

《Python 语言构建机器学习系统》

图书基本信息

书名：《Python 语言构建机器学习系统 第2版（影印版）》

13位ISBN编号：9787564160624

出版时间：2016-1-1

作者：Luis · Pedro · Coelho,WilliRi · chert

页数：301

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《Python 语言构建机器学习系统》

内容概要

运用机器学习获得对于数据的深入洞见，是现代应用开发者和分析师的关键技能。Python是一种可以用于开发机器学习应用的语言。作为一种动态语言，它可以进行快速探索和实验。利用其的开源机器学习库，你可以在快速尝试很多想法的同时专注于手头的任务。

《Python语言构建机器学习系统（第2版 影印版 英文版）》展示了如何在原始数据中寻找模式的具体方法，从复习Python机器学习知识和介绍程序库开始，你将很快进入应对正式而真实的数据集项目环节，运用建模技术，创建推荐系统。然后，《Python语言构建机器学习系统（第2版 影印版 英文版）》介绍了主题建模、篮子分析和云计算等高级主题。这些内容将拓展你的能力，让你能够创建大型复杂系统。

有了《Python语言构建机器学习系统（第2版 影印版 英文版）》，你就能获得构建自有系统所需的工具和知识，定制化解决实际的数据分析相关问题。

书籍目录

- Preface
- Chapter 1 : Getting Started with Python Machine Learning
 - Machine learning and Python - a dream team
 - What the book will teach you (and what it will not)
 - What to do when you are stuck
 - Getting started
 - Introduction to NumPy , SciPy , and matplotlib
 - Installing Python
 - Chewing data efficiently with NumPy and intelligentlywith SciPy
 - Learning NumPy
 - Indexing
 - Handling nonexisting values
 - Comparing the runtime
 - Learning SciPy
 - Our first (tiny) application of machine learning
 - Reading in the data
 - Preprocessing and cleaning the data
 - Choosing the right model and learning algorithm
 - Beforebuilding our first model . . .
 - Starting with a simple straight line
 - Towards some advanced stuff
 - Stepping back to go forward - another look at our data
 - Training and testing
 - Answering our initial question
 - Summary
- Chapter 2 : Classifying with Real-world Examples
 - The Iris dataset
 - Visualization is a good first step
 - Building our first classification model
 - Evaluation - holding out data and cross-validation
 - Building more complex classifiers
 - A more complex dataset and a more complex classifim
 - Learning about the Seeds dataset
 - Features and feature engineering
 - Nearest neighbor classification
 - Classifying with scikit-learn
 - Looking at the decision boundaries
 - Binary and multiclass classification
 - Summary
- Chapter 3 : Clustering - Finding Related Posts
 - Measuring the relatedness of posts
 - How not to do it
 - How to do it
 - Preprocessing - similarity measured as a similar number of common words
 - Converting raw text into a bag of words
 - Counting words
 - Normalizing word count vectors

- Removing less important words
- Stemming
- Stop words on steroids
- Our achievements and goals
- Clustering
- K-means
- Getting test data to evaluate our ideas on
- Clustering posts
- Solving our initial challenge
- Another look at noise
- Tweaking the parameters
- Summary
- Chapter 4 : Topic Modeling
- Latent Dirichlet allocation
- Building a topic model
- Comparing documents by topics
- Modeling the whole of Wikipedia
- Choosing the number of topics
- Summary
- Chapter 5 : Classification - Detecting Poor Answers
- Sketching our roadmap
- Learning to classify classy answers
- Tuning the instance
- Tuning the classifier
- Fetching the data
- Slimming the data down to chewable chunks
- Preselection and processing of attributes
- Defining what is a good answer
- Creating our first classifier
- Starting with kNN
- Engineering the features
- Training the classifier
- Measuring the classifier's performance
- Designing more features
- Deciding how to improve
- Bias-variance and their tradeoff
- Fixing high bias
- Fixing high variance
- High bias or low bias
- Using logistic regression
- A bit of math with a small example
- Applying logistic regression to our post classification problem
- Looking behind accuracy- precision and recall
- Slimming the classifier
- Ship it !
- Summary
- Chapter 6 : Classification II - Sentiment Analysis
- Sketching our roadmap
- Fetching the Twitter data

- Introducing the Naive Bayes classifier
- Getting to know the Bayes' theorem
- Being naive
- Using Naive Bayes to classify
- Accounting for unseen words and other oddities
- Accounting for arithmetic underflows
- Creating our first classifier and tuning it
- Solving an easy problem first
- Using all classes
- Tuning the classifier's parameters
- Cleaning tweets
- Taking the word types into account
- Determining the word types
- Successfully cheating using SentiWordNet
- Our first estimator
- Putting everything together
- Summary
- Chapter 7 : Regression
- Predicting house prices with regression
- Multidimensional regression
- Cross-validation for regression
- Penalized or regularized regression
- L1 and L2 penalties
- Using Lasso or ElasticNet in scikit-learn
- Visualizing the Lasso path
- P-greater-than-N scenarios
- An example based on text documents
- Setting hyperparameters in a principled way
- Summary
- Chapter 8 : Recommendations
- Rating predictions and recommendations
- Splitting into training and testing
- Normalizing the training data
- A neighborhood approach to recommendations
- A regression approach to recommendations
- Combining multiple methods
- Basket analysis
- Obtaining useful predictions
- Analyzing supermarket shopping baskets
- Association rule mining
- More advanced basket analysis
- Summary
- Chapter 9 : Classification - Music Genre Classification
- Sketching our roadmap
- Fetching the music data
- Converting into a WAV format
- Looking at music
- Decomposing music into sine wave components
- Using FFT to build our first classifier

- Increasing experimentation agility
- Training the classifier
- Using a confusion matrix to measure accuracy in multiclass problems
- An alternative way to measure classifier performance using receiver-operator characteristics
- Improving classification performance with Mel Frequency Cepstral Coefficients
- Summary
- Chapter 10 : Computer Vision
 - Introducing image processing
 - Loading and displaying images
 - Thresholding
 - Gaussian blurring
 - Putting the center in focus
 - Basic image classification
 - Computing features from images
 - Writing your own features
 - Using features to find similar images
 - Classifying a harder dataset
 - Local feature representations
 - Summary
- Chapter 11 : Dimensionality Reduction
 - Sketching our roadmap
 - Selecting features
 - Detecting redundant features using filters
 - Correlation
 - Mutual information
 - Asking the model about the features using wrappers
 - Other feature selection methods
 - Feature extraction
 - About principal component analysis
 - Sketching PCA
 - Applying PCA
 - Limitations of PCA and how LDA can help
 - Multidimensional scaling
 - Summary
- Chapter 12 : Bigger Data
 - Learning about big data
 - Using jug to break up your pipeline into tasks
 - An introduction to tasks in jug
 - Looking under the hood
 - Using jug for data analysis
 - Reusing partial results
 - Using Amazon Web Services
 - Creating your first virtual machines
 - Installing Python packages on Amazon Linux
 - Running jug on our cloud machine
 - Automating the generation of clusters with StarCluster

《Python 语言构建机器学习系统》

Summary

Appendix : Where to Learn More Machine Learning

Online courses

Books

Question and answer sites

Blogs

Data sources

Getting competitive

All that was left out

Summary

Index

《Python 语言构建机器学习系统》

精彩短评

1、Step2就靠这本书了，比机器学习实战更实用。缺点是概念相对忽略，SVM和神经网络没有涉及

《Python 语言构建机器学习系统》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu000.com