

Preface	vii
Fundamental Concepts	
Introduction	3
Why the Term Machine Learning?	3
Why Not Just Call It Statistics?	4
Why Do We Need Machine Learning?	5
Machine Learning Thesis	5
How Is This Book Different?	6
Prerequisites for the Book	7
Definitions	8
Model vs. Algorithm	8
Prediction	8
Software	9
Datasets	10
Grit	10
National Survey on Drug Use and Health from 2014	10
Early Childhood Learning Study-Kindergarten Cohort	11
Big Five Inventory	11
Holzinger-Swineford	12
PHE Exposure	12
Professor Ratings	12
The Principles of Machine Learning Research	13
Key Terminology	13
Overview	14
Principle #1; Machine Learning Is Not Just Lazy Induction	15
Complexity	17
Abduction	18
Principle #2: Orienting Our Goals Relative to Prediction, Explanation, and Description	20
Principle #3: Labeling a Study as Exploratory or Confirmatory Is Too Simplistic	25
Model Size	27
Level of Hypothesis	28
Example	29
Types of Relationships	33
Exploratory Data Analysis	35
Principle #4: Report Everything	36
Summary	38
Further Reading	39
The Practices of Machine Learning	40
Key Terminology	40

Comparing Algorithms and Models	41
Model Fit	42
Regression	42
Bias-Variance Trade-Off	47
Resampling	51
k-Fold CV	52
Nested CV	53
Bootstrap Sampling	55
Recommendations	56
Classification	57
Receiver Operating Characteristic (ROC) Curves	64
Imbalanced Outcomes	66
Sampling	67
Conclusion	70
Further Reading	71
Computational Time and Resources	71
Algorithms for Univariate Outcomes	
Regularized Regression	75
Key Terminology	75
Linear Regression	76
Logistic Regression	77
Motivating Example	77
The Logistic Model	77
Regularization	80
Regularization Formulation	81
Choosing a Final Model	85
Rationale for Regularization	87
Bias and Variance	89
Alternative Forms of Regularization	92
Lasso P-Values	92
Stability of Selection	93
Interactions	95
Group Regularization	98
Bayesian Regression	101
Summary	103
Further Reading	103
Computational Time and Resources	104
Decision Trees	105
Key Terminology	105
Introduction	106

Example 1	107
Describing the Tree	112
Example 2	113
Decision Tree Algorithms	119
CART	119
Pruning	120
Conditional Inference Trees	122
Miscellaneous Topics	123
Interactions	123
Pathways	125
Stability	126
Missing Data	131
Variable Importance	132
Summary	133
Further Reading	134
Computational Time and Resources	134
Ensembles	136
Key Terminology	136
Bagging	137
Random Forests	144
Gradient Boosting	148
Variants on Boosting	150
Interpretation	151
Global Interpretation	151
Local Interpretation	155
Empirical Example	157
Important Notes	163
Interactions	163
Other Types of Ensembles	163
Algorithm Comparison More Broadly	164
Summary	165
Further Reading	166
Computational Time and Resources	167
Algorithms for Multivariate Outcomes	
Machine Learning and Measurement	171
Key Terminology	171
Defining Measurement Error	172
Impact of Measurement Error	173
Attenuation	173
Stability	174

Predictive Performance	175
Assessing Measurement Error	178
Indexes of Reliability	178
Factor Analysis	179
Factor Analysis-Based Reliability	184
Factor Scores	185
Factor Score Validity and Reliability	185
Weighting	186
Item Reliability	187
Items versus Scales	188
Alternative Methods	190
Principal Components	190
Variable Network Models	192
Summary	194
Further Reading	195
Computational Time and Resources	195
Machine Learning and Structural Equation Modeling	197
Key Terminology	198
Latent Variables as Predictors	199
Example	200
Application to SEM	202
Drawbacks	203
Predicting Latent Variables	204
Multiple Indicator Multiple Cause (MIMIC) Models	204
Test Performance in SEM	206
Using Latent Variables as Outcomes and Predictors	207
Can Regularization Improve Generalizability in SEM?	209
Regularized SEM	210
Exploratory SEM	213
Nonlinear Relationships and Latent Variables	215
Bayesian SEM	216
Demonstration	216
Summary	221
Further Reading	221
Computational Time and Resources	223
Machine Learning with Mixed-Effects Models	224
Key Terminology	224
Mixed-Effects Models	225
Machine Learning with Clustered Data	227
Recursive Partitioning with Mixed-Effects Models	227

Algorithm	229
Regularization with Mixed-Effects Models	230
Illustrative Example	231
Recursive Partitioning	232
Glmertree	236
REEMtree	237
Regularization	240
Additional Strategies for Mining Longitudinal Data	242
Piecewise Spline Models	242
Generalized Additive Models	243
Summary	244
Further Reading	245
Computational Time and Resources	246
Searching for Groups	247
Key Terminology	247
Finite Mixture Model	249
Search Procedure and Evaluation	251
Illustrative Example	254
Factor Mixture Models	257
Incorporating Covariates	260
Structural Equation Model Trees	261
Demonstration	262
Algorithm Details	264
Focused Assessment	265
SEM Forests	267
Summary	269
Further Reading	272
Computational Time and Resources	273
Alternative Data Types	
Introduction to Text Mining	277
Key Terminology	277
Data	278
Descriptives	278
Basic Text Mining	280
Text Tokenization	281
Text Data Preprocessing	285
Extracting Text Information	286
Text Tokenization	288
Data Matrix Representation	290
Basic Analysis of the Teaching Comment Data	292

Word Frequency	293
Stop Words	293
n-gram Analysis	298
Sentiment Analysis	302
Simple Dictionary-Based Sentiment Analysis	303
Word Sentiment	303
Conducting Sentiment Analysis	306
Sentiment Analysis of Teaching Evaluation Text Data	307
Sentiment Analysis with Valence Shifters	312
Summary and Discussion	314
Topic Models	316
Latent Dirichlet Allocation	317
Topic Modeling of the Teaching Evaluation Data	318
Summary	324
Further Reading	325
Computational Time and Resources	326
Introduction to Social Network Analysis	328
Key Terminology	328
Data	330
Data Collection	330
Network Data Structure	331
Network Visualization	334
Heat Map	334
Network Plot	336
Network Statistics	339
Network Statistics for a Whole Network	339
Network Statistics for Nodes	339
Network Statistics for Dyads	341
Basic Network Analysis	342
Network Modeling	346
Erdos-Rényi Random Graph Model	346
Stochastic Block Model	347
Exponential Random Graph Model	349
Terms Related to the Network	351
Terms Related to Node Covariates	352
Terms Related to the Edge Covariates	354
Latent Space Model	354
Summary	357
Further Reading	357
Computational Time and Resources	357

References	359
Author Index	395
Subject Index	404
About the Authors	415

Table of Contents provided by Blackwell's Book Services and R.R. Bowker. Used with permission.