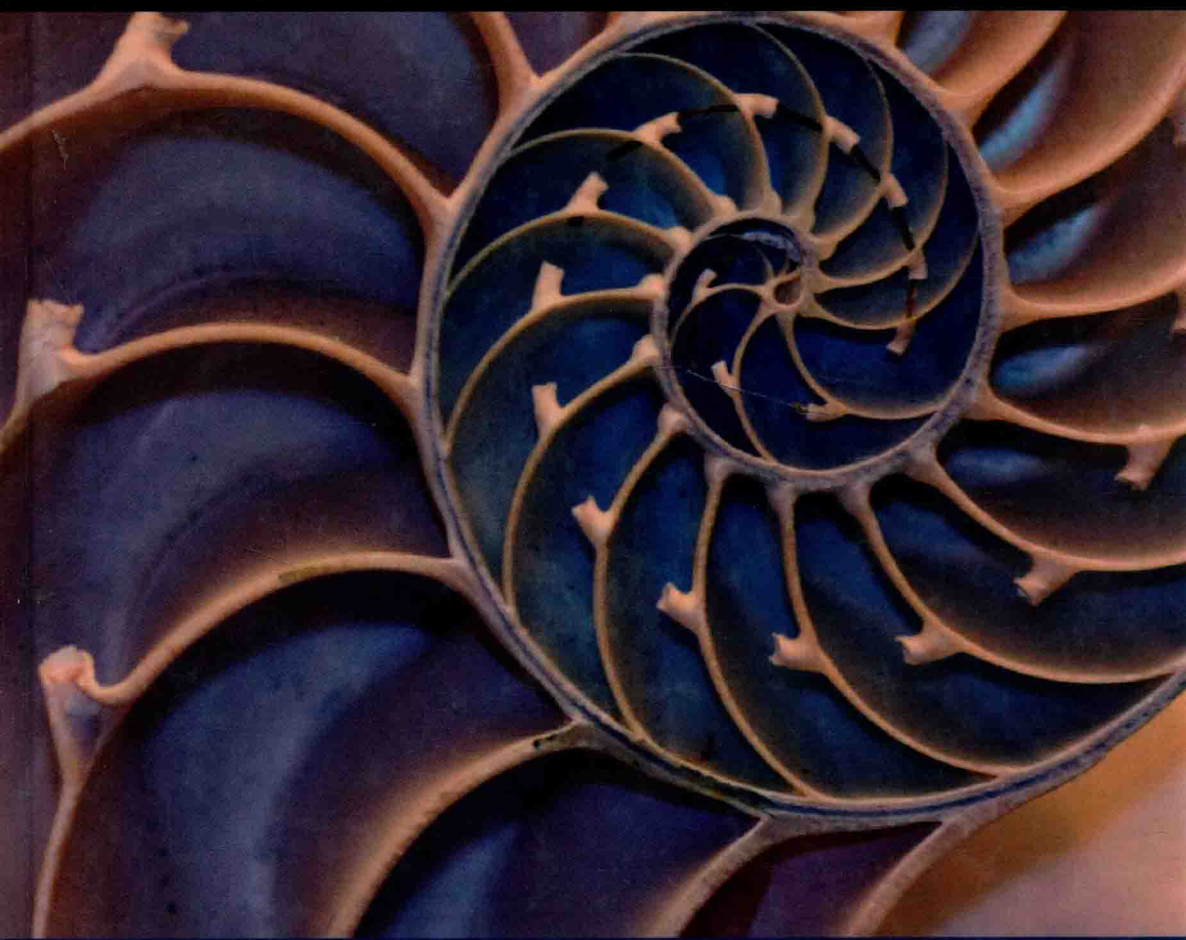
DATA MINING FOR BUSINESS ANALYTICS

CONCEPTS, TECHNIQUES, AND APPLICATIONS WITH JMP PRO®



Galit Shmueli • Peter C. Bruce • Mia L. Stephens • Nitin R. Patel

with website

WILEY

DATA MINING FOR BUSINESS ANALYTICS

Concepts, Techniques, and Applications with JMP Pro®

GALIT SHMUELI
PETER C. BRUCE
MIA L. STEPHENS
NITIN R. PATEL

WILEY

Copyright 2017 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Names: Shmueli, Galit, 1971- author. | Bruce, Peter C., 1953- author. | Stephens, Mia L., author. | Patel, Nitin R. (Nitin Ratilal), author.
Title: Data mining for business analytics : concepts, techniques, and applications with JMP Pro / Galit Shmueli, Peter C. Bruce, Mia L. Stephens, Nitin R. Patel.
Description: Hoboken, New Jersey : John Wiley & Sons, 2017. | Includes index.
Identifiers: LCCN 2015048305| ISBN 9781118877432 (cloth) | ISBN 9781118877524 (epub)
Subjects: LCSH: Business mathematics–Computer programs. | Business–Data processing. | JMP (Computer file) | Data mining.
Classification: LCC HF5691 .S43245 2016 | DDC 006.3/12–dc23 LC record available at http://lccn.loc.gov/2015048305

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

DATA MINING FOR BUSINESS ANALYTICS

To our families

Boaz and Noa
Liz, Lisa, and Allison
Michael, Madi, Olivia, and in memory of E.C. Jr.
Tehmi, Arjun, and in memory of Aneesh

试读结束：需要全本请在线购买： www.ertongbook.com

FOREWORD

No matter what your chosen profession or place of work, your future will almost certainly be saturated with data. The modern world is defined by the bits of data pulsing from billions of keyboards and trillions of card swipes—emanating from every manner of electronic device and system—transmitted instantaneously around the globe. The sheer amount of data is measured in volumes difficult to comprehend. But it's not about how much data you have; it's what you do with it, and how quickly, that counts most. Grappling with this messy world of data and putting it to good use will be key to productive and well-functioning organizations and successful managerial careers, not just in the obvious places circling Silicon Valley such as Google and Facebook but in insurance companies, banks, auto manufacturers, airlines, hospitals, and indeed nearly everywhere.

That's where *Data Mining for Business Analytics: Concepts, Techniques and Applications with JMP Pro®* can help. Professor Shmueli and her coauthors provide a very useful guide for students of business to learn the important concepts and methods for navigating complex datasets. Born out of the authors' years of experience teaching the subject, the book has evolved from earlier editions to keep pace with the changing landscape of business analytics in graduate and undergraduate education. Most important, new with this edition is the integration of JMP Pro®, a statistical tool from SAS Institute, which is provided as the vehicle for working with data in problem sets. Learning analytics is ultimately about doing things to and with data to generate insights. Mastering one's dexterity with powerful statistical tools is a necessary and critical step in the learning process.

If you've set your sights on leading in a digital world, this book is a great place to start preparing yourself for the future.

Michael Rappa
Institute for Advanced Analytics
North Carolina State University

PREFACE

The textbook *Data Mining for Business Intelligence* first appeared in early 2007. Since then, it has been used by numerous practitioners and in many courses, ranging from dedicated data mining classes to more general business analytics courses (including our own experience teaching this material both online and in person for more than 10 years). Following feedback from instructors teaching MBA, undergraduate, and executive courses, and from students, the second edition saw revisions to some of the existing chapters and included two new topics: data visualization and time series forecasting.

This book is the first edition to fully integrate JMP Pro®[1] rather than the Microsoft Office Excel add-in, XLMiner. JMP Pro® is a desktop statistical package from SAS Institute that runs natively on Mac and Windows machines. All examples, special topics boxes, instructions, and exercises presented in this book are based on JMP 12 Pro, the professional version of JMP, which has a rich array of built- in tools for interactive data visualization, analysis, and modeling.[2]

There are other important changes in this edition. The first noticeable change is the title: other than the addition of JMP Pro®, we now use *Business Analytics* in place of *Business Intelligence*. This update reflects the change in terminology since the second edition: BI today refers mainly to reporting and data visualization (what is happening now), while BA has taken over the advanced analytics, which include predictive analytics and data mining. In this new edition we therefore also updated these terms in the book, using them as is currently common.

We added a new chapter, *Combining Methods: Ensembles and Uplift Modeling* (Chapter 13). This chapter, which is the last in Part IV on Prediction and Classification Methods, introduces two important approaches. The first—ensembles—is the combination of multiple models for improving predictive power. Ensembles have routinely proved their

[1] JMP Pro®, Version 12. SAS Institute Inc., Cary, NC 27513. See Chapter 1 for information on how to get JMP Pro®.

[2] Relevant new features in JMP Pro 13 are noted in the chapters.

usefulness in practical applications and in data mining contests. The second topic—uplift modeling—introduces an improved approach for measuring the impact of an intervention or treatment. Similar to other chapters, this new chapter includes real-world examples and end-of-chapter problems.

Other changes include the addition of two new cases based on real data (one on political persuasion and uplift modeling, and another on taxi cancellations), and the removal of one chapter, Association Rules (association rules is a feature not available in JMP 12 Pro, but will be a new feature in JMP 13 Pro).

Since the second edition's appearance, the landscape of courses using the textbook has greatly expanded: whereas initially the book was used mainly in semester-long elective MBA-level courses, it is now used in a variety of courses in Business Analytics degree and certificate programs, ranging from undergraduate programs, to post-graduate and executive education programs. Courses in such programs also vary in their duration and coverage. In many cases, our book is used across multiple courses. The book is designed to continue supporting the general Predictive Analytics or Data Mining course as well as supporting a set of courses in dedicated business analytics programs.

A general "Business Analytics," "Predictive Analytics," or Data Mining course, common in MBA and undergraduate programs as a one-semester elective, would cover Parts I–III, and choose a subset of methods from Parts IV and V. Instructors can choose to use cases as team assignments, class discussions, or projects. For a two-semester course, Part VI might be considered. For a set of courses in a dedicated Business Analytics program, here are a few courses that have been using the second edition of *Data Mining for Business Intelligence*:

Predictive Analytics: Supervised Learning In a dedicated Business Analytics program, the topic of Predictive Analytics is typically instructed across a set of courses. The first course would cover Parts I–IV and instructors typically choose a subset of methods from Part IV according to the course length. We recommend including the new Chapter 13 in such a course.

Predictive Analytics: Unsupervised Learning This course introduces data exploration and visualization, dimension reduction, mining relationships, and clustering (Parts III and V). If this course follows the Predictive Analytics: Supervised Learning course, then it is useful to examine examples and approaches that integrate unsupervised and supervised learning.

Forecasting Analytics A dedicated course on time series forecasting would rely on Part VI.

In all courses, we strongly recommend including a project component, where data are either collected by students according to their interest or provided by the instructor (e.g., from the many data mining competition datasets available). From our experience and the experience of other instructors, such projects enhance learning and provide students with an excellent opportunity to understand the strengths of data mining and the challenges that arise in working with data and solving real business problems.

ACKNOWLEDGMENTS

The authors thank the many people who assisted us in improving the first edition and improving it further in the second edition, and now in this JMP edition. Anthony Babinec, who has been using drafts of this book for years in his Data Mining courses at Statistics.com, provided us with detailed and expert corrections. The Statistics.com team has also provided valuable checking, trouble-shooting, and critiquing: Kuber Deokar, Instructional Operations Supervisor, and Shweta Jadhav and Dhanashree Vishwasrao, Assistant Teachers. We also thank the many students who have used and commented on earlier editions of this text at Statistics.com.

Similarly, Dan Toy and John Elder IV greeted our project with enthusiasm and provided detailed and useful comments on earlier drafts. Boaz Shmueli and Raquelle Azran gave detailed editorial comments and suggestions on the first two editions; Noa Shmueli provided edits on the new edition; Bruce McCullough and Adam Hughes did the same for the first edition. Ravi Bapna, who used an early draft in a Data Mining course at the Indian School of Business, provided invaluable comments and helpful suggestions. Useful comments and feedback have also come from the many instructors, too numerous to mention, who have used the book in their classes.

From the Smith School of Business at the University of Maryland, colleagues Shrivardhan Lele, Wolfgang Jank, and Paul Zantek provided practical advice and comments. We thank Robert Windle, and MBA students Timothy Roach, Pablo Macouzet, and Nathan Birckhead for invaluable datasets. We also thank MBA students Rob Whitener and Daniel Curtis for the heatmap and map charts. And we thank the many MBA students for fruitful discussions and interesting data mining projects that have helped shape and improve the book.

This book would not have seen the light of day without the nurturing support of the faculty at the Sloan School of Management at MIT. Our special thanks to Dimitris Bertsimas, James Orlin, Robert Freund, Roy Welsch, Gordon Kaufmann, and Gabriel Bitran. As teaching assistants for the data mining course at Sloan, Adam Mersereau gave detailed comments on the notes and cases that were the genesis of this book, Romy Shioda helped with the

preparation of several cases and exercises used here, and Mahesh Kumar helped with the material on clustering. We are grateful to the MBA students at Sloan for stimulating discussions in the class that led to refinement of the notes.

Chris Albright, Gregory Piatetsky-Shapiro, Wayne Winston, and Uday Karmarkar gave us helpful advice. Anand Bodapati provided both data and advice. Suresh Ankolekar and Mayank Shah helped develop several cases and provided valuable pedagogical comments. Vinni Bhandari helped write the Charles Book Club case.

We would like to thank Marvin Zelen, L. J. Wei, and Cyrus Mehta at Harvard, as well as Anil Gore at Pune University, for thought-provoking discussions on the relationship between statistics and data mining. Our thanks to Richard Larson of the Engineering Systems Division, MIT, for sparking many stimulating ideas on the role of data mining in modeling complex systems. They helped us develop a balanced philosophical perspective on the emerging field of data mining.

Our thanks to Ajay Sathe, and his Cytel colleagues who helped launch this project: Suresh Ankolekar, Poonam Baviskar, Kuber Deokar, Rupali Desai, YogeshGajjar, Ajit Ghanekar, Ayan Khare, Bharat Lande, Dipankar Mukhopadhyay, S. V.Sabnis, Usha Sathe, Anurag Srivastava, V. Subramaniam, Ramesh Raman, and Sanhita Yeolkar.

Steve Quigley at Wiley showed confidence in this book from the beginning, helped us navigate through the publishing process with great speed, and together with Curt Hinrichs's encouragement and support helped make this JMP Pro® edition possible. Jon Gurstelle, Allison McGinniss, Sari Friedman, and Katrina Maceda at Wiley, and Shikha Pahuja from Thomson Digital, were all helpful and responsive as we finalized this new JMP Pro® edition.

We also thank Catherine Plaisant at the University of Maryland's Human–Computer Interaction Lab, who helped out in a major way by contributing exercises and illustrations to the data visualization chapter, Marietta Tretter at Texas A&M for her helpful comments and thoughts on the time series chapters, and Stephen Few and Ben Shneiderman for feedback and suggestions on the data visualization chapter and overall design tips.

Gregory Piatetsky-Shapiro, founder of KDNuggets.com, has been generous with his time and counsel over the many years of this project. Ken Strasma, founder of the microtargeting firm HaystaqDNA and director of targeting for the 2004 Kerry campaign and the 2008 Obama campaign, provided the scenario and data for the section on uplift modeling.

Finally, we'd like to thank the reviewers of this first JMP Pro® edition for their feedback and suggestions, and members of the JMP Documentation, Education and Development teams, for their support, patience, and responsiveness to our endless questions and requests. We thank L. Allison Jones-Farmer, Maria Weese, Ian Cox, Di Michelson, Marie Gaudard, Curt Hinrichs, Rob Carver, Jim Grayson, Brady Brady, Jian Cao, Chris Gotwalt, and Fang Chen. Most important, we thank John Sall, whose inovatation, inspiration, and continued dedication to providing accessible and user-friendly desktop statistical software made JMP, and this book, possible.

CONTENTS

FOREWORD xvii

PREFACE xix

ACKNOWLEDGMENTS xxi

PART I PRELIMINARIES

1 Introduction 3

1.1 What Is Business Analytics? 3
Who Uses Predictive Analytics? 4
1.2 What Is Data Mining? 5
1.3 Data Mining and Related Terms 5
1.4 Big Data 6
1.5 Data Science 7
1.6 Why Are There So Many Different Methods? 7
1.7 Terminology and Notation 8
1.8 Roadmap to This Book 10
Order of Topics 11
Using JMP Pro, Statistical Discovery Software from SAS 11

2 Overview of the Data Mining Process 14

2.1 Introduction 14
2.2 Core Ideas in Data Mining 15
Classification 15
Prediction 15
Association Rules and Recommendation Systems 15

试读结束：需要全本请在线购买： www.ertongbook.com

Predictive Analytics 16
Data Reduction and Dimension Reduction 16
Data Exploration and Visualization 16
Supervised and Unsupervised Learning 16
2.3 The Steps in Data Mining 17
2.4 Preliminary Steps 19
Organization of Datasets 19
Sampling from a Database 19
Oversampling Rare Events in Classification Tasks 19
Preprocessing and Cleaning the Data 20
Changing Modeling Types in JMP 20
Standardizing Data in JMP 25
2.5 Predictive Power and Overfitting 25
Creation and Use of Data Partitions 25
Partitioning Data for Crossvalidation in JMP Pro 27
Overfitting 27
2.6 Building a Predictive Model with JMP Pro 29
Predicting Home Values in a Boston Neighborhood 29
Modeling Process 30
Setting the Random Seed in JMP 34
2.7 Using JMP Pro for Data Mining 38
2.8 Automating Data Mining Solutions 40
Data Mining Software Tools: the State of the Market by Herb Edelstein 41
Problems 44

PART II DATA EXPLORATION AND DIMENSION REDUCTION

3 Data Visualization **51**

3.1 Uses of Data Visualization 51
3.2 Data Examples 52
Example 1: Boston Housing Data 53
Example 2: Ridership on Amtrak Trains 53
3.3 Basic Charts: Bar Charts, Line Graphs, and Scatterplots 54
Using The JMP Graph Builder 54
Distribution Plots: Boxplots and Histograms 56
Tools for Data Visualization in JMP 59
Heatmaps (Color Maps and Cell Plots): Visualizing Correlations and Missing Values 59
3.4 Multidimensional Visualization 61
Adding Variables: Color, Size, Shape, Multiple Panels, and Animation 62
Manipulations: Rescaling, Aggregation and Hierarchies, Zooming, Filtering 65
Reference: Trend Lines and Labels 68
Adding Trendlines in the Graph Builder 69
Scaling Up: Large Datasets 70

Multivariate Plot: Parallel Coordinates Plot 71
Interactive Visualization 72
3.5 Specialized Visualizations 73
Visualizing Networked Data 74
Visualizing Hierarchical Data: More on Treemaps 75
Visualizing Geographical Data: Maps 76
3.6 Summary of Major Visualizations and Operations, According to Data Mining Goal 77
Prediction 77
Classification 78
Time Series Forecasting 78
Unsupervised Learning 79
Problems 79

4 Dimension Reduction **81**

4.1 Introduction 81
4.2 Curse of Dimensionality 82
4.3 Practical Considerations 82
Example 1: House Prices in Boston 82
4.4 Data Summaries 83
Summary Statistics 83
Tabulating Data (Pivot Tables) 85
4.5 Correlation Analysis 87
4.6 Reducing the Number of Categories in Categorical Variables 87
4.7 Converting a Categorical Variable to a Continuous Variable 90
4.8 Principal Components Analysis 90
Example 2: Breakfast Cereals 91
Principal Components 95
Normalizing the Data 97
Using Principal Components for Classification and Prediction 100
4.9 Dimension Reduction Using Regression Models 100
4.10 Dimension Reduction Using Classification and Regression Trees 100
Problems 101

PART III PERFORMANCE EVALUATION

5 Evaluating Predictive Performance **105**

5.1 Introduction 105
5.2 Evaluating Predictive Performance 106
Benchmark: The Average 106
Prediction Accuracy Measures 107
Comparing Training and Validation Performance 108
5.3 Judging Classifier Performance 109
Benchmark: The Naive Rule 109
Class Separation 109
The Classification Matrix 109

Using the Validation Data 111
Accuracy Measures 111
Propensities and Cutoff for Classification 112
Cutoff Values for Triage 112
Changing the Cutoff Values for a Confussion Matrix in JMP 114
Performance in Unequal Importance of Classes 115
False-Positive and False-Negative Rates 116
Asymmetric Misclassification Costs 116
Asymmetric Misclassification Costs in JMP 119
Generalization to More Than Two Classes 120
5.4 Judging Ranking Performance 120
Lift Curves 120
Beyond Two Classes 122
Lift Curves Incorporating Costs and Benefits 122
5.5 Oversampling 123
Oversampling the Training Set 126
Stratified Sampling and Oversampling in JMP 126
Evaluating Model Performance Using a Nonoversampled Validation Set 126
Evaluating Model Performance If Only Oversampled Validation Set Exists 127
Applying Sampling Weights in JMP 128
Problems 129

PART IV PREDICTION AND CLASSIFICATION METHODS

6 Multiple Linear Regression **133**

6.1 Introduction 133
6.2 Explanatory versus Predictive Modeling 134
6.3 Estimating the Regression Equation and Prediction 135
Example: Predicting the Price of Used Toyota Corolla Automobiles 136
Coding of Categorical Variables in Regression 138
Additional Options for Regression Models in JMP 140
6.4 Variable Selection in Linear Regression 141
Reducing the Number of Predictors 141
How to Reduce the Number of Predictors 142
Manual Variable Selection 142
Automated Variable Selection 142
Coding of Categorical Variables in Stepwise Regression 143
Working with the All Possible Models Output 145
When Using a Stopping Algorithm in JMP 147
Other Regression Procedures in JMP Pro—Generalized Regression 149
Problems 150

7 *k*-Nearest Neighbors (*k*-NN) 155

7.1 The *k*-NN Classifier (Categorical Outcome) 155
Determining Neighbors 155
Classification Rule 156
Example: Riding Mowers 156
Choosing *k* 157
k Nearest Neighbors in JMP Pro 158
The Cutoff Value for Classification 159
k-NN Predictions and Prediction Formulas in JMP Pro 161
k-NN with More Than Two Classes 161
7.2 *k*-NN for a Numerical Response 161
Pandora 161
7.3 Advantages and Shortcomings of *k*-NN Algorithms 163
Problems 164

8 The Naive Bayes Classifier 167

8.1 Introduction 167
Naive Bayes Method 167
Cutoff Probability Method 168
Conditional Probability 168
Example 1: Predicting Fraudulent Financial Reporting 168
8.2 Applying the Full (Exact) Bayesian Classifier 169
Using the "Assign to the Most Probable Class" Method 169
Using the Cutoff Probability Method 169
Practical Difficulty with the Complete (Exact) Bayes Procedure 170
Solution: Naive Bayes 170
Example 2: Predicting Fraudulent Financial Reports, Two Predictors 172
Using the JMP Naive Bayes Add-in 174
Example 3: Predicting Delayed Flights 174
8.3 Advantages and Shortcomings of the Naive Bayes Classifier 179
Spam Filtering 179
Problems 180

9 Classification and Regression Trees 183

9.1 Introduction 183
9.2 Classification Trees 184
Recursive Partitioning 184
Example 1: Riding Mowers 185
Categorical Predictors 186
9.3 Growing a Tree 187
Growing a Tree Example 187
Classifying a New Observation 188
Fitting Classification Trees in JMP Pro 191
Growing a Tree with CART 192

9.4 Evaluating the Performance of a Classification Tree 192
Example 2: Acceptance of Personal Loan 192
9.5 Avoiding Overfitting 193
Stopping Tree Growth: CHAID 194
Growing a Full Tree and Pruning It Back 194
How JMP Limits Tree Size 196
9.6 Classification Rules from Trees 196
9.7 Classification Trees for More Than Two Classes 198
9.8 Regression Trees 199
Prediction 199
Evaluating Performance 200
9.9 Advantages and Weaknesses of a Tree 200
9.10 Improving Prediction: Multiple Trees 204
Fitting Ensemble Tree Models in JMP Pro 206
9.11 CART and Measures of Impurity 207
Problems 207

10 Logistic Regression **211**

10.1 Introduction 211
Logistic Regression and Consumer Choice Theory 212
10.2 The Logistic Regression Model 213
Example: Acceptance of Personal Loan (Universal Bank) 214
Indicator (Dummy) Variables in JMP 216
Model with a Single Predictor 216
Fitting One Predictor Logistic Models in JMP 218
Estimating the Logistic Model from Data: Multiple Predictors 218
Fitting Logistic Models in JMP with More Than One Predictor 221
10.3 Evaluating Classification Performance 221
Variable Selection 222
10.4 Example of Complete Analysis: Predicting Delayed Flights 223
Data Preprocessing 225
Model Fitting, Estimation and Interpretation—A Simple Model 226
Model Fitting, Estimation and Interpretation—The Full Model 227
Model Performance 229
Variable Selection 230
Regrouping and Recoding Variables in JMP 232
10.5 Appendixes: Logistic Regression for Profiling 234
Appendix A: Why Linear Regression Is Problematic for a Categorical Response 234
Appendix B: Evaluating Explanatory Power 236
Appendix C: Logistic Regression for More Than Two Classes 238
Nominal Classes 238
Problems 241

11 Neural Nets **245**

11.1 Introduction 245
11.2 Concept and Structure of a Neural Network 246

11.3 Fitting a Network to Data 246
Example 1: Tiny Dataset 246
Computing Output of Nodes 248
Preprocessing the Data 251
Activation Functions and Data Processing Features in JMP Pro 251
Training the Model 251
Fitting a Neural Network in JMP Pro 254
Using the Output for Prediction and Classification 256
Example 2: Classifying Accident Severity 258
Avoiding overfitting 259
11.4 User Input in JMP Pro 260
Unsupervised Feature Extraction and Deep Learning 263
11.5 Exploring the Relationship between Predictors and Response 264
Understanding Neural Models in JMP Pro 264
11.6 Advantages and Weaknesses of Neural Networks 264
Problems 265

12 Discriminant Analysis 268

12.1 Introduction 268
Example 1: Riding Mowers 269
Example 2: Personal Loan Acceptance (Universal Bank) 269
12.2 Distance of an Observation from a Class 270
12.3 From Distances to Propensities and Classifications 272
Linear Discriminant Analysis in JMP 275
12.4 Classification Performance of Discriminant Analysis 275
12.5 Prior Probabilities 277
12.6 Classifying More Than Two Classes 278
Example 3: Medical Dispatch to Accident Scenes 278
Using Categorical Predictors in Discriminant Analysis in JMP 279
12.7 Advantages and Weaknesses 280
Problems 282

13 Combining Methods: Ensembles and Uplift Modeling 285

13.1 Ensembles 285
Why Ensembles Can Improve Predictive Power 286
The Wisdom of Crowds 287
Simple Averaging 287
Bagging 288
Boosting 288
Creating Ensemble Models in JMP Pro 289
Advantages and Weaknesses of Ensembles 289
13.2 Uplift (Persuasion) Modeling 290
A-B Testing 290
Uplift 290
Gathering the Data 291
A Simple Model 292
Modeling Individual Uplift 293

试读结束：需要全本请在线购买： www.ertongbook.com