کنند؟	۶۶
۴-۴ ویژگی‌های الگوریتم‌های یادگیری	۶۷
فرهنگ لغات تخصصی فصل ۴	۶۸
فصل ۵؛ تمرین پایه‌ای	۶۹
۱-۵ مهندسی ویژگی	۷۱
۱-۱-۵ کدگذاری تک‌فعال	۷۲
۲-۱-۵ پیکال بندی	۷۳
۳-۱-۵ به‌هنجارسازی	۷۴
۴-۱-۵ استانداردسازی	۷۵
۵-۱-۵ برخورد با ویژگی‌های از دست رفته	۷۶
۶-۱-۵ تکنیک‌های جانمایی داده	۷۶
۲-۵ انتخاب الگوریتم یادگیری	۷۷
۳-۵ سه مجموعه	۸۰
۴-۵ کم برآزش و بیش برآزش	۸۱
۵-۵ نظم‌دهی	۸۴
۶-۵ ارزیابی عملکرد مدل	۸۶
۱-۶-۵ ماتریس درهم‌ریختگی	۸۷
۲-۶-۵ دقت / بازخوانی	۸۸
۳-۶-۵ صحت	۹۰
۴-۶-۵ دقت حساس به هزینه	۹۱
۵-۶-۵ سطح زیرمنحنی (ROC) AUC	۹۱
۷-۵ بهینه کردن ابرپارامترها	۹۲
۱-۷-۵ اعتبارسنجی متقاطع	۹۴
فرهنگ لغات تخصصی فصل ۵	۹۶

۹۹.....	فصل ۶؛ شبکه‌های عصبی و یادگیری عمیق
۱۰۱.....	۱-۶ شبکه‌های عصبی
۱۰۳.....	۱-۱-۶ مثال پرسپترون چند لایه
۱۰۴.....	۲-۱-۶ معماری شبکه عصبی پیش خور
۱۰۷.....	۲-۶ یادگیری عمیق
۱۰۸.....	۱-۲-۶ شبکه عصبی پیچیده
۱۱۷.....	۲-۲-۶ شبکه عصبی بازگشتی
۱۲۳.....	فرهنگ لغات تخصصی فصل ۶
۱۲۵.....	فصل ۷؛ مسائل و راه حل‌ها
۱۲۷.....	۱-۷ رگرسیون هسته
۱۲۹.....	۲-۷ دسته‌بندی چندطبقه‌ای
۱۳۰.....	۳-۷ دسته‌بندی یک طبقه‌ای
۱۳۳.....	۴-۷ طبقه‌بندی چند برچسبی
۱۳۵.....	۵-۷ یادگیری گروهی
۱۳۶.....	۱-۵-۷ یادگیری تقویتی ورده‌بندی
۱۳۷.....	۲-۵-۷ جنگل تصادفی
۱۳۸.....	۳-۵-۷ گرادیان تقویتی
۱۴۱.....	۶-۷ یادگیری برچسب زدن دنباله‌ها
۱۴۳.....	۷-۷ یادگیری دنباله به دنباله
۱۴۵.....	۸-۷ یادگیری فعال
۱۴۷.....	۹-۷ یادگیری نیمه نظارتی
۱۵۱.....	۱۰-۷ یادگیری یک-شات
۱۵۳.....	۱۱-۷ یادگیری صفرشات
۱۵۶.....	فرهنگ لغات تخصصی فصل ۷
۱۵۹.....	فصل ۸؛ تمرین پیشرفته
۱۶۱.....	۱-۸ مدیریت مجموعه داده‌های نامتوازن
۱۶۳.....	۲-۸ ترکیب مدل‌ها

۱۶۴	۳-۸ آموزش شبکه‌های عصبی.....
۱۶۶	۴-۸ نظم‌دهی پیشرفته.....
۱۶۷	۵-۸ مدیریت ورودی‌های چندگانه.....
۱۶۸	۶-۸ مدیریت خروجی‌های متعدد.....
۱۶۹	۷-۸ یادگیری انتقالی.....
۱۷۱	۸-۸ کارایی الگوریتمی.....
۱۷۴	فرهنگ لغات تخصصی فصل ۸.....
۱۷۵	فصل ۹؛ یادگیری نظارت‌نشده.....
۱۷۷	۱-۹ تخمین چگالی.....
۱۸۰	۲-۹ خوشه‌بندی.....
۱۸۱	۱-۲-۹ کی-میانگین‌ها.....
۱۸۲	۲-۲-۹ اسکن DB و اسکن HDB.....
۱۸۳	۳-۲-۹ تعیین تعداد خوشه‌ها.....
۱۸۷	۴-۲-۹ الگوریتم‌های خوشه‌بندی دیگر.....
۱۹۱	۳-۹ کاهش ابعاد.....
۱۹۲	۱-۳-۹ تجزیه و تحلیل مؤلفه اصلی.....
۱۹۳	۲-۳-۹ UMAP.....
۱۹۵	۴-۹ تشخیص پرت.....
۱۹۶	فرهنگ لغات تخصصی فصل ۹.....
۱۹۷	فصل ۱۰؛ اشکال دیگر یادگیری.....
۱۹۹	۱-۱۰ یادگیری متریک.....
۲۰۲	۲-۱۰ یادگیری مبتنی بر رتبه‌بندی.....
۲۰۶	۳-۱۰ یادگیری بر مبنای توصیه.....
۲۰۸	۱-۳-۱۰ ماشین فاکتورسازی.....
۲۰۹	۲-۳-۱۰ کدگذار خودکار بدون نویز.....
۲۱۱	۴-۱۰ یادگیری خودنظارتی: تعبیه‌سازی کلمات.....
۲۱۴	فرهنگ لغات تخصصی فصل ۱۰.....

۲۱۵.....	فصل ۱۱؛ نتیجه‌گیری
۲۱۸	۱-۱۱ آن‌چه در کتاب نیامده است
۲۱۸	۱-۱-۱۱ مدل‌سازی موضوع
۲۱۸	۲-۱-۱۱ فرآیندهای گاوسی
۲۱۸	۳-۱-۱۱ مدل خطی تعمیم‌یافته
۲۱۹	۴-۱-۱۱ مدل‌های گرافیکی احتمالی
۲۱۹	۵-۱-۱۱ زنجیره مارکوف مونت‌کارلو
۲۲۰	۶-۱-۱۱ شبکه‌های مولد تخصصی
۲۲۰	۷-۱-۱۱ الگوریتم‌های ژنتیکی
۲۲۱	۸-۱-۱۱ یادگیری تقویتی
۲۲۲.....	۲-۱۱ تشکر و قدردانی
۲۲۳	فرهنگ لغات تخصصی فصل ۱۱

معرفی

فصل ۱

۱-۱ یادگیری ماشین چیست؟

یادگیری ماشین زیر شاخه‌ای از علوم کامپیوتر است که به ساختن الگوریتم‌ها مربوط می‌شود و به منظور اثربخش بودن، به مجموعه‌ای از نمونه‌های یک پدیده مَثکی است. این نمونه‌ها می‌توانند از طبیعت به دست آیند، به وسیله انسان ساخته شوند و یا به وسیله الگوریتم دیگری تولید شوند. یادگیری ماشین را می‌توان برای حلّ یک مسأله عملی نیز به کار گرفت. این کار به روش‌های زیر انجام می‌گیرد:

۱- جمع‌آوری مجموعه داده

۲- الگوریتمی کردن یک مدل آماری بر اساس آن مجموعه داده.

فرض بر این است که مدل آماری به نوعی برای حل مسأله عملی مورد استفاده قرار می‌گیرد. برای صرفه جویی در نگارش، از این پس اصطلاحات "یادگیری" و "یادگیری ماشین" به جای هم استفاده می‌شوند.

۱-۲ انواع یادگیری

یادگیری می‌تواند نظارت‌شده، نیمه نظارتی، نظارت‌نشده و تقویتی باشد.

۱-۲-۱ یادگیری نظارت شده

در یادگیری نظارت شده^۱، مجموعه داده^۲، مجموعه‌ای از نمونه‌های برچسب‌دار^۳ بردار $\{(x_i, y_i)\}_{i=1}^N$ است. هر عنصر x_i از میان N بردار، یک بردار ویژگی^۴ نامیده می‌شود. بردار ویژگی، برداری است که در هر بعد $j=1, \dots, D$ ، حاوی مقداری است که چگونگی نمونه را توصیف می‌کند. این مقدار را یک ویژگی می‌نامند و با $x^{(i)}$ نمایش می‌دهند. به عنوان مثال، اگر هر نمونه، x در گردایه^۵ نمایانگر یک شخص باشد، آنگاه اولین ویژگی، $x^{(1)}$ ، می‌تواند قد به سانتی‌متر و ویژگی دوم، $x^{(2)}$ ، وزن به کیلوگرم، $x^{(3)}$ بیانگر جنسیت باشد و الی آخر. برای همه مثال‌های موجود در مجموعه داده، ویژگی موجود در موقعیت z در بردار ویژگی، همیشه شامل همان نوع اطلاعات است. به عبارتی دیگر، اگر $x_i^{(2)}$ در برگزیده^۶ وزن به کیلوگرم در نمونه x_i باشد، آنگاه $x_k^{(2)}$ نیز بیانگر وزن به کیلوگرم در هر نمونه x_k است که $k=1, \dots, N$ است. برچسب y_i می‌تواند یکی از عناصر متعلق به مجموعه^۷ متناهی از دسته‌های^۸ $\{1, 2, \dots, C\}$ ، یک عدد حقیقی، یا یک ساختار پیچیده‌تر مانند بردار، ماتریس، درخت و یا گراف باشد. در این کتاب y_i مجموعه‌ای متناهی از کلاس‌ها یا یک عدد حقیقی^۹ است مگر این‌که خلاف آن گفته شود. می‌توان یک کلاس را به عنوان گردایه‌ای در نظر گرفت که یک نمونه به آن تعلق دارد. به عنوان مثال، اگر نمونه‌ها پیام‌های ایمیل باشند و مسأله ما شناسایی هرزنانه باشد، پس دو کلاس $\{\text{هرزنانه، غیر هرزنانه}\}$ داریم.

هدف از الگوریتم یادگیری نظارت شده، استفاده از مجموعه داده برای تولید یک مدل است. مدل، بردار ویژگی x را به عنوان ورودی دریافت می‌کند و اطلاعاتی استخراج می‌کند که استنباط یک برچسب برای این بردار ویژگی را میسر می‌کند. به طور نمونه، مدل ایجاد شده با استفاده از مجموعه داده افراد، می‌تواند یک بردار ویژگی را به عنوان ورودی دریافت کند که توصیفی از یک شخص است. در نهایت، احتمال این‌که آن شخص سرطان دارد را در خروجی نشان می‌دهد.

1- supervised learning

2- dataset

3- labeled examples

4- feature vector

5- collection

6- classes

۷- یک عدد حقیقی، کمیتی است که می‌تواند بیانگر فاصله در یک خط باشد. مثال‌ها: ۰، ۳۴، ۲۵۶، ۱۰۰۰، ۲، ۱۰۰۰.

۱-۲-۲ یادگیری نظارت نشده

در یادگیری نظارت نشده^۶، مجموعه داده، مجموعه‌ای از نمونه‌های بدون برچسب $\{x_i\}_{i=1}^N$ است. در این جا نیز x ، یک بردار ویژگی است. هدف از الگوریتم یادگیری نظارت نشده^۷ تولید مدلی است که بردار ویژگی x را به عنوان ورودی دریافت کند و آن را به برداری دیگر یا به یک مقدار بنگارد که در حل مسائل عملی کاربرد دارد. برای مثال، در خوشه بندی^۸، مدل، کد هر خوشه را برای هر بردار ویژگی در مجموعه داده نظیر می‌کند. در کاهش ابعاد^۵، خروجی مدل یک بردار ویژگی است که ویژگی‌های کمتری نسبت به بردار ورودی x دارد. در تشخیص داده‌های پرت^۹، خروجی یک عدد حقیقی است که مشخص می‌کند که x چقدر متفاوت از نمونه "معمولی" از مجموعه داده است.

۱-۲-۳ یادگیری نیمه نظارتی

در یادگیری نیمه نظارتی^۷، مجموعه داده شامل نمونه‌های برچسب دار و بدون برچسب است. به طور معمول، از لحاظ کمی نمونه‌های بدون برچسب بسیار بیشتر از نمونه‌های برچسب دار هستند. هدف الگوریتم یادگیری نیمه نظارتی، مشابه هدف الگوریتم یادگیری نظارت شده است. به این امید که با استفاده از تعداد زیادی نمونه‌های بدون برچسب به الگوریتم یادگیری کمک شود تا مدل بهتری پیدا کند (ما ممکن است از واژه "تولید کردن" یا "محاسبه کردن" استفاده کنیم).

۱-۲-۴ یادگیری تقویتی

یادگیری تقویتی^۸ زیرشاخه‌ای از یادگیری ماشین است که در آن، ماشین مورد نظر در محیط "زندگی می‌کند" و قابلیت این را دارد که حالت^۹ آن محیط را به عنوان یک بردار ویژگی دریافت کند. دستگاه

-
- 1- unsupervised learning
 - 2- unlabeled examples
 - 3- unsupervised learning algorithm
 - 4- clustering
 - 5- dimensionality reduction
 - 6- outlier detection
 - 7- semi-supervised learning
 - 8- Reinforcement learning
 - 9- state

در هر حالتی می‌تواند اعمالی^۱ را اجرا کند. اعمال متفاوت، پاداش‌های^۲ متفاوتی را به همراه دارند و قادر است که ماشین را به حالت دیگری از محیط منتقل کند. هدف از یادگیری تقویتی یادگیری یک خط‌مشی^۳ است. خط‌مشی، تابعی مانند f است (مشابه مدل در یادگیری نظارت‌شده) که بردار ویژگی حالت را به عنوان ورودی دریافت می‌کند و خروجی، بهترین عملکرد اجرایی در آن حالت است. عملکرد در شرایطی بهینه است که امید ریاضی متوسط پاداش^۴ را بیشینه کند.

یادگیری تقویتی، نوع خاصی از مسائل را حل می‌کند که تصمیم‌گیری در آن‌ها متوالی و دف آن‌ها بلند مدت است. از جمله بازی، روباتیک، مدیریت منابع یا لجستیک. در این کتاب، تأکید بر تصمیم‌گیری یک-شات^۵ است که در آن نمونه‌های ورودی مستقل از یکدیگر هستند و پیش‌بینی‌ها قبلاً انجام شده‌اند. یادگیری تقویتی خارج از محدوده این کتاب است.



۱-۳ یادگیری نظارت‌شده چگونه عمل می‌کند؟

در این بخش، به طور خلاصه توضیح می‌دهیم که یادگیری نظارت‌شده چگونه عمل می‌کند. تا قبل از پرداختن به جزئیات، از کلیات فرایند شهودی داشته باشید. در اینجا تصمیم داریم از یادگیری نظارت‌شده مثالی بیاوریم؛ زیرا این نوع یادگیری ماشین در عمل کاربرد گسترده‌ای دارد.

فرایند یادگیری نظارت‌شده با جمع‌آوری داده‌ها شروع می‌شود. داده‌ها برای یادگیری نظارت‌شده، مجموعه‌ای از زوج‌های مرتب (ورودی، خروجی) هستند. ورودی می‌تواند هر چیزی باشد، برای مثال، پیام‌های ایمیل، تصاویر یا اندازه‌گیرهای حس‌گرها. خروجی‌ها معمولاً اعداد

-
- 1- action
 - 2- reward
 - 3- policy
 - 4- expected average reward
 - 5- one-shot decision

حقیقی یا برجسب‌ها هستند (به‌عنوان مثال "هرزنامه"، "غیر هرزنامه"، "گره"، "سگ"، "موش" و غیره). در بعضی مواقع خروجی‌ها، بردار (به‌عنوان مثال، چهار مختصه مستطیل اطراف یک شخص در تصویر)، دنباله (به‌عنوان مثال ["اسم"، "صفت"، "صفت"]), برای ورودی "ماشین بزرگ زیبا" یا ساختارهای دیگر هستند.

فرض کنید مسأله‌ای که می‌خواهیم با استفاده از یادگیری نظارت‌شده حل کنیم، شناسایی هرزنامه باشد. ما داده‌ها را جمع‌آوری می‌کنیم. به‌عنوان مثال، ۱۰,۰۰۰ پیام ایمیل، هر کدام دارای یک برجسب "هرزنامه" یا "غیر هرزنامه" هستند (می‌توان برجسب‌ها را به‌صورت دستی اضافه کرد یا این کار را به کسی واگذار کرد). اکنون، باید هر پیام ایمیل به یک بردار ویژگی تبدیل شود.

تحلیل‌گر داده بر اساس تجربه خود، تصمیم می‌گیرد که چگونه موجودی در دنیای واقعی (یک پیام ایمیل) را به یک بردار ویژگی تبدیل کند. روشی معمول که برای تبدیل متن به بردار ویژگی استفاده می‌شود **کوله‌ی کلمات**^۱ نام دارد. بر اساس این روش، یک فرهنگ لغات از کلمات تهیه کنیم (به فرض، این فرهنگ لغات شامل ۲۰۰۰۰ کلمه مرتب شده بر اساس حروف الفبا باشد) و تصریح می‌کنیم که در بردار ویژگی:

- اگر پیام ایمیل حاوی کلمه "a" باشد، اولین ویژگی برابر ۱ است. در غیر این صورت، این ویژگی صفر است،
- اگر پیام ایمیل حاوی کلمه "aaron" باشد، ویژگی دوم برابر ۱ است. در غیر این صورت، این ویژگی برابر صفر است،
- اگر پیام ایمیل حاوی کلمه "zulu" باشد، ویژگی در جایگاه ۲۰,۰۰۰ برابر با ۱ است، در غیر این صورت، این ویژگی برابر با صفر است.

چنانچه روش بالا را برای هر پیام ایمیل موجود در مجموعه خودمان تکرار کنیم، ۱۰,۰۰۰ بردار ویژگی خواهیم داشت بدین شکل که هر بردار دارای ۲۰,۰۰۰ بعد و یک برجسب ("هرزنامه"/ "غیر هرزنامه") است.

اکنون، داده ورودی قابل خواندن با ماشین در اختیار داریم، اما برجسب‌های خروجی هنوز به‌شکل متن و قابل خواندن برای انسان هستند. برخی از الگوریتم‌های یادگیری نیازمند تبدیل برجسب‌ها به اعداد هستند. به‌عنوان مثال، برخی از الگوریتم‌ها به اعدادی مانند صفر

(برای نشان دادن برجسب "غیرهرزنامه") و ۱ (برای نشان دادن برجسب "هرزنامه") نیاز دارند. الگوریتمی که برای نشان دادن یادگیری نظارت شده استفاده می‌کنیم **ماشین بردار پشتیبان** (SVM) نامیده می‌شود. این الگوریتم نیاز دارد که برجسب مثبت (برای هرزنامه) دارای مقدار عددی +۱ و برجسب منفی (غیر هرزنامه) دارای مقدار -۱ باشد.

در این مرحله، ما یک مجموعه داده و یک **الگوریتم یادگیری**^۲ داریم. بنابراین، آماده هستیم تا الگوریتم یادگیری را روی مجموعه داده به کار ببندیم تا یک مدل دریافت کنیم. SVM، هر بردار ویژگی را به عنوان نقطه‌ای در فضای با ابعاد بالا در نظر می‌گیرد (در نمونه مورد بحث ما فضا ۲۰۰۰۰ بعد دارد). این الگوریتم، همه بردارهای ویژگی را روی طرح فرضی ۲۰۰۰۰ بعدی قرار می‌دهد و یک خط فرضی با بعد ۱۹۹۹۹ (یک ابر صفحه) می‌کشد که نمونه‌های با برجسب‌های مثبت را از نمونه‌های با برجسب‌های منفی افراز می‌کند. در یادگیری ماشین، مرز جدا کننده نمونه‌های کلاس‌های مختلف، **مرز تصمیم‌گیری**^۳ نامیده می‌شود. معادله ابر صفحه، به وسیله دو پارامتر بیان می‌شود که عبارتند از بردار حقیقی w با بعدی برابر بعد بردار ویژگی ورودی x ، و یک عدد حقیقی b :

$$wx - b = 0$$

که در آن عبارت $w \cdot x$ به معنی $w^{(1)}x^{(1)} + w^{(2)}x^{(2)} + \dots + w^{(D)}x^{(D)}$ و D عددی برابر با بعد بردار ویژگی x است. (اگر برخی از معادلات در حال حاضر برای شما روشن نیستند، در فصل ۲ دوباره به بررسی ریاضیات و مفاهیم آماری لازم برای درک آن‌ها می‌پردازیم. در حال حاضر، سعی کنید یک شهود از آن‌چه اتفاق می‌افتد، به دست آورید. همه چیز بعد از خواندن فصل بعدی واضح‌تر می‌شود). اکنون، برجسب پیش‌بینی شده برای برخی از بردارهای ویژگی ورودی x به صورت زیر داده می‌شود:

$$y = \text{sign}(wx - b)$$

که در آن sign عملگری ریاضی است که یک مقدار را به عنوان ورودی دریافت می‌کند و در صورتی که ورودی عددی مثبت باشد آن را به +۱ می‌نگارد. چنانچه ورودی عددی منفی باشد آن را به -۱ می‌نگارد.

1- Support Vector Machine

2- learning algorithm

3- decision boundary

هدف از الگوریتم یادگیری - SVM در این مورد- استفاده از مجموعه داده و یافتن مقادیر بهینه w^* و b^* برای پارامترهای w و b است. هنگامی که الگوریتم یادگیری این مقادیر بهینه را شناسایی کند، آنگاه مدل $f(x)$ به صورت زیر تعریف می‌شود:

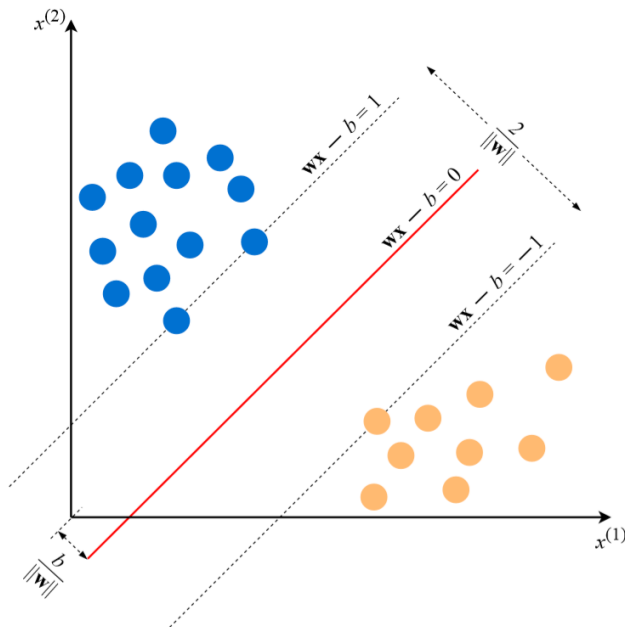
$$f(x) = \text{sign}(w^*x - b^*)$$

بنابراین، برای محاسبه این که پیام ایمیل هرزنامه یا غیرهرزنامه است از مدل SVM استفاده می‌کنیم. روال کار بدین شکل است که ابتدا یک متن از پیام‌ها را دریافت کرده و به بردار ویژگی تبدیل کنیم. سپس آن را در بردار w^* ضرب کنیم و بردار b^* را از آن کم کنیم و علامت نتیجه را دریافت کنیم. این فرایند پیش‌بینی مورد نظر را به ما خواهد داد (+۱ به معنی "هرزنامه"، و -۱ به معنی "غیرهرزنامه"). حال سؤال این است که ماشین چگونه w^* و b^* را می‌یابد؟ در این حالت ماشین فرایند بهینه‌سازی را انجام می‌دهد. ماشین‌ها در بهینه‌سازی توابع با شرایط مشخص شده، عملکرد خوبی دارند.

سؤال بعدی این است که چه شرط‌هایی باید اعمال شوند؟ در درجه اول، مدلی می‌خواهیم که برچسب‌هایی برای ۱۰۰۰۰ نمونه را به درستی به دست آورد. به خاطر آوری که هر نمونه $i = 1, \dots, 10000$ با دوتایی (x_i, y_i) تعریف می‌شود که x_i بردار ویژگی نمونه i و y_i برچسب آن است که یکی از مقادیر +۱ یا -۱ را می‌گیرد. بنابراین، شرایط به‌طور طبیعی به صورت زیر هستند:

- اگر $y_i = +1$ ، آنگاه $w x_i - b \geq 1$ و
- اگر $y_i = -1$ ، آنگاه $w x_i - b \leq -1$.

ما هم‌چنین ترجیح می‌دهیم که ابرصفحه، نمونه‌های مثبت را از موارد منفی با بزرگترین حاشیه^۱ جدا کند. این نوار، فاصله نزدیک‌ترین نمونه‌های دو کلاس است، چرا که به‌عنوان مرز تصمیم‌گیری تعریف شده است. یک حاشیه بزرگ به تعمیم بهتر کمک می‌کند، به عبارتی دیگر، مدل عملکرد خوبی در طبقه‌بندی نمونه‌های جدید خواهد داشت. برای رسیدن به آن، نیاز داریم که نرم اقلیدسی w را که با $\|w\|$ نمایش داده می‌شود و با $\sqrt{\sum_{j=1}^D (w^{(j)})^2}$ تعریف می‌شود، کمینه کنیم. بنابراین، به نظر می‌آید مسأله بهینه‌سازی که علاقه‌مندیم ماشین آن را حل کند به شکل زیر است:



شکل ۱-۱: مثالی از یک مدل SVM برای بردار ویژگی دوبعدی

کمینه کردن $\|w\|$ تحت شرایط $y_i (wx_i - b) \geq 1$ که $i = 1, \dots, N$. عبارت $y_i (wx_i - b) \geq 1$ دقیقاً خلاصه دو شرط فوق است. پاسخ بهینه‌سازی با w^* و b^* نشان داده می‌شود و مدل آماری، یا ساده‌تر، مدل نامیده می‌شود. فرایند ساخت مدل، آموزش نام دارد.

برای بردارهای ویژگی دوبعدی، چنان‌چه در شکل ۱ به تصویر کشیده شده است، مسأله و راه‌حل قابل مشاهده است. دایره‌های آبی و نارنجی، به ترتیب، نمونه‌های مثبت و منفی را نشان می‌دهند و خط با معادله $wx - b = 0$ مرز تصمیم‌گیری است.

چرا با به حداقل رساندن نرم w ، بیشترین حاشیه را بین این دو دسته به دست می‌آوریم.

چنان‌چه در شکل ۱ مشاهده می‌شود، از نظر هندسی، معادلات $wx - b = 1$ و $wx - b = -1$ دو ابرصفحه موازی را تعریف می‌کنند. فاصله بین این ابرصفحه‌ها با $\frac{2}{\|w\|}$ ، نشان داده شده است. بنابراین، $\|w\|$ کوچک‌تر به معنای فاصله بزرگ‌تر بین این دو ابرصفحه است. ماشین‌های بردار پشتیبان این‌گونه کار می‌کنند. این نسخه خاص از الگوریتم، مدل مشهور خطی را ساختار بندی

می‌کند. چون مرز تصمیم‌گیری یک خط مستقیم (صفحه یا ابرصفحه) است، مدل خطی^۱ نامیده می‌شود. SVM هم‌چنین می‌تواند **هسته‌هایی**^۲ را در اختیار ما قرار دهد که می‌توانند مرز تصمیم‌گیری را به‌طور دلخواه غیرخطی کنند. در بعضی موارد، به دلیل وجود نویز در داده‌ها، خطاهای برچسب زدن و یا پرت بودن (نمونه‌های بسیار متفاوت از یک نمونه "معمولی" مجموعه داده)، جداسازی کامل دو دسته از نقاط غیرممکن می‌شود. نسخه دیگری از SVM نیز می‌تواند یک ابرپارامتر اضافی^۳ جریمه برای طبقه‌بندی نادرست نمونه‌های آموزشی کلاس‌های خاص وارد کند. ما الگوریتم SVM را با جزئیات بیشتر در فصل ۳ مطالعه می‌کنیم.

در این مرحله، باید موارد زیر را به ذهن بسپاریم: هر الگوریتم یادگیری طبقه‌بندی شده، یک مدل ایجاد می‌کند که به‌طور ضمنی یا صریح یک مرز تصمیم تولید می‌کند. مرز تصمیم می‌تواند یک خط مستقیم، منحنی، شکلی پیچیده یا منطبق بر یک شکل هندسی باشد. شکل مرز تصمیم‌گیری، **صحت**^۴ مدل را تعیین می‌کند (نسبت تعداد نمونه‌هایی که برچسب‌های آن‌ها به درستی پیش‌بینی شده، به تعداد کل برچسب‌های پیش‌بینی شده). شکل مرز تصمیم‌گیری و نحوه محاسبه الگوریتمی یا ریاضیاتی مبنی بر داده‌های آموزشی، یک الگوریتم یادگیری را از دیگری متمایز می‌کند.

به‌طور ویژه، دو تفاوت اساسی دیگر در الگوریتم‌های یادگیری وجود دارد که به آن می‌پردازیم: سرعت ساخت مدل و زمان پردازش پیش‌بینی. در بسیاری از موارد عملی، الگوریتم یادگیری ای را ترجیح می‌دهیم که مدلی با صحت کمتر اما با سرعت بیشتر تولید کند.

۱-۴ چرا مدل بر روی داده‌های جدید قابل اجرا است؟

چرا مدلی که ماشین یاد گرفته است، قادر به پیش‌بینی درست برچسب‌های جدید (نمونه‌هایی که قبلاً ندیده است) است؟ برای درک این موضوع، به نمودار شکل ۱ نگاه کنید. اگر دو کلاس با مرز تصمیم از هم مجزا شده باشند، آن‌گاه، بدیهی است، نمونه‌هایی که متعلق به آن‌ها هستند در دو زیرفضای مختلف که مرز تصمیم‌گیری ایجاد می‌کند، قرار دارند.

1 - Linear model

2 - kernel

۳ - یک ابرپارامتر خاصیتی از الگوریتم یادگیری است، معمولاً (نه همیشه) مقدار عددی دارد. این مقدار روی روش کار الگوریتم تأثیر می‌گذارد. این مقادیر به وسیله الگوریتم از روی داده‌ها یاد گرفته نشده‌اند. آن‌ها باید به وسیله تحلیل‌گر داده قبل از اجرای الگوریتم تنظیم شوند.

4 - accuracy

اگر نمونه‌هایی که برای آموزش استفاده می‌شوند، به‌طور تصادفی، مستقل از یکدیگر انتخاب شوند و روند یکسانی را دنبال کنند، آن‌گاه، از نظر آماری نمونه جدید منفی، احتمالاً، در شکل جایی واقع می‌شود که چندان از نمونه‌های منفی دیگر دور نباشد. همین مسأله برای نمونه مثبت جدید نیز روی می‌دهد و احتمالاً از حوالی دیگر نمونه‌های مثبت آمده‌اند. در چنین مواردی، مرز تصمیم ما هم‌چنان با احتمال بالایی، نمونه‌های مثبت و منفی جدید را به‌خوبی از هم جدا می‌کند. در موارد نادری مدل، خطا می‌کند، اما از آن‌جا که این موارد نادر هستند، تعداد خطاها از تعداد پیش‌بینی‌های صحیح کوچک‌تر خواهد بود.

به‌طور شهودی، مجموعه‌ای شامل نمونه‌های آموزشی فراوان سبب می‌گردد که نمونه‌های جدید با نمونه‌هایی که برای آموزش استفاده می‌شوند، چندان متفاوت (و از لحاظ طرح، دور از آن‌ها) نباشند. برای به حداقل رساندن احتمال خطا در نمونه‌های جدید، الگوریتم SVM، با جست‌وجوی حاشیه پهن‌تر، به‌طور صریح سعی می‌کند مرز تصمیم را به گونه‌ای رسم کند که تا آن‌جا که ممکن است به یک فاصله از نمونه‌های هر دو کلاس قرار گیرد.

خواننده‌ای که علاقه‌مند است دانش بیشتری در مورد توانایی یادگیری و درک رابطه نزدیک بین خطای مدل، اندازه مجموعه آموزش، شکل معادله ریاضی مدل و زمان لازم برای ساخت مدل کسب کند، توصیه می‌گردد تا یادگیری CAP را مطالعه کند. نظریه یادگیری CAP^۱ ("احتمال تقریباً صحیح"^۲) یا به عبارتی یادگیری با احتمال تقریباً درست به ما کمک می‌کند تا دریابیم کجا و تحت چه شرایطی الگوریتم یادگیری، با احتمال خوبی، خروجی را درست دسته‌بندی می‌کند.



1 - PAC learning

2 - probably approximately correct

فرهنگ لغات تخصصی فصل ۱

Action	عمل
Accuracy	صحت
Bag of words	کوله‌ی کلمات
Classes	کلاس‌ها
Clustering	خوشه‌بندی
Dataset	مجموعه داده
Dimensionality reduction	کاهش بعد
Decision boundary	مرز تصمیم‌گیری
Expected average reward	متوسط پاداش مورد انتظار
Feature vector	بردار ویژگی
Labeled examples	نمونه‌های برچسب‌دار
Linear model	مدل خطی
Learning algorithm	الگوریتم یادگیری
Margin	حاشیه
one-shot decision	تصمیم‌گیری یک-شات
Outlier detection	تشخیص پرت
Policy	خط‌مشی
PAC learning	یادگیری PAC
Probably approximately correct	احتمال تقریباً صحیح
Reinforcement learning	یادگیری تقویتی
State	حالت
Reward	پاداش
Supervised learning	یادگیری نظارت‌شده
Semi-supervised learning	یادگیری نیمه‌نظارتی
Support Vector Machine	ماشین بردار پشتیبان
Training	آموزش

