

به نام خدا

# یادگیری ماشین با Python

نویسنده: جیسن براونلی

مترجم:

مهندس رامین مولاناپور

(دانشجوی دکترای مهندسی فناوری اطلاعات)

## یادگیری ماشین با Python

مترجم: مهندس رامین مولاناپور

ناشر: انتشارات آتی‌نگر

ناشر همکار: انتشارات وینا

چاپ اول، ۱۳۹۶

شمارگان: ۱۰۰۰ نسخه

قیمت: ۲۵۰,۰۰۰ ریال

شابک: ۹۷۸-۶۰۰-۷۶۳۱-۸۷-۴

ISBN: 978-600-7631-87-4

حق چاپ برای انتشارات آتی‌نگر محفوظ است.

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

نمابر: ۶۶۵۶۵۳۳۷

تلفن: ۸-۶۶۵۶۵۳۳۶



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

سرشناسه: براونلی، جیسن، Brownlee Jason

یادگیری ماشین با Python / نویسنده: جیسن براونلی؛ مترجم: رامین مولاناپور.

تهران: آتی‌نگر؛ وینا، ۱۳۹۶

۲۴۸ ص.: مصور، جدول.

ISBN: 978-600-7631-87-4

فیبا.

یادداشت: عنوان اصلی: Machine learning mastery with Python : understand your data, create accurate models and work projects end-to-end, c 2016

موضوع: پیتون (زبان برنامه‌نویسی کامپیوتر) - Python (Computer program language)

موضوع: شبکه‌های عصبی (کامپیوتر) - Neural networks (Computer science)

شناسه افزوده: مولاناپور، رامین، ۱۳۵۲ - مترجم

QAY۶/۷۳/پ۹۵ب۴ ۱۳۹۶

۰۰۵/۱۳۳

۴۹۶۲۸۰۴

رده‌بندی کنگره

رده‌بندی دیویی

شماره کتابشناسی ملی

# فهرست مطالب

۱۱	پیشگفتار
۱۳	بخش اول: مقدمه
۱۵	فصل اول: خوش آمدید
۱۵	۱-۱ آموختن یادگیری ماشین Python به طور نادرست
۱۶	۱-۲ یادگیری ماشین در Python
۲۰	۱-۳ این کتاب چه چیزی را ارائه نمی دهد
۲۱	۱-۴ خلاصه
۲۳	بخش دوم: دروس
۲۵	فصل دوم: اکوسیستم Python برای یادگیری ماشین
۲۵	۲-۱ Python
۲۶	۲-۲ SciPy
۲۷	۲-۳ Scikit-Learn
۲۷	۲-۴ نصب اکوسیستم Python
۳۰	۲-۵ خلاصه
۳۳	فصل سوم: دوره های فشرده برای Python و SciPy
۳۳	۳-۱ دوره فشرده Python
۴۰	۳-۲ دوره فشرده NumPy
۴۲	۳-۳ دوره فشرده Matplotlib
۴۴	۳-۴ دوره فشرده Pandas
۴۷	۳-۵ خلاصه

## فصل چهارم: چگونگی بارگذاری داده‌های یادگیری ماشین

۴۹

- ۴-۱ ملاحظات زمان بارگذاری داده‌های CSV ..... ۴۹
- ۴-۲ مجموعه داده Pima Indians ..... ۵۰
- ۴-۳ بارگذاری فایل‌های CSV با کتابخانه استاندارد Python ..... ۵۱
- ۴-۴ بارگذاری فایل‌های CSV با NumPy ..... ۵۲
- ۴-۵ بارگذاری فایل‌های CSV با Pandas ..... ۵۳
- ۴-۶ خلاصه ..... ۵۴

## فصل پنجم: شناخت داده‌ها با آمار توصیفی

۵۷

- ۵-۱ ورنانداز داده‌ها ..... ۵۷
- ۵-۲ ابعاد داده‌ها ..... ۵۹
- ۵-۳ نوع داده هر ستون ..... ۵۹
- ۵-۴ آمار توصیفی ..... ۶۰
- ۵-۵ توزیع کلاس (فقط دسته‌بندی) ..... ۶۱
- ۵-۶ همبستگی‌های بیین صفات ..... ۶۲
- ۵-۷ چوله توزیع‌های یک‌متغیره ..... ۶۳
- ۵-۸ نکاتی که باید به خاطر سپرد ..... ۶۴
- ۵-۹ خلاصه ..... ۶۵

## فصل ششم: شناخت داده‌ها با مصورسازی

۶۷

- ۶-۱ ترسیمات یک‌متغیره ..... ۶۷
- ۶-۲ نمودارهای چند متغیره ..... ۷۱
- ۶-۳ خلاصه ..... ۷۵

## فصل هفتم: آماده‌سازی داده‌ها برای یادگیری ماشین

۷۷

- ۷-۱ نیاز به پیش‌پردازش داده‌ها ..... ۷۷
- ۷-۲ تبدیلات داده‌ای ..... ۷۸
- ۷-۳ تغییر مقیاس داده‌ها ..... ۷۸
- ۷-۴ استاندارد کردن داده‌ها ..... ۸۰

۸۱	۷-۵ نرمال سازی داده‌ها
۸۲	۷-۶ دودویی کردن داده‌ها
۸۳	۷-۷ خلاصه

## ۸۵ فصل هشتم: انتخاب ویژگی برای یادگیری ماشین

۸۵	۸-۱ انتخاب ویژگی
۸۶	۸-۲ انتخاب یک متغیره
۸۷	۸-۳ حذف ویژگی بازگشتی
۸۸	۸-۴ تحلیل مؤلفه اصلی
۸۹	۸-۵ اهمیت ویژگی
۹۰	۸-۶ خلاصه

## ۹۳ فصل نهم: ارزیابی عملکرد الگوریتم‌های یادگیری ماشین با باز نمونه‌گیری

۹۳	۹-۱ ارزیابی الگوریتم‌های یادگیری ماشین
۹۴	۹-۲ تقسیم‌بندی به مجموعه‌های آموزشی و آزمایشی
۹۶	۹-۳ اعتبارسنجی متقابل K تایی
۹۷	۹-۴ اعتبارسنجی متقابل Leave One Out
۹۸	۹-۵ تقسیم‌بندی‌های آزمایشی - آموزشی تصادفی تکراری
۹۹	۹-۶ چه تکنیک‌هایی را چه هنگام استفاده کنیم
۹۹	۹-۷ خلاصه

## ۱۰۱ فصل دهم: معیارهای عملکرد الگوریتم یادگیری ماشین

۱۰۱	۱۰-۱ معیارهای ارزیابی الگوریتم
۱۰۲	۱۰-۲ معیارهای دسته‌بندی
۱۰۷	۱۰-۳ معیارهای رگرسیون
۱۱۱	۱۰-۴ خلاصه

## ۱۱۳ فصل یازدهم: الگوریتم‌های دسته‌بندی نمونه‌سنجی

۱۱۳	۱۱-۱ نمونه‌سنجی الگوریتم
-----	--------------------------

۱۱۴.....	۱۱-۲	مروری بر الگوریتم‌ها
۱۱۵.....	۱۱-۳	الگوریتم‌های یادگیری ماشین خطی
۱۱۶.....	۱۱-۴	الگوریتم‌های یادگیری ماشین غیرخطی
۱۲۰.....	۱۱-۵	خلاصه

## فصل دوازدهم: نمونه‌سنجی الگوریتم‌های رگرسیون

۱۲۱.....	۱۲-۱	مروری بر الگوریتم‌ها
۱۲۲.....	۱۲-۲	الگوریتم‌های یادگیری ماشین خطی
۱۲۶.....	۱۲-۳	الگوریتم‌های یادگیری ماشین غیرخطی
۱۲۹.....	۱۲-۴	خلاصه

## فصل سیزدهم: مقایسه الگوریتم‌های یادگیری ماشین

۱۳۱.....	۱۳-۱	انتخاب بهترین مدل یادگیری ماشین
۱۳۲.....	۱۳-۲	مقایسه منسجم الگوریتم‌های یادگیری ماشین
۱۳۵.....	۱۳-۳	خلاصه

## فصل چهاردهم: خودکار کردن گردش کارهای یادگیری ماشین با Pipelines

۱۳۷.....	۱۴-۱	خودکار کردن گردش کارهای یادگیری ماشین
۱۳۸.....	۱۴-۲	آماده‌سازی داده‌ها و مدل‌سازی مجرای ارتباطی
۱۳۹.....	۱۴-۳	استخراج ویژگی و مجرای مدل‌سازی
۱۴۱.....	۱۴-۴	خلاصه

## فصل پانزدهم: بهبود عملکرد با مجموعه‌ها

۱۴۳.....	۱۵-۱	مدل‌های ترکیبی در پیش‌بینی‌های کلیت
۱۴۴.....	۱۵-۲	الگوریتم‌های کیسه کردن
۱۴۷.....	۱۵-۳	الگوریتم‌های تقویت
۱۴۹.....	۱۵-۴	کلیت رأی‌دهی
۱۵۱.....	۱۵-۵	خلاصه

فصل شانزدهم: بهبود عملکرد با میزان‌سازی الگوریتم ۱۵۳

- ۱۵۳..... ۱۶-۱ پارامترهای الگوریتم یادگیری ماشین
- ۱۵۴..... ۱۶-۲ میزان‌سازی پارامتر جستجوی مشبک
- ۱۵۵..... ۱۶-۳ میزان‌سازی پارامتر جستجوی تصادفی
- ۱۵۶..... ۱۶-۴ خلاصه

فصل هفدهم: ذخیره‌سازی و بارگذاری مدل‌های یادگیری ماشین ۱۵۷

- ۱۵۷..... ۱۷-۱ نهایی کردن مدل با Pickle
- ۱۵۹..... ۱۷-۲ نهایی کردن مدل با Joblib
- ۱۶۰..... ۱۷-۳ نکاتی برای نهایی کردن مدل
- ۱۶۱..... ۱۷-۴ خلاصه

بخش سوم: پروژه‌ها ۱۶۳

فصل هجدهم: قالب پروژه مدل‌سازی پیش‌بینانه ۱۶۵

- ۱۶۵..... ۱۸-۱ یادگیری ماشین عملی با پروژه‌ها
- ۱۶۶..... ۱۸-۲ قابل پروژه یادگیری ماشین در Python
- ۱۶۸..... ۱۸-۳ مراحل قالب پروژه یادگیری ماشین
- ۱۷۰..... ۱۸-۴ نکاتی برای استفاده خوب از قالب
- ۱۷۱..... ۱۸-۵ خلاصه

فصل نوزدهم: اولین پروژه یادگیری ماشین در Python به صورت گام به گام ۱۷۳

- ۱۷۳..... ۱۹-۱ سلام به دنیای یادگیری ماشین
- ۱۷۴..... ۱۹-۲ بارگذاری داده‌ها
- ۱۷۵..... ۱۹-۳ خلاصه‌سازی مجموعه داده
- ۱۷۸..... ۱۹-۴ مصورسازی داده‌ها
- ۱۸۱..... ۱۹-۵ ارزیابی چند الگوریتم
- ۱۸۵..... ۱۹-۶ انجام پیش‌بینی‌ها
- ۱۸۶..... ۱۹-۷ خلاصه

## فصل بیستم: پروژه مطالعه موردی یادگیری ماشین رگرسیون

۱۸۷

- ۱۸۷-۱ تعریف مسأله ..... ۱۸۷
- ۱۸۸-۲ بارگذاری مجموعه داده ..... ۱۸۸
- ۱۸۹-۳ تحلیل داده‌ها ..... ۱۸۹
- ۱۹۳-۴ مصورسازی‌های داده ..... ۱۹۳
- ۱۹۹-۵ مجموعه داده اعتبارسنجی ..... ۱۹۹
- ۱۹۹-۶ ارزیابی الگوریتم‌ها: خط مبنا ..... ۱۹۹
- ۲۰۱-۷ ارزیابی الگوریتم‌ها: استانداردسازی ..... ۲۰۱
- ۲۰۴-۸ بهبود نتایج با میزان‌سازی ..... ۲۰۴
- ۲۰۶-۹ روش‌های کلیت ..... ۲۰۶
- ۲۰۸-۱۰ میزان‌سازی روش‌های کلیت ..... ۲۰۸
- ۲۱۰-۱۱ نهایی کردن مدل ..... ۲۱۰
- ۲۱۰-۱۲ خلاصه ..... ۲۱۰

## فصل بیست و یکم: پروژه مطالعه موردی یادگیری ماشین دسته‌بندی دودویی

۲۱۳

- ۲۱۳-۱ تعریف مسأله ..... ۲۱۳
- ۲۱۴-۲ بارگذاری مجموعه داده ..... ۲۱۴
- ۲۱۵-۳ تحلیل داده‌ها ..... ۲۱۵
- ۲۲۲-۴ مجموعه داده اعتبارسنجی ..... ۲۲۲
- ۲۲۳-۵ ارزیابی الگوریتم‌ها: خط مبنا ..... ۲۲۳
- ۲۲۵-۶ ارزیابی الگوریتم‌ها: داده‌های استاندارد شده ..... ۲۲۵
- ۲۲۸-۷ میزان‌سازی الگوریتم ..... ۲۲۸
- ۲۳۲-۸ روش‌های کلیت ..... ۲۳۲
- ۲۳۴-۹ نهایی کردن مدل ..... ۲۳۴
- ۲۳۵-۱۰ خلاصه ..... ۲۳۵

## فصل بیست و دوم: پروژه‌های مدل‌سازی پیش‌بینانه بیشتر

۲۳۷

- ۲۳۷-۱ ساخت و نگهداری دستورالعمل‌ها ..... ۲۳۷
- ۲۳۸-۲ پروژه‌های کوچک روی مجموعه داده‌های کوچک ..... ۲۳۸
- ۲۳۸-۳ یادگیری ماشین رقابتی ..... ۲۳۸



۲۳۹..... خلاصه ۲۲-۴

۲۴۱ بخش چهارم: نتایج

---

۲۴۳ فصل بیست و سوم: چقدر به شما می آید

---

۲۴۵ فصل بیست و چهارم: گرفتن راهنمایی بیشتر

---

۲۴۵..... ۲۴-۱ توصیه کلی

۲۴۶..... ۲۴-۲ راهنمای Python

۲۴۶..... ۲۴-۳ راهنمای با NumPy و SciPy

۲۴۶..... ۲۴-۴ راهنمای Matplotlib

۲۴۷..... ۲۴-۵ راهنمای Pandas

۲۴۷..... ۲۴-۶ راهنمای Scikit-Learn