

Table of Contents

About the Authors..... xiii

About the Technical Reviewerxv

Acknowledgmentsxvii

Introductionxix

Section 1: Python for Machine Learning 1

Chapter 1: Getting Started with Python 3 and Jupyter Notebook..... 3

 Python 3 Programming Language..... 4

 History of Python Programming Language..... 4

 Where Python Is Used..... 5

 Installing Python 6

 Python on Linux Distributions..... 8

 Python on macOS 8

 Python Modes 8

 Interactive Mode..... 9

 Script Mode 15

 Pip3 Utility..... 17

 Scientific Python Ecosystem..... 18

 Python Implementations and Distributions 19

 Anaconda 20

 Summary..... 21

Chapter 2: Getting Started with NumPy..... 23

 Getting Started with NumPy..... 24

 Multidimensional Ndarrays..... 26

 Indexing of Ndarrays 26

TABLE OF CONTENTS

Ndarray Properties 28

NumPy Constants..... 29

Summary..... 30

Chapter 3: Introduction to Data Visualization..... 31

NumPy Routines for Ndarray Creation 31

Matplotlib Data Visualization..... 34

Summary..... 43

Chapter 4: Introduction to Pandas..... 45

Pandas Basics..... 45

Series in Pandas 46

 Properties of Series 47

Pandas Dataframes..... 48

Visualizing the Data in Dataframes 50

Summary..... 61

Section 2: Machine Learning Approaches 63

Chapter 5: Introduction to Machine Learning with Scikit-learn 65

Learning from Data 66

 Supervised Learning..... 66

 Unsupervised Learning..... 67

Structure of a Machine Learning System..... 68

 Problem Understanding 69

 Data Collection 69

 Data Annotation and Data Preparation 70

 Data Wrangling 70

 Model Development, Training, and Evaluation 70

 Model Deployment..... 71

Scikit-Learn 72

Installing Scikit-Learn..... 72

Understanding the API..... 73

Your First Scikit-learn Experiment 75

Summary..... 77

Chapter 6: Preparing Data for Machine Learning 79

Types of Data Variables 79

 Nominal Data 80

 Ordinal Data..... 80

 Interval Data 81

 Ratio Data 81

Transformation 81

 Transforming Nominal Attributes..... 82

 Transforming Ordinal Attributes..... 84

Normalization..... 86

 Min-Max Scaling..... 87

 Standard Scaling 88

Preprocessing Text..... 89

 Preparing NLTK 89

 Five-Step NLP Pipeline 90

Preprocessing Images 94

Summary..... 97

Chapter 7: Supervised Learning Methods: Part 1 99

Linear Regression 99

 Finding the Regression Line 100

Logistic Regression..... 107

 Line vs. Curve for Expression Probability 108

 Learning the Parameters 109

 Logistic Regression Using Python 110

 Visualizing the Decision Boundary 112

TABLE OF CONTENTS

Decision Trees	114
Building a Decision Tree	114
Decision Tree in Python	116
Summary.....	120
Chapter 8: Tuning Supervised Learners.....	121
Training and Testing Processes.....	121
Measures of Performance.....	122
Confusion Matrix	122
Precision.....	124
Accuracy.....	124
F-Measure	124
Performance Metrics in Python	125
Cross Validation	128
Why Cross Validation?	129
Cross Validation in Python	129
ROC Curve.....	131
Overfitting and Regularization	134
Bias and Variance	138
Regularization.....	139
Hyperparameter Tuning.....	142
Effect of Hyperparameters	143
Summary.....	148
Chapter 9: Supervised Learning Methods: Part 2	149
Naive Bayes	149
Bayes Theorem.....	150
Conditional Probability.....	150
How Naive Bayes Works	151
Multinomial Naive Bayes	151
Naive Bayes in Python	152

Support Vector Machines	155
How SVM Works	155
Nonlinear Classification.....	157
Kernel Trick in SVM.....	158
Support Vector Machines in Python.....	159
Summary.....	165
Chapter 10: Ensemble Learning Methods	167
Bagging and Random Forest.....	168
Random Forest in Python	170
Boosting	172
Boosting in Python.....	175
Stacking Ensemble	180
Stacking in Python.....	181
Summary.....	184
Chapter 11: Unsupervised Learning Methods.....	185
Dimensionality Reduction	185
Understanding the Curse of Dimensionality	186
Principal Component Analysis	187
Principal Component Analysis in Python	189
Clustering.....	192
Clustering Using K-Means	193
K-Means in Python	194
Frequent Pattern Mining	205
Market Basket Analysis	206
Frequent Pattern Mining in Python.....	207
Summary.....	211

TABLE OF CONTENTS

Section 3: Neural Networks and Deep Learning 213

Chapter 12: Neural Network and PyTorch Basics 215

 Installing PyTorch..... 216

 PyTorch Basics 217

 Creating a Tensor 217

 Tensor Operations..... 219

 Perceptron 221

 Perceptron in Python 223

 Artificial Neural Networks 225

 Summary..... 226

Chapter 13: Feedforward Neural Networks 227

 Feedforward Neural Network..... 228

 Training Neural Networks 229

 Loss Functions..... 233

 ANN for Regression..... 234

 Activation Functions..... 239

 ReLU Activation Function 240

 Sigmoid Activation Function 240

 Tanh Activation Function 241

 Multilayer ANN 242

 NN Class in PyTorch..... 248

 Overfitting and Dropouts 251

 Classifying Handwritten Digits..... 253

 Summary..... 260

Chapter 14: Convolutional Neural Networks..... 261

 Convolution Operation..... 261

 Structure of a CNN 266

 Padding and Stride 267

 CNN in PyTorch..... 268

Image Classification Using CNN..... 271

 What Did the Model Learn? 282

Deep Networks of CNN 284

Summary..... 284

Chapter 15: Recurrent Neural Networks..... 285

 Recurrent Unit..... 286

 Types of RNN..... 288

 One to One..... 288

 One to Many 289

 Many to One 289

 Many to Many 289

 RNN in Python 290

 Long Short-Term Memory 291

 LSTM Cell..... 292

 Time Series Prediction 293

 Gated Recurrent Unit..... 304

 Summary..... 305

Chapter 16: Bringing It All Together 307

 Data Science Life Cycle 308

 CRISP-DM Process 309

 How ML Applications Are Served 312

 Learning with an Example 312

 Defining the Problem..... 313

 Data 313

 Preparing the Model 319

 Serializing for Future Predictions 325

 Hosting the Model 327

 What's Next..... 328

Index..... 329