

یادگیری ماشین عملی با Keras، Scikit-Learn و TensorFlow

مفاهیم، ابزارها و تکنیک‌هایی برای
ساخت سیستم‌های هوشمند

(یادگیری ماشین با پایتون)

مؤلف:

اورلین جرون

مترجمین:

دکتر جواد وحیدی – (عضو هیات علمی دانشگاه علم و صنعت ایران)

مهندس مزگان الیکائی آهاری



فن‌آوری نوین

سرشناسه	: گرون، اورلین
عنوان و نام پدیدآور	: یادگیری ماشین عملی با Scikit-learn و Keras و TensorFlow: مفاهیم، ابزارها و تکنیک‌هایی برای ساخت سیستم‌های هوشمند (یادگیری ماشین با پایتون) / مولف اورلین جرون؛ مترجمین جواد وحیدی، مزگان الیکانی آهاری.
مشخصات نشر	: بابل: فناوری نوین، ۱۴۰۰.
مشخصات ظاهری	: ۴۷۴ ص.: مصور، جدول، نمودار.
شابک	: 978-622-7393-61-3
وضعیت فهرست نویسی	: فیبا
یادداشت	: عنوان اصلی: Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow ۲nd ed, ۲: concepts, tools, and techniques to build intelligent systems, ۲۰۱۹.
عنوان دیگر	: یادگیری ماشین با پایتون.
موضوع	: تنسورفلو
موضوع	: پایتون (زبان برنامه‌نویسی کامپیوتر) TensorFlow پایتون (زبان برنامه‌نویسی کامپیوتر) (Python (Computer program language
شناسه افزوده	: یادگیری ماشینی
شناسه افزوده	: Machine learning
شناسه افزوده	: هوش مصنوعی
شناسه افزوده	: Artificial intelligence
رده بندی کنگره	: و وحیدی، جواد، ۱۳۶۸ - مترجم
رده بندی دیویی	: Vahidi, Javad
شماره کتابشناسی ملی	: الیکانی آهاری، مزگان، ۱۳۶۸ - مترجم
اطلاعات رکورد کتابشناسی	: ۷۶/۷۳۴A
	: ۰۰۵/۱۳۳
	: ۸۵۷۲۰۲۶
	: فیبا

www.novinpub.ir

تلفن: ۰۱۱-۳۲۲۵۶۶۸۷

بابل، کد پستی ۴۷۱۶۷-۷۳۴۴۸

فن آوری نوین

یادگیری ماشین عملی با Scikit-Learn و Keras و TensorFlow: مفاهیم، ابزارها و تکنیک‌هایی برای ساخت سیستم‌های هوشمند (یادگیری ماشین با پایتون)
ترجمه: جواد وحیدی، مزگان الیکانی آهاری.

نوبت چاپ: چاپ اول

سال چاپ: پاییز ۱۴۰۰

شمارگان: ۱۰۰۰

قیمت: ۲۰۵۰۰۰ تومان

نام چاپخانه و صحافی: دفتر فنی سورنا

شابک چاپی: 978-622-7393-61-3

نشانی ناشر: بابل، چهارراه نواب، کاظم‌بیگی، جنب مسجد منصور کاظم‌بیگی، طبقه اول

طراح جلد: مهندس الهام بیده

تهران، خ اردیبهشت، نبش وحید نظری، پلاک ۱۴۲ تلفن: ۰۱۱-۳۲۲۵۶۶۸۷

۱-۱. یادگیری ماشین چیست؟ ۱۵

۱-۲. چرا باید از یادگیری ماشینی استفاده کرد؟ ۱۶

۱-۳. انواع سیستم‌های یادگیری ماشینی ۱۹

۱-۳-۱. یادگیری با ناظر / بدون ناظر ۲۰

۱-۳-۲. یادگیری برخط و دسته‌ای ۲۷

۱-۳-۳. یادگیری مبتنی بر نمونه در مقابل مبتنی مدل ۳۰

۱-۴. چالش‌های اصلی یادگیری ماشینی ۳۶

۱-۴-۱. اثر بیخشی نامعقول داده‌ها ۳۶

۱-۴-۲. داده‌های آموزشی غیر معرف ۳۷

۱-۴-۳. داده‌های کم کیفیت ۳۹

۱-۴-۴. ویژگی‌های نامربوط ۳۹

۱-۴-۵. بیش برآزش داده‌های آموزش ۴۰

۱-۴-۶. کم برآزش داده‌های آموزش ۴۲

۱-۴-۷. عقب‌نشینی ۴۳

۱-۵. تست و اعتبار سنجی ۴۳

۱-۵-۱. تنظیم ابر پارامترها و انتخاب مدل ۴۳

۱-۵-۲. عدم تطابق داده‌ها ۴۴

۱-۶. تمرینات ۴۶

۲-۱-۱. کار با داده‌های واقعی ۴۷

۲-۱-۲. به تصویر بزرگ نگاه کنید ۴۸

۲-۱-۳. مسئله را قاب‌بندی کنید ۴۸

۲-۱-۴. یک معیار عملکرد را انتخاب کنید ۵۰

۲-۱-۵. مقروضات را بررسی کنید ۵۳

۲-۲. اطلاعات را دریافت کنید ۵۴

۲-۲-۱. ایجاد محیطی جداگانه ۵۴

۲-۲-۲. بارگیری اطلاعات ۵۷

۲-۲-۳. نگاهی گذرا به ساختار داده بیندازید ۵۸

۲-۲-۴. یک مجموعه تست ایجاد کنید ۶۲

۲-۳. برای به دست آوردن بیش، داده‌ها را کشف و تجسم کنید ۶۷

۲-۳-۱. مصورسازی داده‌های جغرافیایی ۶۷

۲-۳-۲. جستجو برای همبستگی‌ها ۷۰

۲-۳-۳. تجربه کردن با ترکیبات ویژگی ۷۳

۲-۳-۴. داده‌ها را برای الگوریتم‌های یادگیری ماشین آماده کنید ۷۴

۲-۴-۱. پاک‌سازی داده‌ها ۷۵

۲-۴-۲. جستجو برای همبستگی‌ه رسیدگی به متن و خصوصیات طبقه‌ای ۷۸

۲-۴-۳. مبدل‌های سفارشی ۸۰

۲-۴-۴. مقیاس‌گذاری ویژگی ۸۱

۲-۴-۵. تبدیل پایلین ۸۲

۲-۵. یک مدل را انتخاب و آموزش دهید ۸۴

۲-۵-۱. آموزش و ارزیابی روی مجموعه آموزش ۸۴

۲-۵-۲. ارزیابی بهتر با استفاده از اعتبارسنجی متقابل ۸۶

۲-۶. مدل خود را خوب تنظیم کنید ۸۸

۲-۶-۱. جستجوی شبکه‌ای ۸۸

۲-۶-۲. جستجوی تصادفی ۹۰

۲-۶-۳. روش‌های جمعی ۹۱

۲-۶-۴. بهترین مدل‌ها و خطاهای آن‌ها را تحلیل کنید ۹۱

۲-۶-۵. سیستم خود را در مجموعه آزمایش ارزیابی کنید ۹۲

۲-۶-۶. سیستم خود را راه‌اندازی، نظارت و نگهداری کنید ۹۳

۲-۷. تمرینات ۹۴

۳-۱. MNIST ۹۵

۳-۲. آموزش کلاس باینری ۹۷

۹۸.....	۳-۳. اندازه گیری عملکرد.....
۹۸.....	۳-۳-۱. اجرای اعتبار سنجی متقابل.....
۱۰۰.....	۳-۳-۲. ماتریس درهم ریختگی.....
۱۰۲.....	۳-۳-۳. دقت و بازیابی.....
۱۰۳.....	۳-۳-۴. مصالحه دقت / بازیابی.....
۱۰۷.....	۳-۳-۵. منحنی ROC.....
۱۱۰.....	۳-۴. طبقه بندی چند کلاسه.....
۱۱۳.....	۳-۵. تجزیه و تحلیل خطا.....
۱۱۷.....	۳-۶. طبقه بندی چند برجسته.....
۱۱۸.....	۳-۷. طبقه بندی چند خروجی.....
۱۱۹.....	۳-۸. تمرینات.....

فصل چهارم: مدل های یادگیری

۱۲۲.....	۴-۱. رگرسیون خطی.....
۱۲۳.....	۴-۱-۱. معادله نورمال.....
۱۲۶.....	۴-۱-۲. پیچیدگی محاسباتی.....
۱۲۶.....	۴-۲. گرادینان کاهشی.....
۱۳۰.....	۴-۲-۱. گرادینان کاهشی Batch.....
۱۳۱.....	۴-۲-۲. نرخ همگرایی.....
۱۳۱.....	۴-۲-۳. گرادینان کاهشی تصادفی.....
۱۳۲.....	۴-۲-۴. گرادینان کاهشی Mini-batch.....
۱۳۶.....	۴-۳. رگرسیون چند جمله ای.....
۱۳۸.....	۴-۴. منحنی های یادگیری.....
۱۴۲.....	۴-۵. تعادل بین بایاس و واریانس.....
۱۴۲.....	۴-۶. مدل های خطی منظم.....
۱۴۳.....	۴-۶-۱. رگرسیون ستینی.....
۱۴۴.....	۴-۶-۲. رگرسیون لاسو.....
۱۴۷.....	۴-۶-۳. نت الاستیک.....
۱۴۷.....	۴-۶-۴. توقف زودهنگام.....
۱۴۹.....	۴-۷. رگرسیون لجستیک.....
۱۴۹.....	۴-۷-۱. تخمین احتمالات.....
۱۵۰.....	۴-۷-۲. تابع هزینه و آموزش.....
۱۵۱.....	۴-۷-۳. مرزهای تصمیم گیری.....
۱۵۵.....	۴-۷-۴. آنتروپی متقاطع.....
۱۵۷.....	۴-۸. تمرینات.....

فصل پنجم: ماشین بردار پشتیبان

۱۵۹.....	۵-۱. طبقه بندی خطی SVM.....
۱۶۰.....	۵-۱-۱. طبقه بندی حاشیه نرم.....
۱۶۲.....	۵-۲. طبقه بندی غیر خطی SVM.....
۱۶۳.....	۵-۲-۱. هسته چند جمله ای.....
۱۶۴.....	۵-۲-۲. اضافه کردن ویژگی های شباهت.....
۱۶۵.....	۵-۲-۳. هسته تابع شعاعی بایه گوسین.....
۱۶۷.....	۵-۲-۴. پیچیدگی محاسباتی.....
۱۶۷.....	۵-۳. رگرسیون SVM.....
۱۶۹.....	۵-۴. در زیر Hood.....
۱۷۰.....	۵-۴-۱. تابع تصمیم گیری و پیش بینی ها.....
۱۷۱.....	۵-۴-۲. هدف آموزش.....
۱۷۲.....	۵-۴-۳. برنامه نویسی درجه دوم.....
۱۷۳.....	۵-۴-۴. مساله دوگان.....
۱۷۴.....	۵-۴-۵. SVM هسته دار شده.....
۱۷۵.....	۵-۴-۶. تئوری مرسر.....

۱۷۷	۴-۵. ائتلاف هینچ
۱۷۷	۵. تمرینات

فصل ششم: درخت تصمیم گیری

۱۷۹	۱-۷. آموزش و تجسم درخت تصمیم
۱۸۰	۲-۷. پیش‌بینی کردن
۱۸۲	۳-۷. تخمین احتمالات کلاس
۱۸۳	۴-۷. الگوریتم آموزش CART
۱۸۳	۵-۷. پیچیدگی محاسباتی
۱۸۴	۶-۷. ناخالصی جینی یا آنتروپی؟
۱۸۴	۷-۷. تنظیم ابر پارامترها
۱۸۶	۸-۷. رگرسیون
۱۸۸	۹-۷. بی‌ثباتی
۱۸۹	۱۰-۷. تمرینات

فصل هفتم: یادگیری جمعی و جنگل تصادفی

۱۹۱	۱-۷. طبقه‌بندی کننده رأی گیری
۱۹۵	۲-۷. بگینگ و Pasting
۱۹۶	۱-۲-۷. بگینگ و Pasting در Scikit-Learn
۱۹۷	۲-۲-۷. محاسبات Out-of-Bag
۱۹۸	۳-۷. بچ‌های تصادفی و زیر فضاهای تصادفی
۱۹۹	۴-۷. جنگل‌های تصادفی
۲۰۰	۱-۴-۷. درختان اضافی
۲۰۰	۲-۴-۷. اهمیت ویژگی
۲۰۲	۵-۷. تقویت
۲۰۲	۱-۵-۷. آداپوست
۲۰۵	۲-۵-۷. گرادینت تقویتی
۲۱۰	۶-۷. پشته‌سازی

فصل هشتم: کاهش ابعاد

۲۱۵	۱-۸. شفت بعد چندی
۲۱۶	۲-۸. رویکردهای اصلی کاهش ابعاد
۲۱۷	۱-۲-۸. تکاشت
۲۱۸	۲-۲-۸. یادگیری منیفولد
۲۲۰	۳-۸. تحلیل مؤلفه‌های اصلی
۲۲۰	۱-۳-۸. Preserving واریانس
۲۲۱	۲-۳-۸. مؤلفه‌های اصلی
۲۲۲	۳-۳-۸. تکاشت به بعد پایین‌تر تا d بعد
۲۲۳	۴-۳-۸. استفاده از Scikit-Learn
۲۲۳	۵-۳-۸. نرخ واریانس توضیح داده‌شده
۲۲۳	۶-۳-۸. انتخاب تعداد مناسب ابعاد
۲۲۴	۷-۳-۸. PCA برای فشرده‌سازی
۲۲۵	۸-۳-۸. PCA تصادفی
۲۲۶	۹-۳-۸. PCA افزایشی
۲۲۷	۴-۸. آتالیز مؤلفه‌های اصلی مبتنی بر هسته
۲۲۷	۱-۴-۸. انتخاب و تنظیم ابر پارامترهای هسته
۲۳۰	۵-۸. جاسازی خطی محلی
۲۳۲	۶-۸. سایر تکنیک‌های کاهش ابعاد
۲۳۳	۷-۸. تمرینات

فصل نهم: تکنیک‌های یادگیری بدون نظارت

۲۳۶	۱-۹. خوشه‌بندی
۲۳۸	۱-۱-۹. K-Means
۲۴۷	۲-۱-۹. محدودیت‌های K-Means

۲۴۸	۹-۱-۳. استفاده از خوشه‌بندی برای بخش‌بندی تصویر
۲۵۰	۹-۱-۴. استفاده از خوشه‌بندی برای پیش‌پردازش
۲۵۱	۹-۱-۵. استفاده از خوشه‌بندی برای یادگیری نیمه نظارت‌شده
۲۵۵	۹-۱-۶. DBSCAN
۲۵۷	۹-۲. مخلوط‌های گاوسی
۲۶۳	۹-۲-۱. تشخیص ناهنجاری با استفاده از مخلوط‌های گاوسی
۲۶۵	۹-۲-۲. انتخاب تعداد خوشه‌ها
۲۶۸	۹-۲-۳. مدل‌های مخلوط گاوسی یزی
۲۷۱	۹-۲-۴. سایر الگوریتم‌های تشخیص ناهنجاری و تازگی

فصل دهم: مقدمه‌ای بر شبکه‌های عصبی مصنوعی با Keras

۲۷۳	۱۰-۱. از سلول‌های عصبی بیولوژیکی گرفته تا مصنوعی
۲۷۴	۱۰-۱-۱. نورون‌های بیولوژیکی
۲۷۵	۱۰-۱-۲. محاسبات منطقی با سلول‌های عصبی
۲۷۶	۱۰-۱-۳. پرسپترون
۲۸۱	۱۰-۱-۴. پرسپترون چند لایه و پس‌انتشار
۲۸۳	۱۰-۱-۵. MLPs
۲۸۴	۱۰-۱-۶. طبقه‌بندی MLPها
۲۸۶	۱۰-۲. پیاده‌سازی MLPها با Keras
۲۸۶	۱۰-۲-۱. نصب TensorFlow ۲
۲۸۷	۱۰-۲-۲. ساخت یک طبقه‌بندی تصویر با استفاده از API پی‌درپی (متوالی)
۲۸۹	۱۰-۲-۳. Sequential API
۲۹۶	۱۰-۲-۴. ساخت MLP با پرسپترون با استفاده از API متوالی
۲۹۷	۱۰-۲-۵. ساخت مدل‌های پیچیده با استفاده از API عملکردی
۳۰۲	۱۰-۲-۶. ساخت مدل‌های پویا با استفاده از Subclassing API
۳۰۳	۱۰-۲-۷. ذخیره و بازیابی یک مدل
۳۰۴	۱۰-۲-۸. استفاده از بازخوانی
۳۰۶	۱۰-۲-۹. TensorBoard
۳۰۸	۱۰-۳. تنظیم خوب ابر پارامترهای شبکه عصبی
۳۱۲	۱۰-۳-۱. تعداد لایه‌های پنهان
۳۱۳	۱۰-۳-۲. تعداد نورون‌ها در هر لایه پنهان
۳۱۴	۱۰-۳-۳. نرخ یادگیری، اندازه دسته و سایر پارامترهای دیگر
۳۱۶	۱۰-۴. تمرینات

فصل یازدهم: آموزش شبکه‌های عصبی عمیق

۳۱۹	۱۱-۱. مشکل محوشدگی/انفجار گرادینان
۳۲۱	۱۱-۱-۱. مقداردهی اولیه HE و Glorot
۳۲۲	۱۱-۱-۲. توابع Nonsaturating Activation
۳۲۶	۱۱-۱-۳. نرمال‌سازی دسته‌ای
۳۳۱	۱۱-۱-۴. برش گرادینان
۳۳۲	۱۱-۲. استفاده مجدد از لایه‌های پیش آموزش دیده
۳۳۳	۱۱-۲-۱. یادگیری انتقال با Keras
۳۳۵	۱۱-۲-۲. پیش آموزش بدون نظارت
۳۳۶	۱۱-۲-۳. پیش آموزش روی یک وظیفه کمکی
۳۳۷	۱۱-۳. بهینه‌سازهای سریع تر
۳۳۷	۱۱-۳-۱. بهینه‌سازی ممتوم
۳۳۸	۱۱-۳-۲. گرادینان شتاب‌یافتهی نتروف
۳۳۹	۱۱-۳-۳. AdaGrad
۳۴۱	۱۱-۳-۴. RMSProp
۳۴۱	۱۱-۳-۵. بهینه‌سازی Adam و Nadam
۳۴۳	۱۱-۳-۶. آموزش مدل‌های اسپارس
۳۴۴	۱۱-۳-۷. زمان‌بندی نرخ یادگیری

۳۴۸	۱۱-۴. جلوگیری از بیش برآزش هنگام تنظیم پارامتر
۳۴۸	۱۱-۴-۱. تنظیم ℓ_1 و ℓ_2
۳۴۹	۱۱-۴-۲. Dropout
۳۵۱	۱۱-۴-۳. Dropout مونت کارلو
۳۵۳	۱۱-۴-۴. تنظیم کننده Max-Norm
۳۵۴	۱۱-۵. رهنمودهای خلاصه و عملی
۳۵۵	۱۱-۶. تمرینات

فصل دوازدهم: مدل‌های سفارشی و آموزش با TensorFlow

۳۵۹	۱۲-۱. گشت‌وگذار سریع در TensorFlow
۳۶۲	۱۲-۲-۱. NumPy استفاده از TensorFlow مانند NumPy
۳۶۲	۱۲-۲-۱. تصورها و عملیات
۳۶۴	۱۲-۲-۲. حسگرها و NumPy
۳۶۴	۱۲-۲-۳. تبدیل نوع
۳۶۵	۱۲-۲-۴. متغیرها
۳۶۵	۱۲-۲-۵. سایر ساختارهای داده
۳۶۷	۱۲-۲. سفارشی کردن مدل‌ها و الگوریتم‌های آموزش
۳۶۷	۱۲-۳-۱. توابع از دست دادن سفارشی
۳۶۸	۱۲-۳-۲. ذخیره و بارگیری مدل‌هایی که دارای اجزای سفارشی هستند
۳۷۰	۱۲-۳-۳. توابع فعال‌سازی سفارشی، مقدار اولیه دهنده، تنظیم‌کننده‌های پارامتر و محدودیت‌ها
۳۷۱	۱۲-۳-۴. معیارهای سفارشی
۳۷۴	۱۲-۳-۵. لایه‌های سفارشی
۳۷۷	۱۲-۳-۶. مدل‌های سفارشی
۳۷۹	۱۲-۳-۷. معیار و زیان بر اساس مدل داخلی
۳۸۱	۱۲-۳-۸. محاسبه گرادیان‌ها با استفاده از Autodiff
۳۸۵	۱۲-۳-۹. حلقه‌های آموزشی سفارشی
۳۸۸	۱۲-۴. توابع و نمودارهای TensorFlow
۳۸۹	۱۲-۴-۱. رونویسی و ردیابی
۳۹۱	۱۲-۴-۲. قوانین تابع TF

فصل سیزدهم: بارگیری و بیش بردارش داده‌ها با TensorFlow

۳۹۴	۱۳-۱. API داده
۳۹۵	۱۳-۱-۱. تبدیل زنجیره‌ای
۳۹۶	۱۳-۱-۲. بر زدن داده
۳۹۸	۱۳-۱-۳. جایگذاری خطوط از چندین فایل
۳۹۹	۱۳-۱-۴. بیش پردازش داده‌ها
۴۰۰	۱۳-۱-۵. همه چیز را کنار هم قرار دهید
۴۰۱	۱۳-۱-۶. پیش‌واکشی
۴۰۲	۱۳-۱-۷. استفاده از مجموعه داده با tf.keras
۴۰۴	۱۳-۲. قالب TFRecord
۴۰۵	۱۳-۲-۱. فایل‌های فشرده TFRecord
۴۰۵	۱۳-۲-۲. مقدمه ای مختصر بر Protocol Buffers
۴۰۷	۱۳-۲-۳. TensorFlow Protobufs
۴۰۸	۱۳-۲-۴. بارگیری و تجزیه نمونه‌ها
۴۱۰	۱۳-۲-۵. مدیریت لیست‌ها با استفاده از SequenceExample protobuf
۴۱۱	۱۳-۳. ویژگی‌های API
۴۱۳	۱۳-۳-۱. ویژگی‌های طبقه بندی شده
۴۱۳	۱۳-۳-۲. ویژگی‌های طبقه بندی متقاطع
۴۱۴	۱۳-۳-۳. رمزگذاری ویژگی‌های طبقه بندی با استفاده از بردارهای One-Hot
۴۱۴	۱۳-۳-۴. رمزگذاری ویژگی‌های دسته بندی با استفاده از جاسازی‌ها
۴۱۵	۱۳-۳-۵. جاسازی کلمه
۴۱۷	۱۳-۳-۶. استفاده از ستون‌های ویژگی برای تجزیه

مقدمه

سونامی یادگیری ماشین

در سال ۲۰۰۶، جفری هینتون و همکارانش^۱ مقاله‌ای منتشر کردند که نشان می‌داد چگونه آموزش یک شبکه عصبی عمیق قادر به تشخیص ارقام دست‌نویس با دقت پیشرفته (> ۹۸٪) را است. آن‌ها این تکنیک را یادگیری عمیق نام‌گذاری کردند. در آن زمان آموزش شبکه عصبی عمیق غیرممکن تلقی می‌شد و اکثر محققان از دهه ۱۹۹۰ این ایده را کنار گذاشته بودند. نتایج این مقاله علاقه جامعه علمی را در مورد یادگیری عمیق دوباره زنده کرد و نشان داد که یادگیری عمیق نه تنها امکان‌پذیر است، بلکه می‌تواند دستاوردهای حیرت‌انگیزی داشته باشد که هیچ تکنیک یادگیری ماشین (ML) نمی‌تواند با آن‌ها مطابقت داشته باشد. این اشتیاق به‌زودی به بسیاری از زمینه‌های دیگر یادگیری ماشین گسترش یافت.

۱۰ یادگیری ماشین صنعت را تسخیر کرده است و اکنون در قلب بسیاری از جادوهای محصولات با تکنولوژی قرار دارد. یادگیری ماشین نتایج جستجوی وب را رتبه‌بندی کرده، تشخیص گفتار تلفن هوشمند را تقویت و به شما ویدیو توصیه می‌کند.

یادگیری ماشین در پروژه‌های شما

آیا به یادگیری ماشین علاقه دارید و دوست دارید به این داستان بپیوندید؟

شاید شرکت شما دارای تعداد زیادی داده (گزارشات کاربر، داده‌های مالی، داده‌های تولید، داده‌های حسگر ماشین، آمار خط تلفن، گزارشات منابع انسانی و غیره) باشد و به احتمال زیاد اگر بدانید کجا باید نگاه کنید برخی مشکلات حل می‌شوند:

- ❖ مشتریان را تقسیم‌بندی کرده و بهترین استراتژی بازاریابی را برای هر گروه پیدا کنید.
- ❖ محصولات را برای هر مشتری بر اساس خرید مشتریان مشابه توصیه کنید.
- ❖ تشخیص دهید کدام تراکنش‌ها احتمالاً تقلبی هستند.
- ❖ پیش‌بینی درآمد سال آینده.
- ❖ و غیره

به هر دلیلی تصمیم گرفته‌اید یادگیری ماشین را بیاموزید و آن را در پروژه‌های خود پیاده‌سازی کنید.

هدف و رویکرد

این کتاب فرض می‌کند که شما تقریباً هیچ چیز درباره یادگیری ماشین نمی‌دانید. هدف آن ارائه مفاهیم، شهود و ابزارهایی است که برای اجرای برنامه‌های قابل یادگیری از داده‌ها نیاز دارید. ما تکنیک‌های زیادی را پوشش می‌دهیم، از ساده‌ترین و متداول‌ترین (مانند رگرسیون خطی) تا برخی از تکنیک‌های یادگیری عمیق که به‌طور منظم در مسابقات برنده می‌شوند.

به‌جای اجرای نسخه‌های خودمان از هر الگوریتم، از چارچوب‌های آماده تولید پایتون استفاده می‌کنیم:

- ❖ استفاده از Scikit-Learn بسیار آسان است، با این حال بسیاری از الگوریتم‌های یادگیری ماشین را

^۱ <http://www.cs.toronto.edu/~hinton/>

به طور کارآمد پیاده سازی می کند، بنابراین یک نقطه ورود عالی برای یادگیری ماشین است.

- ❖ TensorFlow یک کتابخانه پیچیده تر برای محاسبه عددی توزیع شده است. این امکان را فراهم می کند که شبکه های عصبی بسیار بزرگ را با تقسیم محاسبات در صدها سرور چند GPU به طور کارآمد آموزش داده و اجرا کنیم. TensorFlow در Google ایجاد شده و از بسیاری از برنامه های بزرگ یادگیری ماشین را پشتیبانی می کند. این منبع در نوامبر ۲۰۱۵ باز شد.
- ❖ Keras یک API یادگیری عمیق سطح بالا است که آموزش و اجرای شبکه های عصبی را بسیار ساده می کند. می تواند روی TensorFlow، Theano یا Micro soft Cognitive Toolkit (که قبلاً به عنوان CNTK شناخته می شد) اجرا شود. TensorFlow با پیاده سازی خود از این API، به نام tf.keras پشتیبانی می کند که از برخی ویژگی های پیشرفته TensorFlow (به عنوان مثال برای بارگذاری کارآمد داده ها) پشتیبانی می کند.

این کتاب از یک رویکرد عملی استفاده می کند و درک شهودی از یادگیری ماشین را از طریق مثال های عملی و کمی تئوری افزایش می دهد. توصیه می کنیم نمونه های کد موجود در اینترنت را به عنوان نوتبوک Jupyter در <https://github.com/ageron/handson-ml2> بررسی کنید.

پیش نیازها

این کتاب فرض می کند که شما برخی از تجربه برنامه نویسی پایتون را دارید و با کتابخانه های اصلی علمی پایتون، به ویژه NumPy، Pandas و Matplotlib آشنا هستید.

همچنین، اگر به موارد زیر توجه دارید، باید از ریاضیات سطح دانشگاه نیز درک خوبی داشته باشید (حساب، جبر خطی، احتمالات و آمار).

اگر هنوز پایتون را نمی شناسید، <http://learnpython.org/> مکان عالی برای شروع است. آموزش رسمی در python.org نیز بسیار خوب است.

اگر هرگز از Jupyter استفاده نکرده اید، فصل ۲ شما را در مورد نصب و اصول اولیه راهنمایی می کند. اگر با کتابخانه های علمی پایتون آشنایی ندارید، کتاب های ارائه شده include Jupyter note شامل چند آموزش است. همچنین یک آموزش ریاضی سریع برای جبر خطی وجود دارد.

نقشه راه

این کتاب در دو بخش تنظیم شده است. **قسمت اول، مبانی یادگیری ماشین، موضوعات زیر را پوشش**

می دهد:

- ❖ یادگیری ماشین چیست؟ سعی می کند چه مشکلاتی را حل کند؟ مقوله های اصلی و مفاهیم اساسی سیستم های یادگیری ماشین چیست؟
- ❖ مراحل اصلی در یک پروژه معمول یادگیری ماشین.
- ❖ یادگیری با متناسب سازی مدل با داده ها.