R语言机器学习参考手册（影印版）

Machine Learning with R Cookbook

Yu-Wei Chiu (David Chiu) 著

[PACKT] PUBLISHING

使用简单且易于使用的R代码，
通过110多个参考方案分析数据并构建可预测模型

东南大学出版社
SOUTHEAST UNIVERSITY PRESS

R语言机器学习参考手册(影印版)

*Yu-Wei Chiu (David Chiu)*著

南京　东南大学出版社

图书在版编目(CIP)数据

R 语言机器学习参考手册:英文/丘祐玮(Yu - wei, Chiu)著. —影印本. —南京:东南大学出版社,2016.1
书名原文:Machine Learning with R Cookbook
ISBN 978 - 7 - 5641 - 6063 - 0

Ⅰ. ①R… Ⅱ. ①丘… Ⅲ. ①程序语言—程序设计—技术手册—英文 Ⅳ. ①TP312 - 62

中国版本图书馆 CIP 数据核字(2015)第 243405 号

R 语言机器学习参考手册(影印版)

出版发行:东南大学出版社
地　　址:南京四牌楼 2 号　　邮编:210096
出 版 人:江建中
网　　址:http://www.seupress.com
电子邮件:press@seupress.com
印　　刷:常州市武进第三印刷有限公司
开　　本:787 毫米×980 毫米　16 开本
印　　张:27.5
字　　数:539 千字
版　　次:2016 年 1 月第 1 版
印　　次:2016 年 1 月第 1 次印刷
书　　号:ISBN 978 - 7 - 5641 - 6063 - 0
定　　价:78.00 元

本社图书若有印装质量问题,请直接与营销部联系。电话(传真):025 - 83791830

About the Author

Yu-Wei, Chiu (David Chiu) is the founder of LargitData (`www.LargitData.com`). He has previously worked for Trend Micro as a software engineer, with the responsibility of building big data platforms for business intelligence and customer relationship management systems. In addition to being a start-up entrepreneur and data scientist, he specializes in using Spark and Hadoop to process big data and apply data mining techniques for data analysis. Yu-Wei is also a professional lecturer and has delivered lectures on Python, R, Hadoop, and tech talks at a variety of conferences.

In 2013, Yu-Wei reviewed *Bioinformatics with R Cookbook*, *Packt Publishing*. For more information, please visit his personal website at `www.ywchiu.com`.

I have immense gratitude for my family and friends for supporting and encouraging me to complete this book. I would like to sincerely thank my mother, Ming-Yang Huang (Miranda Huang); my mentor, Man-Kwan Shan; the proofreader of this book, Brendan Fisher; Taiwan R User Group; Data Science Program (DSP); and other friends who have offered their support.

Credits

Author

Yu-Wei, Chiu (David Chiu)

Reviewers

Tarek Amr

Abir Datta (data scientist)

Saibal Dutta

Ratanlal Mahanta
(senior quantitative analyst)

Ricky Shi

Jithin S.L

Commissioning Editor

Akram Hussain

Acquisition Editor

James Jones

Content Development Editor

Arvind Koul

Technical Editors

Tanvi Bhatt

Shashank Desai

Copy Editor

Sonia Cheema

Project Coordinator

Nikhil Nair

Proofreaders

Simran Bhogal

Joanna McMahon

Jonathan Todd

Indexer

Mariammal Chettiyar

Graphics

Sheetal Aute

Abhinash Sahu

Production Coordinator

Melwyn D'sa

Cover Work

Melwyn D'sa

About the Reviewers

Tarek Amr currently works as a data scientist at bidx in the Netherlands. He has an MSc degree from the University of East Anglia in knowledge discovery and data mining. He also volunteers at the Open Knowledge Foundation and School of Data, where he works on projects related to open data and gives training in the field of data journalism and data visualization. He has reviewed another book, *Python Data Visualization Cookbook, Packt Publishing*, and is currently working on writing a new book on data visualization using D3.js.

You can find out more about him at `http://tarekamr.appspot.com/`.

Abir Datta (data scientist) has been working as a data scientist in Cognizant Technology Solutions Ltd. in the fields of insurance, financial services, and digital analytics verticals. He has mainly been working in the fields of analytics, predictive modeling, and business intelligence/analysis in designing and developing end-to-end big data integrated analytical solutions for different verticals to cater to a client's analytical business problems. He has also developed algorithms to identify the latent characteristics of customers so as to take channelized strategic decisions for much more effective business success.

Abir is also involved in risk modeling and has been a part of the team that developed a model risk governance platform for his current organization, which has been widely recognized across the banking and financial service industry.

Saibal Dutta is presently researching in the field of data mining and machine learning at the Indian Institute of Technology, Kharagpur, India. He also holds a master's degree in electronics and communication from the National Institute of Technology, Rourkela, India. He has also worked at HCL Technologies Limited, Noida, as a software consultant. In his 4 years of consulting experience, he has been associated with global players such as IKEA (in Sweden), Pearson (in the U.S.), and so on. His passion for entrepreneurship has led him to start his own start-up in the field of data analytics, which is in the bootstrapping stage. His areas of expertise include data mining, machine learning, image processing, and business consultation.

Ratanlal Mahanta (senior quantitative analyst) holds an MSc in computational finance and is currently working at the GPSK Investment Group as a senior quantitative analyst. He has 4 years of experience in quantitative trading and strategy developments for sell-side and risk consultation firms. He is an expert in high frequency and algorithmic trading.

He has expertise in the following areas:

- Quantitative trading: FX, equities, futures, options, and engineering on derivatives
- Algorithms: Partial differential equations, Stochastic Differential Equations, Finite Difference Method, Monte-Carlo, and Machine Learning
- Code: R Programming, C++, MATLAB, HPC, and Scientific Computing
- Data analysis: Big-Data-Analytic [EOD to TBT], Bloomberg, Quandl, and Quantopian
- Strategies: Vol-Arbitrage, Vanilla and Exotic Options Modeling, trend following, Mean reversion, Co-integration, Monte-Carlo Simulations, ValueatRisk, Stress Testing, Buy side trading strategies with high Sharpe ratio, Credit Risk Modeling, and Credit Rating

He has already reviewed two books for Packt Publishing: *Mastering Scientific Computing with R* and *Mastering Quantitative Finance with R*.

Currently, he is reviewing a book for Packt Publishing: *Mastering Python for Data Science.*

Ricky Shi is currently a quantitative trader and researcher, focusing on large-scale machine learning and robust prediction techniques. He obtained a PhD in the field of machine learning and data mining with big data. Concurrently, he conducts research in applied math. With the objective to apply academic research to real-world practice, he has worked with several research institutes and companies, including Yahoo! labs, AT&T Labs, Eagle Seven, Morgan Stanley Equity Trading Lab (ETL), and Engineers Gate Manager LP, supervised by Professor Philip S. Yu.

His research interest covers the following topics:

- Correlation among heterogeneous data, such as social advertising from both the users' demographic features and users' social networks
- Correlation among evolving time series objects, such as finding dynamic correlations, finding the most influential financial products (shaker detection, cascading graph), and using the correlation in hedging and portfolio management
- Correlation among learning tasks, such as transfer learning

Jithin S.L completed his BTech in information technology from Loyola Institute of Technology and Science. He started his career in the field of analytics and then moved to various verticals of big data technology. He has worked with reputed organizations, such as Thomson Reuters, IBM, and Flytxt, under different roles. He has worked in the banking, energy, healthcare, and telecom domains and has handled global projects on big data technology.

He has submitted many research papers on technology and business at national and international conferences.

His motto in life is that learning is always a neverending process that helps in understanding, modeling, and presenting new concepts to the modern world.

I surrender myself to God almighty who helped me to review this book in an effective way. I dedicate my work on this book to my dad, Mr. N. Subbian Asari, my lovable mom, Mrs. M. Lekshmi, and my sweet sister, Ms. S.L Jishma, for coordinating and encouraging me to write this book.

Last but not least, I would like to thank all my friends.

---

www.PacktPub.com

Support files, eBooks, discount offers, and more

For support files and downloads related to your book, please visit www.PacktPub.com.

Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.PacktPub.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at service@packtpub.com for more details.

At www.PacktPub.com, you can also read a collection of free technical articles, sign up for a range of free newsletters and receive exclusive discounts and offers on Packt books and eBooks.

https://www2.packtpub.com/books/subscription/packtlib

Do you need instant solutions to your IT questions? PacktLib is Packt's online digital book library. Here, you can search, access, and read Packt's entire library of books.

Why subscribe?

- Fully searchable across every book published by Packt
- Copy and paste, print, and bookmark content
- On demand and accessible via a web browser

Free access for Packt account holders

If you have an account with Packt at www.PacktPub.com, you can use this to access PacktLib today and view 9 entirely free books. Simply use your login credentials for immediate access.

Table of Contents

Table of Contents