

Contents

About the Authors	xvii
Foreword	xix
Preface	xxi
Acknowledgements	xxv
Introduction	1
Development of Financial Data Analytics	1
Organization of the Book	5
References	7
PART ONE	
<hr/>	
Data Cleansing and Analytical Models	
CHAPTER 1	
Mathematical and Statistical Preliminaries	11
1.1 Random Vector	12
1.2 Matrix Theory	16
1.3 Vectors and Matrix Norms	23
1.4 Common Probability Distributions	24
1.5 Introductory Bayesian Statistics	30
References	40
CHAPTER 2	
Introduction to Python and R	41
2.1 What is Python?	41
2.2 What is R?	42
2.3 Package Management in Python and R	42
2.4 Basic Operations in Python and R	44
2.5 One-Way ANOVA and Tukey's HSD for Stock Market Indices	49
References	64

CHAPTER 3		
Statistical Diagnostics of Financial Data		67
3.1 Normality Assumption for Relative Stock Price Changes		67
3.2 Student's t_v -distribution for Stock Price Changes		76
3.3 Testing for Multivariate Normality		81
3.4 Sample Correlation Matrix		84
3.5 Empirical Properties of Stock Prices		86
3.A Appendix		93
References		97
CHAPTER 4		
Financial Forensics		99
4.1 Benford's Law		99
4.2 Scaling Invariance and Benford's Law		101
4.3 Benford's Law in Business Reports		104
4.4 Benford's Law in Growth Figures		117
4.5 Zipf's Law		125
4.6 Zipf's Law and COVID-19 Figures		127
4.A Appendix		132
References		136
CHAPTER 5		
Numerical Finance		139
5.1 Fundamentals of Simulation		139
5.2 Variance Reduction Technique		146
5.3 A Review of Financial Calculus and Derivative Pricing		158
*5.4 Greeks and their Approximations		179
References		199
CHAPTER 6		
Approximation for Model Inference		201
6.1 EM Algorithm		201
6.2 MM Algorithm		216
*6.3 A Short Course on the Theory of Markov Chains		222
*6.4 Markov Chain Monte Carlo		236
*6.A Appendix		261
References		268
CHAPTER 7		
Time-Varying Volatility Matrix and Kelly Fraction		271
7.1 Fluctuation of Volatilities		271
7.2 Exponentially Weighted Moving Average		275
7.3 ARIMA Time Series Model		277
7.4 ARCH and GARCH Models		291
*7.5 Kelly Fraction		317

7.6	Calendar Effects	330
*7.A	Appendix	335
	References	343

CHAPTER 8

	Risk Measures, Extreme Values, and Copulae	345
8.1	Value-at-Risk and Expected Shortfall	345
8.2	Basel Accords and Risk Measures	348
8.3	Historical Simulation (Bootstrapping)	350
8.4	Statistical Model Building Approach	354
8.5	Use of Extreme Value Theory	356
8.6	Backtesting	359
8.7	Estimates of Expected Shortfall	364
8.8	Dependence Modelling via Copulae	369
*8.A	Appendix	402
	References	404

PART TWO**Linear Models****CHAPTER 9**

	Principal Component Analysis and Recommender Systems	409
9.1	US Zero-Coupon Rates	409
9.2	PCA Algorithm	411
9.3	Financial Interpretation of PCs for US Zero-Coupon Rates	417
9.4	PCA as an Eigenvalue Problem	421
9.5	Factor Models via PCA	422
9.6	Value-at-Risk via PCA	424
9.7	Portfolio Immunization	427
9.8	Facial Recognition via PCA	430
9.9	Non-Life Insurance via PCA	439
9.10	Investment Strategies using PCA	442
*9.11	Recommender System	447
*9.A	Appendix	456
	References	465

CHAPTER 10

	Regression Learning	467
10.1	Simple and Multiple Linear Regression Models and Beyond	467
10.2	Polynomial Regression	473
10.3	Generalized Linear Models	478
10.4	Logistic Regression	484
10.5	Poisson Regression	497
10.6	Model Evaluation and Considerations in Practice	501

*10.7	Principal Component Regression	510
*10.A	Appendix	518
	References	522
CHAPTER 11		
	Linear Classifiers	525
11.1	Perceptron	526
11.2	Support Vector Machine	533
*11.A	Appendix	545
	References	567
PART THREE		
	Nonlinear Models	
CHAPTER 12		
	Bayesian Learning	571
12.1	Simple Credibility Theory	571
*12.2	Bayesian Asymptotic Inference	573
12.3	Revisiting Polynomial Regression	575
12.4	Bayesian Classifiers	578
12.5	Comonotone-Independence Bayes Classifier (CIBer)	580
12.A	Appendix	609
	References	612
CHAPTER 13		
	Classification and Regression Trees, and Random Forests	613
13.1	Classification (Decision) Trees	613
*13.2	Concepts of Entropies	615
13.3	Information Gain	623
13.4	Other Impurity Measures for Information	626
13.5	Splitting Against Continuous Attributes	629
13.6	Overfitting in Classification Tree	630
13.7	Classification Trees in Python and R	633
13.8	Regression Trees	641
13.9	Random Forest	649
13.A	Appendix	654
	References	659
CHAPTER 14		
	Cluster Analysis	661
14.1	K-Means Clustering	661
14.2	K-Nearest Neighbour	694
*14.3	Kernel Regression	703
*14.A	Appendix	714
	References	725

CHAPTER 15	
Applications of Deep Learning in Finance	727
15.1 Human Brains and Artificial Neurons	727
15.2 Feedforward Network	729
15.3 ANN with Linear Outputs	730
15.4 ANN with Logistic Outputs	737
15.5 Adaptive Learning Rate	740
15.6 Training Neural Networks via Backpropagation	742
15.7 Multilayer Perceptron	746
15.8 Universal Approximation Theorem	752
15.9 Long Short-Term Memory (LSTM)	754
References	764
Postlude	767
Index	769