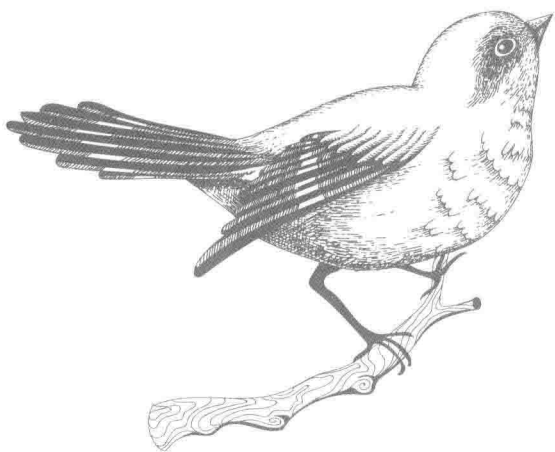


R语言预测实战

游皓麟◎著

本书注重算法理论与实际案例相结合，将枯燥无味的预测算法原理用R语言重现，
并通过案例让读者掌握预测模型的应用。

CDA数据分析师系列丛书



R语言预测实战

游皓麟 著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

R 语言具有上手快、效率高的特点，它横跨金融、生物、医学、互联网等多个领域，主要用于统计、建模及可视化。预测是数据挖掘的主要作用之一，也是大数据时代的核心价值所在。基于 R 语言来做预测，兼具效率和价值的双重属性，这是本书的一大亮点。

本书共分为三部分。第一部分讲预测基础，主要涵盖预测概念理解、预测方法论、分析方法、特征技术、模型优化及评价，读者通过这部分内容的学习，可以掌握进行预测的基本步骤和方法思路；第二部分讲预测算法，该部分包含了多元回归分析、复杂回归分析、时间序列及进阶算法，内容比较有难度，需要细心体会；第三部分讲预测案例，包括短期日负荷曲线预测和股票价格预测两个实例，读者可以了解到实施预测时需要关注的技术细节。

希望读者在看完本书后，能够将本书的精华融会贯通，进一步在工作和学习实践中提炼价值。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

R 语言预测实战 / 游皓麟著. —北京：电子工业出版社，2016.10

（CDA 数据分析师系列丛书）

ISBN 978-7-121-29854-7

I. ①R… II. ①游… III. ①程序语言—程序设计 IV. ①TP312

中国版本图书馆 CIP 数据核字(2016)第 211982 号

策划编辑：石 倩

责任编辑：石 倩

印 刷：三河市双峰印刷装订有限公司

装 订：三河市双峰印刷装订有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编：100036

开 本：787×980 1/16 印张：30 字数：768 千字 彩插：2

版 次：2016 年 10 月第 1 版

印 次：2016 年 10 月第 1 次印刷

定 价：79.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888, 88258888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819 faq@phei.com.cn。

序 一

“我已经工作满一年了”，多年以后我仍然会想起那一幕，一个很朴实的小伙子，说的一句很朴实的话。在简单的交谈后，我认识了皓麟。当时感觉他是一个比较平淡的人，但是多年的经验却让我感觉到他是一个有想法的人。在以后的工作中，我们经常在一起合作，研究、讨论并实践许多机器学习模型，从分类模型、关联模型到预测模型，有幸的是，这些模型都取得了预期效果，得到了用户的肯定。整个过程中，感觉皓麟聪慧、思路敏锐、责任心强，软件操作能力特别是 R 语言能力一流，是一个较为理想的工作伙伴。更可贵的是，皓麟还表现出了比同龄人成熟的一面。

“我打算写一本有关预测实战的书”，记得一年前的一天接到皓麟的电话。我知道这几年皓麟在机器学习算法，特别是预测专题研究、实践方面取得了很多成果。他做过演讲，在网校讲过课，已在圈内小有名气。以前听他说过要写一本书，这次真的下定决心做了，而且当时他决定回成都发展，已经辞职，这种情况下，我知道他的压力还是蛮大的。在后继的时间里，皓麟经常打电话和我探讨我知道的、不知道的、懂的或不懂的各种预测、优化算法、数学理论基础，也包括以效果为导向的各种模型的组合实践，感觉他在理论修养、预测模型的把控方面已经达到了较高水准。期间，我能感觉到皓麟在写作过程中表现出的克服理论、实践中诸多困难的很强的定力、毅力与良好的心态。

现如今，智能浪潮如火如荼，算法，特别是有效的算法发挥着越来越大的作用。在这其中，预测算法始终是重要的算法方向。它可以让我们知道谁才是我们的消费者，某客户打算购买的产品是什么，哪类用户最可能流失，下个月景区的人流量是多少，哪支球队更有可能获胜，智能聊天中下一句更可能的语句组合是什么等诸多预测问题。如果预测模型精度足够高，它能给业务带来很多的可能性，并进一步创造价值。

本书的案例代码以 R 语言为主要实现工具。R 语言有着丰富的软件包，截至目前 R 语言包的数量已超 7000 个，包含诸多行业的诸多算法，涉及分析、挖掘、机器学习、智能算法类、优化类以及微分类等各种算法和操作，在处理中小数据时是最佳选择，在处理大数据时也有诸多优势。R 语言与 Hadoop、Spark 等主流平台也有接口，具有很好的扩展性。

本书首先讲述了预测的概述，包括预测的优势、特征等。随后从方法论的角度，让读者更加能够理解并掌握处理预测问题时的思路。接着，进入具体分析技术的讲解。特别是皓麟有深入研究的特征构建和特征选择技术，这部分是较为关键的环节，没有良好的特征或更广泛的输入，后继算法再好，也难出很好的效果，故作者在此花了较大精力。在模型参数优化阶段主要用了在工业界应用广泛的智能类算法，选择遗传、粒子群与模拟退火等几种经典算法，同时将原理与实现步骤也一一说明。

在算法阶段，主要讲解了常见的回归算法，也讲解了复杂的回归算法，展示了常见的各种时间序列算法以及非线性方面。最后以真实数据为例，详细讲解了两个案例，这部分对生产的模型有较大的借鉴作用。

本书思路清晰，算法较为全面新颖，案例真实且效果不错，对从事分析、挖掘与机器学习的读者具有较大的参考价值。

刘爱松

亚信科技大数据平台部人工智能专家

序 二

随着大数据的概念变得越来越流行，对数据的探索、分析和预测已经成为大数据分析领域的基本技能之一。作为探索和分析数据的基本理论和工具，利用科学的数据挖掘技术进行数据预处理和借助成熟的机器学习算法进行模型拟合成为时下的热门技术之一。R 语言作为功能强大并且免费的数据挖掘工具，在数据挖掘领域获得了越来越多用户的青睐。近几年，中国 R 语言大会在越来越多的高校举行，参会人数也呈现爆发式增长，尤其是业界人士占据了相当大的比例。如今年举办的“第九届中国 R 语言大会（北京会场）”比往届有了更大的突破，会议共设有 22 个分会场，126 场主题报告，覆盖大数据技术、机器学习、可视分析、社交网络、互联网金融、量化投资、人网物联、生物信息、自然语言等诸多当下热门话题。报名非常火爆，人数突破 5000 人，报名单位超过 1500 个。不仅创下历届之最，也使本次会议成为亚洲地区规模最大的数据科学盛会之一。

本书作者游浩麟先生是多届“中国 R 语言大会”的演讲嘉宾，其在广州会场分享的主题《R 语言并行计算实践》广受学生们喜爱，也为业界的数据分析人员在利用 R 语言处理大数据时如何解决内存问题提供了很好的解决方案。作者多年专注于数据分析、挖掘、大数据领域，在互联网/电信/电力等行业具有丰富的数据分析与挖掘建模经验。

非常荣幸能够受邀为本书写序。虽然本人也玩 R 语言多年，但当我拿到本书样稿时，看目录就有很强的阅读欲望，整本书的条理清晰、思维缜密，内容由浅入深。阅读内容之后，发现本书注重算法理论与实际案例相结合，作者将枯燥无味的诸多预测算法原理用 R 语言重现，并通过案例让读者掌握预测模型的应用。

本书分为三部分共 12 章。

第一部分为预测入门篇，包括预测入门、方法论、特征构建及选择、模型参数优化及效果评估等内容。我们做任何事情都需要有方法论的指导，数据预测也不例外。作者根据从事预测工作多年的经验，从数据挖掘 CRISP-DM 标准过程提炼出预测基本流程：确定主题、收集数据、选择方法、分析规律、建立模型、评估效果及发布模型的预测工作闭环。在实际生产环境中，Raw Data（原始数据）基本达不到直接建模的要求，需要经过数据处理技术或业务经验进行特征构建及选择，将 Raw Data 转换成适合建模的数据。特征的构建与选择在提供预测精度方面越来越受到重视，甚至影响到预测项目的成败。本书详细介绍了特征构建技术以及特征选择方法。记得曾经有学员问我为什么他自己构建的系谱聚类结果看起来特别怪异，相信他如果看完本书的特征构建技术等章节内容就会找到答案。模型评估是最重要也是最容易被大家忽视的一个环节。我的团队曾经有同事利用神经网络模型对玩家付费进行流失预测，得出准确率高达 97% 的预测模型，结果用其他数据预测出来的结果

却惨不忍睹，是因为其未对模型进行是否存在过拟合的检验。故模型评估的环节至关重要，它就像一道关口，只有通过它的模型，才能给予被用于真实环境的机会；只有不断通过它的模型，才能被认为是稳定可靠的。

第二部分是预测算法篇，包括线性回归、复杂回归分析及时间序列分析三大主题。其中线性回归包括经典的线性回归及其模型优化，当出现多重共线性问题时的 Ridge 及 Lasso 回归；复杂回归分析包括梯度提升回归树、神经网络、支持向量机等近几年非常流行的机器学习算法。

第三部分是预测应用篇，其中一个案例是作者在做电力数据时的一个短期负荷曲线预测：从业务背景介绍开始，包括数据采集、数据处理、数据探索、数据建模及模型评估一整套完整的数据预测解决方案，读者可以很好地将此方法移植到其他的生产数据中。

本书既强调深入的算法理论，又强调实际应用。书中有大量的 R 语言代码帮助读者理解算法原理及数据预测。如果是有兴趣进入数据挖掘行业的读者，这本书非常有参考和指导意义，会使你成为数据处理及数据建模预测的多面手，不至于拿到原始数据时一筹莫展。如果是已具有一定数据挖掘经验的专业人士，本书亦能作为一面“哈哈镜”，照出自己平时工作中需要完善的地方。

谢佳标
乐逗游戏

前 言

为什么要写这本书

2014 年对我来说具有特别的意义，这一年我含辛如苦地恶补了 R 语言、数据挖掘的进阶算法及应用，同年还多次参加了大数据、R 语言的会议，收获良多，认识了不少志同道合的朋友。也是在这一年，我成为了小象学院的兼职讲师，并录制了《数据挖掘之 R 语言实践》的课程，让我对数据挖掘的理解又上了一个台阶。这年的 10 月，我回四川老家参加了小妹的婚礼，了却了一桩心愿，从此小妹不再一个人奋斗了，衷心祝愿他们能够一生幸福。时值国庆，我带着全家去了趟九寨沟，虽然路途疲惫，但是号称童话世界的九寨沟给我留下了深刻的印象，水是从来没有见过的圣洁，风景更是美不胜收。也就是在这一年，因为工作需要，我开始接触实实在在的预测。起初我便觉得这事不好做，因为当时做的是短期日负荷曲线预测，并不像收入预测、景区人流量预测等情况只需要预测一个值，而是要预测一条曲线。我当时基本把国内关于负荷预测的书籍、博士硕士论文、期刊看了个遍，每天下班回去看，周末也看，该用的时间都用上了，这样我对负荷预测业务及算法的理解很快地提升了一个档次。但是这还不能让我满意，因为最终要拿数据来说话。我用业余时间写了一套 R 语言并行计算框架，可以同时支持成千上万台变压器数据的建模和预测。可是真正影响效果的不是平台，而是预测背后的那一整套逻辑。当时我就想有没有一套通用的预测方法论，可以直接拿过来使用。这个问题一直在我心底装了很久，后来通过与团队的沟通我们确立了自己的预测体系，虽然不能通用，但是可以解决目前的问题。接着我花了大量时间调测优化模型，期望以统一的框架搞定所有问题，那段时间我研究了许多的预测算法，这些算法包括高斯过程回归、神经网络、卡尔曼滤波等，为了提高预测精度，我使用了很多模型，并使用组合预测的方法来提升效果，甚至考虑通过预测误差来学习优化参数。就是这样的反复折腾，使我在模型调优上的水平很快提升，并且效果也达到了预期的预测目标。这个经历更让我意识到，现在不只是电力行业，其实在其他很多行业，包括电商、物流、金融等都对预测有着天生的依赖性。我们可以通过预测了解到各区域用户的需求趋势，为战略布局提供决策依据；我们也可以通过预测提前预知，使其按正常方式进行下去，我们所关注的指标的变化，并通过有效的途径，牵引向好的方向发展。预测的魅力在于控制、在于引导、在于把握未来。基于以上思考，我觉得很有必要做一个总结，将长期以来混乱的知识体系进行有条理的整理，以供大家学习、交流、讨论，一起成长。

阅读对象

- 对数据挖掘、机器学习、预测算法及商业预测应用感兴趣的大专院校师生
- 数据挖掘工程师或从事数据挖掘相关工作
- 各行各业的数据分析师
- 对数据挖掘、预测专题感兴趣的读者

勘误和支持

由于作者的水平有限，编写的时间也很仓促，书中难免会出现一些错误或者不准确的地方，恳请读者批评指正。读者可以把意见或建议直接发至我的邮箱 cador@sina.com。我会定期发布 Bug 勘误表，并统一回复。同时，如果你有什么问题，也可以发邮件来提问，我将尽力为读者提供最满意的解答，期待你们的反馈。

如何阅读这本书

本书包括三个部分，共有 12 章。

第 1 章介绍预测的基本概念，以及大数据时代预测的特点，并结合案例进行讲解，最后基于 R 语言讲解一个预测案例。本章适合初学者入门。

第 2 章介绍预测的方法论，其中预测流程是基础，它说明了预测实施的各个阶段。然后介绍了预测的指导原则，它是预测工作者必知必会的。另外，还介绍了预测工作的团队要求。本章内容适合长期实践，活学活用。

第 3 章介绍分析方法，本章内容是数据分析、数据挖掘常见的分析方法，出现在这里，主要是为预测技术的数据处理做铺垫。如果预测工作者没有掌握有效的分析思路和方法，去提炼有用的指标和特征，那么预测工作是很难进行下去的。本章提供了规律发现的常用方法和技巧。

第 4 章介绍特征构建技术，不仅介绍了常见的特征变换方法，还介绍了特征组合的方法，特别值得一提的是，本章包含了自动生成特征的方法，它是基于遗传编程实现的。从事数据挖掘的朋友都很清楚，好的特征在建模时是非常重要的，然而，有时我们拿到的基础数据直接拿去建模效果不见得好，如果进行规律挖掘，也比较费时费力，比较好的做法就是特征自动生成。感兴趣的朋友，可以认真学习这个章节。

第 5 章介绍特征选择的方法，特征选择适合于特征很多（成千上万个，甚至千百万个）的情况。本章介绍了使用单变量和多变量特征选择技术，读者可根据需要选择性阅读。

第 6 章介绍模型参数的优化。我们在建立数据挖掘和预测模型时，参数的确定通常不是一步到位的，往往需要做一些优化或改进，以提升最终的效果。本章介绍的遗传算法、粒子群优化、模拟

退火等问题求解算法，有助于找到模型的最优或接近最优的参数。

第7章介绍预测模型效果评估的方法，主要包括分类预测模型评估方法和数值预测模型评估方法。本章的内容包含了评估指标和评估图表，读者要活学活用，正确地选择和使用。

从第1章到第7章为本书的第一部分，主要介绍预测的入门知识，如果读者对预测有一定的功底，可以跳过本部分，直接进入第二部分，了解预测算法的基本原理和实现。

第8章介绍线性回归技术，主要包括多元线性回归、Ridge回归、Lasso回归、分位数回归、稳健回归的内容。在实际工作或实践中，读者应该有选择地使用对应的回归方法，以确保对回归问题的有效性。

第9章介绍复杂回归技术，主要包括梯度提升回归树（GBRT）、神经网络、支持向量机、高斯过程回归的内容。这是回归技术的进阶部分，涉及统计学习以及机器学习的内容，想挑战难度的朋友，一定要好好读一读这部分。

第10章介绍时间序列分析技术，主要包括Box-Jenkins方法、门限自回归模型、GARCH模型族、向量自回归模型、卡尔曼滤波器算法的内容。本章不仅介绍了常见的Box-Jenkins方法，还介绍了门限自回归等高阶时序分析技术。

从第8章到第10章为本书的第二部分，主要介绍预测算法，本部分的算法选择有一定的难度，基本包含了常见的以及部分高阶的预测回归算法，读者仔细阅读。

第11章介绍短期日负荷曲线预测技术，首先介绍电力行业负荷预测的行业知识，接着从预测的基本要求出发，经过预测的建模准备，进入预测建模的环节。本章使用了RBF神经网络和LS-SVMR两种算法来建立预测模型，并对预测效果进行了评估。

第12章介绍股票价格预测技术，首先从股票市场的简介开始，接着介绍股票数据的获取，然后基于VAR算法对预测模型进行了实现，最后检验了预测的准确性。

最后两章为本书的第三部分，主要介绍预测案例，由于写书太过仓促，案例部分略显单薄。另外，由于商业关系，有些案例分析细节内容不便在书中全面展开介绍，有兴趣的读者，可以邮件联系笔者。

致谢

首先，感谢电子工业出版社的编辑石倩，没有你的主动联系，我也不会这么快写这本书，也同时感谢电子工业出版社给我这次出版的机会！

其次，要感谢长期以来一直鼓励我、支持我的刘爱松，如果没有和你的长期分享交流，我也不会长得这么快。在此，也衷心感谢所有关心、支持、帮助过我的良师益友。

最后，感谢造物主给我一颗孜孜不倦的心，让我在学习的道路上不至于因工作忙碌而有所懈怠，也不至于因有所成就而不思进取。

青山不改，绿水长流。谨以此书，献给我最亲爱的家人和朋友，以及热爱 R 语言和从事数据分析、数据挖掘、预测相关工作的朋友们。

游皓麟
中国 成都

目 录

第一部分 预测入门篇

第 1 章 预测入门.....	2
1.1 什么是预测.....	3
1.1.1 预测的定义.....	3
1.1.2 预测的特点.....	4
1.1.3 预测的分类.....	4
1.1.4 预测的基本原则.....	5
1.2 大数据与预测.....	7
1.2.1 什么是大数据.....	7
1.2.2 大数据预测的优势.....	9
1.2.3 大数据预测的特征.....	9
1.2.4 大数据预测案例.....	12
1.3 预测利器之 R 语言.....	22
1.3.1 R 语言简介.....	22
1.3.2 R 语言预测初步.....	33
1.3.3 R 语言预测常用包.....	42
第 2 章 预测方法论.....	44
2.1 预测流程.....	45
2.1.1 确定主题.....	45
2.1.2 收集数据.....	47
2.1.3 选择方法.....	49
2.1.4 分析规律.....	53
2.1.5 建立模型.....	57
2.1.6 评估效果.....	60
2.1.7 发布模型.....	60
2.2 指导原则.....	61

2.2.1	界定问题	61
2.2.2	判断预测法	63
2.2.3	外推预测法	64
2.2.4	因果预测法	66
2.3	团队构成	66
2.3.1	成员分类	67
2.3.2	数据氛围	68
2.3.3	团队合作	70
第 3 章	分析方法	72
3.1	相关分析	73
3.1.1	自相关分析	73
3.1.2	偏相关分析	74
3.1.3	简单相关分析	75
3.1.4	互相关分析	83
3.1.5	典型相关分析	84
3.2	对应分析	88
3.3	频谱分析	92
3.4	趋势分析	94
3.5	聚类分析	96
3.5.1	K-Means 算法	96
3.5.2	系统聚类算法	99
3.6	关联分析	105
3.6.1	关联规则挖掘: Apriori 与 Eclat 算法	105
3.6.2	序列模式挖掘: SPADE 算法	119
第 4 章	特征构建技术	132
4.1	特征变换	133
4.1.1	概念分层	134
4.1.2	标准化	136
4.1.3	离散化	139
4.1.4	函数变换	142
4.1.5	深入表达	143
4.2	特征组合	143
4.2.1	基于特定的领域知识	143
4.2.2	二元组合	144

4.2.3	高阶多项式	148
4.3	自动生成：基于遗传编程的方法	152
4.3.1	基本思路	153
4.3.2	特征表达式	154
4.3.3	产生初始种群	162
4.3.4	计算适应度	163
4.3.5	选择、交叉和变异	165
4.3.6	实例分析	171
第 5 章	特征选择方法	176
5.1	直接法	177
5.2	单元法	177
5.2.1	Pearson 相关系数	177
5.2.2	距离相关系数	179
5.2.3	单因素方差分析	181
5.2.4	信息增益	184
5.2.5	卡方检验	191
5.2.6	Gini 系数	194
5.3	多元法	198
5.3.1	逐步回归	198
5.3.2	随机森林	203
5.3.3	遗传算法	210
第 6 章	模型参数优化	213
6.1	交叉验证	214
6.2	网格搜索	215
6.3	遗传算法	217
6.3.1	基本概念	217
6.3.2	遗传算法算例	218
6.3.3	遗传算法实现步骤	223
6.3.4	遗传算法 R 语言实现	223
6.3.5	R 语言 mcga 和 genalg 包的应用	228
6.4	粒子群优化	233
6.4.1	基本概念及原理	233
6.4.2	粒子群算法 R 语言实现	235
6.4.3	粒子群算法实现步骤	238

6.4.4	R 语言 pso 包的应用	239
6.5	模拟退火	241
6.5.1	基本概念及原理	241
6.5.2	模拟退火算法 R 语言实现	242
6.5.3	模拟退火算法实现步骤	244
6.5.4	R 语言 GenSA 和 stats 包的应用	245
第 7 章	预测效果评估	250
7.1	概率预测评估方法	251
7.1.1	混淆矩阵	251
7.1.2	ROC 曲线	255
7.1.3	KS 曲线	261
7.1.4	累计收益图	263
7.1.5	累计提升图	264
7.1.6	累计响应图	266
7.2	数值预测评估方法	267
7.2.1	常见评估指标	267
7.2.2	ASD 累计收益图	270

第二部分 预测算法篇

第 8 章	线性回归及其优化	274
8.1	多元线性回归	275
8.1.1	回归模型和基本假定	275
8.1.2	最小二乘估计	276
8.1.3	回归方程和回归系数的显著性检验	276
8.1.4	多重共线性	277
8.2	Ridge 回归	280
8.2.1	基本概念	281
8.2.2	岭迹曲线	281
8.2.3	基于 GCV 准则确定岭参数	283
8.2.4	Ridge 回归的 R 语言实现	284
8.3	Lasso 回归	285
8.3.1	基本概念	285
8.3.2	使用 LAR 求解 Lasso	286
8.3.3	Lasso 算法的 R 语言实现	288

8.3.4	R 语言 lars 包的应用	290
8.4	分位数回归	292
8.4.1	基本概念	292
8.4.2	分位数回归的计算	294
8.4.3	用单纯形法求解分位数回归及 R 语言实现	296
8.4.4	R 语言 quantreg 包的应用	298
8.5	稳健回归	300
8.5.1	基本概念	301
8.5.2	M-估计法及其 R 语言实现	301
8.5.3	应用 R 语言 MASS 包实现稳健回归	304
第 9 章	复杂回归分析	307
9.1	梯度提升回归树 (GBRT)	308
9.1.1	Boosting 方法简介	308
9.1.2	AdaBoost 算法	308
9.1.3	提升回归树算法	311
9.1.4	梯度提升	312
9.1.5	GBRT 的 R 语言实现	314
9.1.6	R 语言 gbm 包的应用	316
9.2	神经网络	320
9.2.1	基本概念	320
9.2.2	单层感知器学习算法	322
9.2.3	SLP 回归算法的 R 语言实现	323
9.2.4	BP 神经网络学习算法	325
9.2.5	BP 回归算法的 R 语言实现	327
9.2.6	RBF 神经网络学习算法	330
9.2.7	RBF 回归算法的 R 语言实现	332
9.2.8	Elman 神经网络学习算法	334
9.2.9	Elman 回归算法的 R 语言实现	336
9.2.10	使用 R 语言包构建神经网络	338
9.3	支持向量机回归	343
9.3.1	基本问题	344
9.3.2	LS-SVMR 算法	347
9.3.3	LS-SVMR 算法的 R 语言实现	348
9.4	高斯过程回归	349
9.4.1	GPR 算法	350

9.4.2	GPR 算法的 R 语言实现	352
9.4.3	R 语言 kernlab 包的应用	355
第 10 章	时间序列分析	358
10.1	Box-Jenkins 方法	359
10.1.1	p 阶自回归模型	359
10.1.2	q 阶移动平均模型	361
10.1.3	自回归移动平均模型	363
10.1.4	ARIMA 模型	365
10.1.5	ARIMA 模型的 R 语言实现	367
10.1.6	R 语言 forecast 包的应用	373
10.2	门限自回归模型	376
10.2.1	TAR 模型的基本原理	376
10.2.2	TAR 模型的 R 语言实现	377
10.2.3	R 语言 TSA 包的应用	380
10.3	GARCH 模型族	382
10.3.1	线性 ARCH 模型	382
10.3.2	GRACH 模型	383
10.3.3	EGARCH 模型	384
10.3.4	Power ARCH 模型	384
10.3.5	PARCH 模型的 R 语言实现	385
10.3.6	R 语言 fGarch 包的应用	395
10.4	向量自回归模型	398
10.4.1	VAR 模型基本原理	398
10.4.2	VAR 模型的 R 语言实现	399
10.4.3	R 语言 vars 包的应用	403
10.5	卡尔曼滤波器算法	405
10.5.1	Kalman 滤波算法初步	406
10.5.2	Kalman 滤波的 R 语言实现	407
10.5.3	R 语言 FKF 包的应用	409

第三部分 预测应用篇

第 11 章	短期日负荷曲线预测	414
11.1	电力行业负荷预测介绍	415
11.2	短期日负荷曲线预测的基本要求	415