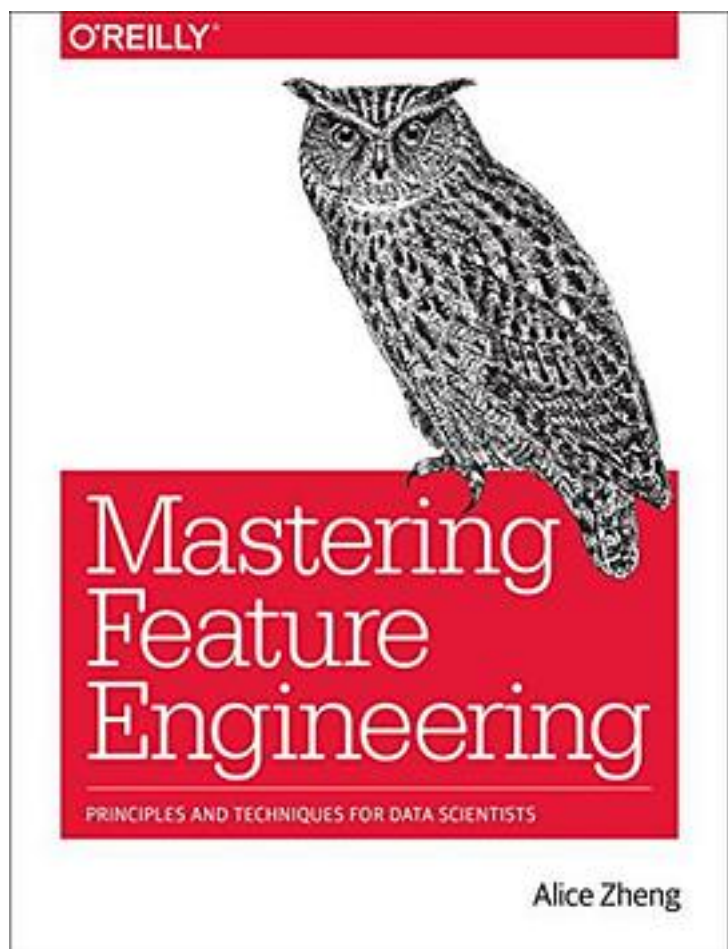


# Feature Engineering for Machine Learning Models



[Feature Engineering for Machine Learning Models\\_ 下载链接1](#)

著者:Alice Zheng

出版者:O'Reilly

出版时间:2017-12-31

装帧:Paperback

isbn:9781491953242

特征工程对于应用机器学习来说是基础的，但是使用域知识来加强你的预测模型既困难成本又高。为了弥补特征工程现有资料的不足，本书将会为初中级数据科学家讲解如何处理这项广泛应用却鲜见讨论的技术。

作者Alic

Zheng会讲解常用的练习和数学原理，以帮助工程师分析新数据和任务的特征。如果你理解基本的机器学习概念，如有监督学习和无监督学习，那么你已经准备好学习本书了。你不仅会学习到如何以一种系统化和原理化的方式部署特征工程，并且还会学习如何更好地实践数据科学。

作者介绍:

Alice是一家位于西雅图的提供可扩展数据分析工具的创业公司GraphLab的数据科学部门的负责人。Alice喜欢处理数据，以方便他人能够使用数据。她是一名在机器学习领域的工具开发者和专家。她的研究领域有软件诊断、计算机网络安全以及社会网络分析。在加入GraphLab之前，她作为一名研究院就职于在Redmond的微软研究院。她拥有数学专业的文学学士学位以及计算机科学的博士学位，两者皆获得与加州大学伯克利分校。

目录: Chapter 1 Introduction

Chapter 2 Fancy Tricks with Simple Numbers

Chapter 3 Basic Feature Engineering for Text Data: Flatten and Filter

Chapter 4 The Effects of Feature Scaling: From Bag-of-Words to Tf-Idf

Chapter 5 Counts and Categorical Variables: Counting Eggs in the Age of Robotic Chickens

Chapter 6 Dimensionality Reduction: Squashing the Data Pancake with PCA

Chapter 7 Non-Linear Featurization and Model Stacking

Chapter 8 Automating the Featurizer: Image Feature Extraction and Deep Learning

Appendix A Linear Modeling and Linear Algebra Basics

• • • • • ([收起](#))

[Feature Engineering for Machine Learning Models\\_ 下载链接1](#)

## 标签

机器学习

特征工程

数据科学

数据分析

ML

计算机

大数据

数据挖掘

## 评论

入门还行。但有些东西还是有点旧

-----  
非常简单，连入门都说不上

-----  
还可以。基本都是大众方法。

-----  
一般吧，工业级超大数据集就一般不用这些方法了。计算广告领域，如果不说深度学习，LR的特征工程，基本就是离散化一切，找新特征和交叉特征。

-----  
特征工程入门介绍

-----  
这本书好像改名叫mastering feature engineering了？  
对于入门者来说讲得比较系统，对于我这小白来说查缺补漏温故知新，对bin-counting很有兴趣，可惜没有找到可用的代码。先knearest cluster features再建立linear model可在某些情况下匹敌ensemble models，下次试试。

-----  
常见的特征处理trick介绍的还比较系统。

-----

需要复习线性代数

-----  
不值一读，废话多

-----  
基本都是跳着看的，太啰嗦了。。

-----  
practical feature engineering with some mathematical explanations

-----  
太入门了。。应该叫Introduction To Feature Engineering

-----  
因为书里的一段话直接让我们的模型从复赛资格边缘慢慢地冲进了现在前十，感谢

-----  
[Feature Engineering for Machine Learning Models\\_下载链接1](#)

## 书评

<https://github.com/apacheecn/feature-engineering-for-ml-zh>  
=====

-----  
特征工程是数据科学工程的核心，目前关于这个话题专门的书籍不多。本书通过概念（不是理论）和案例代码相结合的方式，还该了特征工程中的一些基础技术。包括分类型变量编码，数值型数据的分箱，变换。文本处理，PCA以及基于模型的特征工程。模型堆叠和k-均值特征化。最后简单介...

在图书馆看到的，感觉内容很棒，来豆瓣mark一下，上班有钱后买一本。吐个槽，书有点薄，59元略贵。虽然知识无价，不过对比国内出版物环境，嗯...相对有点点问题。声明一下，这本书不是入门书籍，不适合机器学习入门/python入门的来看。虽然英文名叫Feature Engineering for ...

-----  
[Feature Engineering for Machine Learning Models\\_下载链接1](#)