

# 机器学习实战：基于Scikit-Learn和TensorFlow



[机器学习实战：基于Scikit-Learn和TensorFlow\\_下载链接1](#)

著者:Aurélien Géron

出版者:机械工业出版社

出版时间:2018-8

装帧:平装

isbn:9787111603023

本书主要分为两个部分。第一部分为第1章到第8章，涵盖机器学习的基础理论知识和基本算法——从线性回归到随机森林等，帮助读者掌握Scikit-Learn的常用方法；第二部分为第9章到第16章，探讨深度学习和常用框架TensorFlow，一步一个脚印地带领读者使

用TensorFlow搭建和训练深度神经网络，以及卷积神经网络。

作者介绍:

Aurelien Geron

是机器学习方面的顾问。他曾是Google软件工程师，在2013年到2016年主导了YouTube视频分类工程。2002年和2012年，他还是Wifirst公司（一家法国的无线ISP）的创始人和首席技术官，2001年是Ployconseil公司（现在管理电动汽车共享服务Autolib）的创始人和首席技术官。

目录: 前言1

第一部分 机器学习基础

第1章 机器学习概览11

什么是机器学习12

为什么要使用机器学习12

机器学习系统的种类15

监督式/无监督式学习16

批量学习和在线学习21

基于实例与基于模型的学习24

机器学习的主要挑战29

训练数据的数量不足29

训练数据不具代表性30

质量差的数据32

无关特征32

训练数据过度拟合33

训练数据拟合不足34

退后一步35

测试与验证35

练习37

第2章 端到端的机器学习项目39

使用真实数据39

观察大局40

框架问题41

选择性能指标42

检查假设45

获取数据45

创建工作区45

下载数据48

快速查看数据结构49

创建测试集52

从数据探索和可视化中获得洞见56

将地理数据可视化57

寻找相关性59

试验不同属性的组合61

机器学习算法的数据准备62

数据清理63

处理文本和分类属性65

自定义转换器67

特征缩放68

转换流水线68

选择和训练模型70

培训和评估训练集70

使用交叉验证来更好地进行评估72

微调模型74

网格搜索74

随机搜索76

集成方法76

分析最佳模型及其错误76

通过测试集评估系统77

启动、监控和维护系统78

试试看79

练习79

第3章 分类80

MNIST80

训练一个二元分类器82

性能考核83

使用交叉验证测量精度83

混淆矩阵84

精度和召回率86

精度/召回率权衡87

ROC曲线90

多类别分类器93

错误分析95

多标签分类98

多输出分类99

练习100

第4章 训练模型102

线性回归103

标准方程104

计算复杂度106

梯度下降107

批量梯度下降110

随机梯度下降112

小批量梯度下降114

多项式回归115

学习曲线117

正则线性模型121

岭回归121

套索回归123

弹性网络125

早期停止法126

逻辑回归127

概率估算127

训练和成本函数128

决策边界129

Softmax回归131

练习134

第5章 支持向量机136

线性SVM分类136

软间隔分类137

非线性SVM分类139

多项式核140

添加相似特征141

高斯RBF核函数142

计算复杂度143

SVM回归144

工作原理145  
决策函数和预测146  
训练目标146  
二次规划148  
对偶问题149  
核化SVM149  
在线SVM151  
练习152  
第6章 决策树154  
决策树训练和可视化154  
做出预测155  
估算类别概率157  
CART训练算法158  
计算复杂度158  
基尼不纯度还是信息熵159  
正则化超参数159  
回归161  
不稳定性162  
练习163  
第7章 集成学习和随机森林165  
投票分类器165  
bagging和pasting168  
Scikit-Learn的bagging和pasting169  
包外评估170  
Random Patches和随机子空间171  
随机森林172  
极端随机树173  
特征重要性173  
提升法174  
AdaBoost175  
梯度提升177  
堆叠法181  
练习184  
第8章 降维185  
维度的诅咒186  
数据降维的主要方法187  
投影187  
流形学习189  
PCA190  
保留差异性190  
主成分191  
低维度投影192  
使用Scikit-Learn192  
方差解释率193  
选择正确数量的维度193  
PCA压缩194  
增量PCA195  
随机PCA195  
核主成分分析196  
选择核函数和调整超参数197  
局部线性嵌入199  
其他降维技巧200  
练习201  
第二部分 神经网络和深度学习

- 第9章 运行TensorFlow205
  - 安装207
  - 创建一个计算图并在会话中执行208
  - 管理图209
  - 节点值的生命周期210
  - TensorFlow中的线性回归211
  - 实现梯度下降211
  - 手工计算梯度212
  - 使用自动微分212
  - 使用优化器214
  - 给训练算法提供数据214
  - 保存和恢复模型215
  - 用TensorBoard来可视化图和训练曲线216
  - 命名作用域219
  - 模块化220
  - 共享变量222
  - 练习225
- 第10章 人工神经网络简介227
  - 从生物神经元到人工神经元227
  - 生物神经元228
  - 具有神经元的逻辑计算229
  - 感知器230
  - 多层感知器和反向传播233
  - 用TensorFlow的高级API来训练MLP236
  - 使用纯TensorFlow训练DNN237
  - 构建阶段237
  - 执行阶段240
  - 使用神经网络241
  - 微调神经网络的超参数242
  - 隐藏层的个数242
  - 每个隐藏层中的神经元数243
  - 激活函数243
  - 练习244
- 第11章 训练深度神经网络245
  - 梯度消失/爆炸问题245
  - Xavier初始化和He初始化246
  - 非饱和激活函数248
  - 批量归一化250
  - 梯度剪裁254
  - 重用预训练图层255
  - 重用TensorFlow模型255
  - 重用其他框架的模型256
  - 冻结低层257
  - 缓存冻结层257
  - 调整、丢弃或替换高层258
  - 模型动物园258
  - 无监督的预训练259
  - 辅助任务中的预训练260
  - 快速优化器261
  - Momentum优化261
  - Nesterov梯度加速262
  - AdaGrad263
  - RMSProp265
  - Adam优化265

学习速率调度267  
通过正则化避免过度拟合269  
提前停止269  
1和2正则化269  
dropout270  
最大范数正则化273  
数据扩充274  
实用指南275  
练习276  
第12章 跨设备和服务器的分布式TensorFlow279  
一台机器上的多个运算资源280  
安装280  
管理GPU RAM282  
在设备上操作284  
并行执行287  
控制依赖288  
多设备跨多服务器288  
开启一个会话290  
master和worker服务290  
分配跨任务操作291  
跨多参数服务器分片变量291  
用资源容器跨会话共享状态292  
使用TensorFlow队列进行异步通信294  
直接从图中加载数据299  
在TensorFlow集群上并行化神经网络305  
一台设备一个神经网络305  
图内与图间复制306  
模型并行化308  
数据并行化309  
练习314  
第13章 卷积神经网络31  
· · · · · (收起)

[机器学习实战：基于Scikit-Learn和TensorFlow\\_下载链接1\\_](#)

标签

机器学习

tensorflow

sklearn

深度学习

Python

人工智能

TensorFlow

计算机

## 评论

WOC，中文版缺代码，执行出来结果和书上都不一样，而且深度学习部分的翻译也很差劲，建议直接参看长评“川上月的github”

-----  
原版非常好，后面翻译的什么玩意！

-----  
太乱了（包括原版）。

-----  
忍不住从五分降低到四分，后面深度学习那块的翻译者是机器翻译后根本没调吧。读起来根本就晦涩曲解，出书可是会一直流传的，这骂名且背着吧。  
等看具体内容介绍部分的感觉吧，反正感觉tf介绍很不好，还不如网贴。  
后续深度学习部分，有些像文献综述，作者的深度和广度的确是有的，只是不太适合初学者扎实基础用，很多方法先有个概览，后面再找书扎扎实实推导吸收。  
机器学习或者深度学习，是一个需要实践推动的学科，需要跑模型，思考，再去理解某些优化方式或者模型的特点，以及表现优秀的原因。

-----  
当得起”实战“二字。

-----  
刚看了两页，中文版这翻的什么玩意儿？机翻都不至于吧？  
“在你意识到之前，它甚至会驾驶你的汽车”？  
“你们公司有大量的数据，如果知道去哪儿找，你会挖掘出一些隐藏的宝石”？

“我们会覆盖很多技术”？ “我们将使用真正生产就绪的Python框架中的算法”？  
“而不是每个算法的玩具版本”？  
准确地说到这句话才一页半，看得十分尴尬……生怕读者不知道是译本？  
对不起，看中文版是我的错，我还是去看英文版吧  
补：刚刚本来实事求是标记的是“在读”，仔细想了一下应该不会再读了，所以改为标记“已读”

-----  
在熟pandas/numpy/python的前提下，还需要配合专门讲算法的书才能看的下去看得懂吧

-----  
简单翻看。内容挺全面的，线性回归、SVM、决策树、集成学习、PCA、Tensorflow的使用，CNN、RNN、AutoEncoder、强化学习，大部分都有样例代码，虽然讲的不算太深，但入门不错

-----  
略读了一遍

-----  
刚学AI 就买了这本书，血赚

-----  
神经网络之前认真看了一下，果然还是看到代码会有体会

-----  
名为实战，但不是项目的实战，而是用基于python讲解理论知识的实战。上半卷sklearn相见恨晚，下半卷tensorflow似读天书。阅读门槛其实比较高，要能理解，必须对从sklearn到tensorflow的调包，以及从机器学习到深度学习的理论都有基本的了解。

-----  
我感觉还挺好的 可惜自己理解能力有限。

-----  
可以了解反人类工具时代的tensorflow

-----  
后半部分翻译欠妥，可以对照github阅读。

-----  
市面上最好的相关书籍资料了

-----  
最新一版的翻译已经非常好了

-----  
对于入门来说很棒，不劝退，很基础实用。感谢网友的github译本。

-----  
6星给前半讲传统机器学习的部分，简洁明了，字字珠玑；4星给后半讲深度学习的部分，内容和翻译都有待提高。

-----  
机器学习实践第一书

-----  
[机器学习实战：基于Scikit-Learn和TensorFlow 下载链接1](#)

## 书评

比一些照着pakcage的API tutorial抄出来的书姿势水平不知道高到哪里去了。  
个人认为这本书最精华的部分在于Appendix B  
机器学习项目清单，基本上工业界做一套Machine Learning解决方案顺着这个checklist问一遍自己就够了，需要Presentation的场合按照这个结构来组织也非常合适...

-----  
<https://github.com/it-ebooks/hands-on-ml-zh>  
=====

-----  
=====

<https://github.com/DeqianBai/Hands-on-Machine-Learning>

=====

自己翻译的版本，还在更新，打开一个Jupyter  
文件就可以一边学习理论，一遍进行操作验证 原书的代码示例部...

-----

挺不错的，推荐做ML的同学都拿来看看，一定能学到不少东西，尤其是接触没多久的  
不足之处是例子还是稍显不足，我个人更想要Kaggle真题解析  
一些我比较喜欢的地方如下 1. 2-3章适合所有刚接触数据科学的同学 第2章 California  
housing（加州区域房价）的例子非常实际，能学到很...

-----

第一部分写scikit的还行，后面第二部分关于神经网络部分，原文写的就乱，很多术语  
代码该解释的不解释，写的稀里糊涂，翻译更是糊涂，完全当不起5星。  
举个例子，第13章330页最下面，“最后一层是不言而喻的:放弃正则化”，翻译的人你  
给我出来，解释一下什么是放弃正则化，那tm...

-----

目前读到第5章，对于小白来说虽然略难但是受益匪浅啊，至于书的质量前面褒扬的话  
绝对不为过，说说我的自虐历程，到第二章实例分析的时候作者说撸起袖子别怕脏手  
，我就乖乖的开始跟着撸代码 python不太好，各种不懂有没有  
但是也不管那么多就是强撸，到现在很多以前不懂...

-----

tensorflow的官方文档写的比较乱，这本书的出现，恰好拯救了一批想入门tf，又看不  
进去官方文档的人。行文非常棒，例子丰富，有助于工程实践。这本书上提到了一些理  
论，简单形象；但是，理论不是此书的重点，也不应是此书的重点。这本书对于机器学  
习小白十分友好，读完了也就差...

-----

明年才出版，结合这两者的书很顺应时代需求啊，希望能出中文版，虽然英文读起来更  
好，但是为了效率，学习技术还是中文的来得更快些。

-----

[机器学习实战：基于Scikit-Learn和TensorFlow\\_下载链接1](#)