Table of Contents

INTRODUCTION 1
About This Book 1
Foolish Assumptions 2
Icons Used in This Book 3
Beyond the Book 3
Where to Go from Here 4

PART 1: INTRODUCING HOW MACHINES LEARN 5

CHAPTER 1: Getting the Real Story about AI 7
Moving beyond the Hype 8
Dreaming of Electric Sheep 9
Understanding the history of AI and machine learning 10
Exploring what machine learning can do for AI 11
Considering the goals of machine learning 12
Defining machine learning limits based on hardware 12
Overcoming AI Fantasies 13
Discovering the fad uses of AI and machine learning 14
Considering the true uses of AI and machine learning 15
Being useful; being mundane 16
Considering the Relationship between AI and Machine Learning 17
Considering AI and Machine Learning Specifications 18
Defining the Divide between Art and Engineering 19
Predicting the Next AI Winter 20

CHAPTER 2: Learning in the Age of Big Data 23
Considering the Machine Learning Essentials 24
Defining Big Data 25
Considering the Sources of Big Data 26
Building a new data source 26
Using existing data sources 29
Locating test data sources 29
Specifying the Role of Statistics in Machine Learning 30
Understanding the Role of Algorithms 31
Defining what algorithms do 32
Considering the five main techniques 32
Defining What Training Means 34

CHAPTER 3: **Having a Glance at the Future** 37
Creating Useful Technologies for the Future 38
Considering the role of machine learning in robots. 38
Using machine learning in health care. 39
Creating smart systems for various needs 40
Using machine learning in industrial settings 40
Understanding the role of updated processors and other hardware 41
Discovering the New Work Opportunities with Machine Learning . . . 42
Working for a machine 42
Working with machines 43
Repairing machines. 44
Creating new machine learning tasks. 44
Devising new machine learning environments. 45
Avoiding the Potential Pitfalls of Future Technologies 46

PART 2: PREPARING YOUR LEARNING TOOLS 47

CHAPTER 4: **Installing a Python Distribution** 49
Using Anaconda for Machine Learning 50
Getting Anaconda 50
Defining why Anaconda is used in this book. 51
Installing Anaconda on Linux. 52
Installing Anaconda on Mac OS X 53
Installing Anaconda on Windows 54
Downloading the Datasets and Example Code. 57
Using Jupyter Notebook 57
Defining the code repository 59
Understanding the datasets used in this book 64

CHAPTER 5: **Beyond Basic Coding in Python** 67
Defining the Basics You Should Know 68
Considering Python basics 68
Working with functions. 72
Working with modules 76
Storing Data Using Sets, Lists, and Tuples 78
Creating sets. 78
Performing operations on sets 78
Using lists 79
Creating and using tuples 82
Defining Useful Iterators 83
Working with ranges 83
Iterating multiple lists using zip. 84
Working with generators using yield 84

Indexing Data Using Dictionaries 85
Creating dictionaries 85
Storing and retrieving data from dictionaries 85

CHAPTER 6: **Working with Google Colab** 87
Defining Google Colab 88
Understanding what Google Colab does 88
Considering the online coding difference 90
Using local runtime support 91
Working with Google Colab features 91
Getting a Google Account 94
Creating the account 94
Signing in 95
Working with Notebooks 96
Creating a new notebook 96
Opening existing notebooks 97
Uploading a notebook 99
Saving notebooks 100
Downloading notebooks 103
Performing Common Tasks 103
Creating code cells 104
Creating text cells 106
Creating special cells 107
Editing cells 108
Moving cells 108
Using Hardware Acceleration 108
Viewing Your Notebook 109
Displaying the table of contents 110
Getting notebook information 110
Checking code execution 110
Executing the Code 111
Sharing Your Notebook 112
Getting Help 113

PART 3: GETTING STARTED WITH THE MATH BASICS 115

CHAPTER 7: **Demystifying the Math Behind Machine Learning** 117
Working with Data 118
Learning the terminology 119
Understanding scalar and vector operations 120
Performing vector multiplication 123
Creating a matrix 123
Understanding basic operations 125

Performing matrix multiplication 126
Glancing at advanced matrix operations 128
Using vectorization effectively 129
Exploring the World of Probabilities 130
Getting an overview of probability 130
Operating on probabilities 131
Conditioning chance by Bayes' theorem 132
Describing the Use of Statistics 135

CHAPTER 8: **Descending the Gradient** 139
Acknowledging Different Kinds of Learning 140
Supervised learning 140
Unsupervised learning 141
Reinforcement learning 141
The learning process 142
Mapping an unknown function 142
Exploring cost functions 145
Descending the optimization curve 147
Optimizing with big data 148
Leveraging sampling 149
Using parallelism 150
Learning out-of-core 151

CHAPTER 9: **Validating Machine Learning** 153
Considering the Use of Example Data 154
Checking Out-of-Sample Errors 155
Understanding the concept of samples 155
Looking for the holy grail of generalization 156
Experimenting how bias and variance work 158
Keeping model complexity in mind 160
Keeping solutions balanced 162
Depicting learning curves 163
Training, Validating, and Testing 165
Considering the split 165
Resorting to cross-validation 166
Looking for alternatives in validation 167
Optimizing by Cross-Validation 169
Sources of predictive performance 169
Exploring the hyper-parameter space 170
Selecting relevant features 171
Avoiding Sample Bias and Leakage Traps 173

CHAPTER 10: **Starting with Simple Learners** 175
Discovering the Incredible Perceptron 176
Falling short of a miracle 176
Hitting the nonseparability limit 179

Growing Greedy Classification Trees 180
Predicting outcomes by splitting data 181
Pruning overgrown trees 185
Taking a Probabilistic Turn 188
Understanding Naïve Bayes 189
Estimating response with Naïve Bayes 192

PART 4: LEARNING FROM SMART AND BIG DATA 197

CHAPTER 11: Preprocessing Data 199
Gathering and Cleaning Data 200
Repairing Missing Data 201
Identifying missing data 201
Choosing the right replacement strategy 203
Transforming Distributions 205
Creating Your Own Features 207
Understanding the need to create features 207
Creating features automatically 208
Explaining the basics of SVD 210
Reorganizing data 212
Delimiting Anomalous Data 215
Using a univariate strategy 215
Resorting to Multivariate Models 217

CHAPTER 12: Leveraging Similarity 221
Measuring Similarity between Vectors 222
Understanding similarity 222
Computing distances for learning 223
Using Distances to Locate Clusters 224
Checking assumptions and expectations 226
Inspecting the gears of the K-means algorithm 227
Tuning the K-Means Algorithm 229
Experimenting with K-means reliability 230
Experimenting with how centroids converge 233
Finding Similarity by K-Nearest Neighbors 238
Understanding the k parameter 238
Experimenting with a flexible algorithm 240

CHAPTER 13: Working with Linear Models the Easy Way 243
Starting to Combine Features 244
Getting an overview of regression 244
Solving problems with a machine learning approach 247
Understanding R squared 249
Mixing Features of Different Types 251
Switching to Probabilities 255

Specifying a binary response 255
Handling multiple classes 259
Guessing the Right Features 259
Defining the outcome of features that don't work together 259
Solving overfitting by using greedy selection 260
Addressing overfitting by regularization 262
Learning One Example at a Time 264
Using gradient descent 264
Understanding how SGD is different 265

CHAPTER 14: **Hitting Complexity with Neural Networks** 271
Revising the Perceptron 272
Pushing forth with feed-forward 274
Going even deeper down the rabbit hole 276
Pulling back with backpropagation 280
Representing the Way of Learning of a Network 283
Understanding the problem with overfitting 283
Choosing a framework 285
Getting your copy of TensorFlow and Keras 286
Opening the black box 289
Introducing Deep Learning 294
Understanding some deep learning essentials 295
Explaining the magic of convolutions 296
Understanding recurrent neural networks 300

CHAPTER 15: **Going a Step Beyond Using Support Vector Machines** 307
Revisiting the Separation Problem 308
Explaining the Algorithm 309
Avoiding the pitfalls of nonseparability 311
Applying nonlinearity 312
Explaining the kernel trick by example 314
Classifying and Estimating with SVM 316

CHAPTER 16: **Resorting to Ensembles of Learners** 319
Leveraging Decision Trees 320
Growing a forest of trees 322
Understanding the importance measures 327
Working with Almost Random Guesses 330
Bagging predictors with Adaboost 330
Boosting Smart Predictors 333
Meeting again with gradient descent 334
Considering the state of the art in tabular data 335

Averaging Different Predictors 336
Blending solutions 337
Stacking diverse solutions 337

PART 5: APPLYING LEARNING TO REAL PROBLEMS 339

CHAPTER 17: Classifying Images 341
Working with a Set of Images 342
Revising the State of the Art in Computer Vision 347
Extracting Visual Features 350
Recognizing Faces Using Eigenfaces 352
Classifying Images 356

CHAPTER 18: Scoring Opinions and Sentiments 361
Introducing Natural Language Processing 362
Revising the State of the Art in NLP 363
Understanding How Machines Read 364
Defining the input data 364
Processing and enhancing text 366
Scraping textual datasets from the web 371
Handling problems with raw text 374
Using Scoring and Classification 375
Performing classification tasks 375
Analyzing reviews from e-commerce 378

CHAPTER 19: Recommending Products and Movies 383
Realizing the Revolution of E-Commerce 384
Downloading Rating Data 386
Trudging through the MovieLens dataset 386
Navigating through anonymous web data 390
Encountering the limits of rating data 392
Considering collaborative filtering 392
Catching the Limits of Behavioral Data 397
Integrating Text and Behaviors 399
Viewing the attributes 399
Obtaining statistics 400
Leveraging SVD 400
Understanding the SVD connection 400
Seeing SVD in action 401

PART 6: THE PART OF TENS 405

CHAPTER 20: Ten Ways to Improve Your Machine Learning Models 407
Studying Learning Curves 408
Using Cross-Validation Correctly 409

Choosing the Right Error or Score Metric 410
Searching for the Best Hyper-Parameters 410
Testing Multiple Models 411
Averaging Models 411
Stacking Models 412
Applying Feature Engineering 412
Selecting Features and Examples 413
Looking for More Data 414

CHAPTER 21: **Ten Guidelines for Ethical Data Usage** 415
Obtaining Permission 416
Using Sanitization Techniques 417
Avoiding Data Inference 418
Using Generalizations Correctly 418
Shunning Discriminatory Practices 419
Detecting Black Swans in Code 420
Understanding the Process 420
Considering the Consequences of an Action 421
Balancing Decision Making 421
Verifying a Data Source 422

CHAPTER 22: **Ten Machine Learning Packages to Master** 423
Gensim 423
imbalanced-learn 424
OpenCV 424
SciPy 425
SHAP 426
Statsmodels 427
Modin 427
PyTorch 428
Poetry 429
Snorkel 429

INDEX 431