
Contents

Preface		xiii
	Structure of the Book	xiv
	How to Use this Book	xv
	Online Resources	xvi
	Acknowledgements	xvi
Notation		xix
	Notational Conventions	xix
	Notational Conventions for Probabilities	xxi
1	Machine Learning for Predictive Data Analytics	1
	1.1 What Is Predictive Data Analytics?	1
	1.2 What Is Machine Learning?	3
	1.3 How Does Machine Learning Work?	5
	1.4 What Can Go Wrong with Machine Learning?	11
	1.5 The Predictive Data Analytics Project Lifecycle: CRISP-DM	12
	1.6 Predictive Data Analytics Tools	15
	1.7 The Road Ahead	17
	1.8 Exercises	19
2	Data to Insights to Decisions	21
	2.1 Converting Business Problems into Analytics Solutions	21
	2.1.1 Case Study: Motor Insurance Fraud	23
	2.2 Assessing Feasibility	24
	2.2.1 Case Study: Motor Insurance Fraud	26
	2.3 Designing the Analytics Base Table	27
	2.3.1 Case Study: Motor Insurance Fraud	31
	2.4 Designing and Implementing Features	32
	2.4.1 Different Types of Data	34
	2.4.2 Different Types of Features	34
	2.4.3 Handling Time	37
	2.4.4 Legal Issues	41
	2.4.5 Implementing Features	43
	2.4.6 Case Study: Motor Insurance Fraud	43
	2.5 Summary	46
	2.6 Further Reading	49
	2.7 Exercises	50

3	Data Exploration	55
3.1	The Data Quality Report	56
3.1.1	Case Study: Motor Insurance Fraud	57
3.2	Getting to Know the Data	61
3.2.1	The Normal Distribution	64
3.2.2	Case Study: Motor Insurance Fraud	65
3.3	Identifying Data Quality Issues	66
3.3.1	Missing Values	67
3.3.2	Irregular Cardinality	68
3.3.3	Outliers	69
3.3.4	Case Study: Motor Insurance Fraud	70
3.4	Handling Data Quality Issues	73
3.4.1	Handling Missing Values	73
3.4.2	Handling Outliers	74
3.4.3	Case Study: Motor Insurance Fraud	76
3.5	Advanced Data Exploration	77
3.5.1	Visualizing Relationships Between Features	77
3.5.2	Measuring Covariance and Correlation	86
3.6	Data Preparation	92
3.6.1	Normalization	92
3.6.2	Binning	94
3.6.3	Sampling	98
3.7	Summary	100
3.8	Further Reading	101
3.9	Exercises	103
4	Information-based Learning	117
4.1	Big Idea	117
4.2	Fundamentals	120
4.2.1	Decision Trees	121
4.2.2	Shannon's Entropy Model	124
4.2.3	Information Gain	128
4.3	Standard Approach: The ID3 Algorithm	134
4.3.1	A Worked Example: Predicting Vegetation Distributions	137
4.4	Extensions and Variations	144
4.4.1	Alternative Feature Selection and Impurity Metrics	144

4.4.2	Handling Continuous Descriptive Features	150
4.4.3	Predicting Continuous Targets	153
4.4.4	Tree Pruning	158
4.4.5	Model Ensembles	163
4.5	Summary	167
4.6	Further Reading	169
4.7	Exercises	170
5	Similarity-based Learning	179
5.1	Big Idea	179
5.2	Fundamentals	180
5.2.1	Feature Space	181
5.2.2	Measuring Similarity Using Distance Metrics	183
5.3	Standard Approach: The Nearest Neighbor Algorithm	186
5.3.1	A Worked Example	186
5.4	Extensions and Variations	190
5.4.1	Handling Noisy Data	190
5.4.2	Efficient Memory Search	195
5.4.3	Data Normalization	204
5.4.4	Predicting Continuous Targets	209
5.4.5	Other Measures of Similarity	212
5.4.6	Feature Selection	226
5.5	Summary	234
5.6	Further Reading	237
5.7	Epilogue	238
5.8	Exercises	240
6	Probability-based Learning	247
6.1	Big Idea	247
6.2	Fundamentals	249
6.2.1	Bayes' Theorem	252
6.2.2	Bayesian Prediction	256
6.2.3	Conditional Independence and Factorization	262
6.3	Standard Approach: The Naive Bayes Model	267
6.3.1	A Worked Example	269
6.4	Extensions and Variations	272
6.4.1	Smoothing	272

	6.4.2	Continuous Features: Probability Density Functions	276
	6.4.3	Continuous Features: Binning	289
	6.4.4	Bayesian Networks	292
	6.5	Summary	312
	6.6	Further Reading	315
	6.7	Exercises	317
7		Error-based Learning	323
	7.1	Big Idea	323
	7.2	Fundamentals	324
	7.2.1	Simple Linear Regression	324
	7.2.2	Measuring Error	327
	7.2.3	Error Surfaces	330
	7.3	Standard Approach: Multivariable Linear Regression with Gradient Descent	332
	7.3.1	Multivariable Linear Regression	332
	7.3.2	Gradient Descent	334
	7.3.3	Choosing Learning Rates and Initial Weights	341
	7.3.4	A Worked Example	343
	7.4	Extensions and Variations	346
	7.4.1	Interpreting Multivariable Linear Regression Models	347
	7.4.2	Setting the Learning Rate Using Weight Decay	349
	7.4.3	Handling Categorical Descriptive Features	351
	7.4.4	Handling Categorical Target Features: Logistic Regression	353
	7.4.5	Modeling Non-linear Relationships	365
	7.4.6	Multinomial Logistic Regression	373
	7.4.7	Support Vector Machines	376
	7.5	Summary	383
	7.6	Further Reading	386
	7.7	Exercises	388
8		Evaluation	397
	8.1	Big Idea	397
	8.2	Fundamentals	398
	8.3	Standard Approach: Misclassification Rate on a Hold-out Test Set	399
	8.4	Extensions and Variations	405
	8.4.1	Designing Evaluation Experiments	405

	8.4.2	Performance Measures: Categorical Targets	413
	8.4.3	Performance Measures: Prediction Scores	423
	8.4.4	Performance Measures: Multinomial Targets	440
	8.4.5	Performance Measures: Continuous Targets	442
	8.4.6	Evaluating Models after Deployment	447
	8.5	Summary	455
	8.6	Further Reading	456
	8.7	Exercises	457
9		Case Study: Customer Churn	463
	9.1	Business Understanding	463
	9.2	Data Understanding	467
	9.3	Data Preparation	471
	9.4	Modeling	477
	9.5	Evaluation	479
	9.6	Deployment	482
10		Case Study: Galaxy Classification	483
	10.1	Business Understanding	483
		10.1.1 Situational Fluency	486
	10.2	Data Understanding	488
	10.3	Data Preparation	495
	10.4	Modeling	500
		10.4.1 Baseline Models	500
		10.4.2 Feature Selection	503
		10.4.3 The 5-level Model	505
	10.5	Evaluation	508
	10.6	Deployment	509
11		The Art of Machine Learning for Predictive Data Analytics	511
	11.1	Different Perspectives on Prediction Models	513
	11.2	Choosing a Machine Learning Approach	518
		11.2.1 Matching Machine Learning Approaches to Projects	521
		11.2.2 Matching Machine Learning Approaches to Data	522
	11.3	Your Next Steps	523

A	Descriptive Statistics and Data Visualization for Machine Learning	525
	A.1 Descriptive Statistics for Continuous Features	525
	A.1.1 Central Tendency	525
	A.1.2 Variation	527
	A.2 Descriptive Statistics for Categorical Features	530
	A.3 Populations and Samples	532
	A.4 Data Visualization	534
	A.4.1 Bar Plots	534
	A.4.2 Histograms	535
	A.4.3 Box Plots	538
B	Introduction to Probability for Machine Learning	541
	B.1 Probability Basics	541
	B.2 Probability Distributions and Summing Out	546
	B.3 Some Useful Probability Rules	548
	B.4 Summary	550
C	Differentiation Techniques for Machine Learning	551
	C.1 Derivatives of Continuous Functions	552
	C.2 The Chain Rule	554
	C.3 Partial Derivatives	555
	Bibliography	557
	List of Figures	565
	List of Tables	581
	Index	589