

CONTENTS IN DETAIL

FOREWORD	xv
-----------------	-----------

ACKNOWLEDGMENTS	xvii
------------------------	-------------

INTRODUCTION	xix
---------------------	------------

Who Is This Book For?	xx
About This Book	xxi
Downloading and Installing R	xxi
Installing R Packages	xxii
Getting Help with R	xxii
Support for Python Users	xxiii
Summary	xxiii

1	
THE GEOMETRIC STRUCTURE OF DATA	1

Machine Learning Categories	2
Supervised Learning	2
Unsupervised Learning	3
Matching Algorithms and Other Machine Learning	4
Structured Data	4
The Geometry of Dummy Variables	5
The Geometry of Numerical Spreadsheets	8
The Geometry of Supervised Learning	10
Unstructured Data	17
Network Data	17
Image Data	18
Text Data	20
Summary	21

2	
THE GEOMETRIC STRUCTURE OF NETWORKS	23

Introduction to Network Science	24
The Basics of Network Theory	25
Directed Networks	26
Networks in R	26
Paths and Distance in a Network	28
Network Centrality Metrics	30
The Degree of a Vertex	30
The Closeness of a Vertex	31
The Betweenness of a Vertex	32
Eigenvector Centrality	33
PageRank Centrality	34
Katz Centrality	35
Hub and Authority	35

Measuring Centrality in an Example Social Network	36
Additional Quantities of a Network	42
The Diversity of a Vertex	42
Triadic Closure	43
The Efficiency and Eccentricity of a Vertex	44
Forman–Ricci Curvature	45
Global Network Metrics	47
The Interconnectivity of a Network	48
Spreading Processes on a Network	49
Spectral Measures of a Network	49
Network Models for Real-World Behavior	51
Erdős–Renyi Graphs	51
Scale-Free Graphs	52
Watts–Strogatz Graphs	52
Summary	54

3 NETWORK ANALYSIS 55

Using Network Data for Supervised Learning	56
Making Predictions with Social Media Network Metrics	56
Predicting Network Links in Social Media	58
Using Network Data for Unsupervised Learning	59
Applying Clustering to the Social Media Dataset	59
Community Mining in a Network	61
Comparing Networks	64
Analyzing Spread Through Networks	66
Tracking Disease Spread Between Towns	67
Tracking Disease Spread Between Windsurfers	69
Disrupting Communication and Disease Spread	72
Summary	74

4 NETWORK FILTRATION 75

Graph Filtration	76
From Graphs to Simplicial Complexes	81
Introduction to Homology	85
Examples of Betti Numbers	85
The Euler Characteristic	87
Persistent Homology	88
Comparison of Networks with Persistent Homology	90
Summary	94

5 GEOMETRY IN DATA SCIENCE 95

Introduction to Distance Metrics in Data	96
Common Distance Metrics	98
Simulating a Small Dataset	98
Using Norm-Based Distance Metrics	99
Comparing Diagrams, Shapes, and Probability Distributions	105

K-Nearest Neighbors with Metric Geometry	116
Manifold Learning	119
Using Multidimensional Scaling	120
Extending Multidimensional Scaling with Isomap	121
Capturing Local Properties with Locally Linear Embedding	122
Visualizing with t-Distributed Stochastic Neighbor Embedding	124
Fractals	125
Summary	129

**6
NEWER APPLICATIONS OF GEOMETRY
IN MACHINE LEARNING 131**

Working with Nonlinear Spaces	132
Introducing dgLARS	133
Predicting Depression with dgLARS	136
Predicting Credit Default with dgLARS	138
Applying Discrete Exterior Derivatives	140
Nonlinear Algebra in Machine Learning Algorithms	146
Comparing Choice Rankings with HodgeRank	149
Summary	153

**7
TOOLS FOR TOPOLOGICAL DATA ANALYSIS 155**

Finding Distinctive Groups with Unique Behavior	156
Validating Measurement Tools	159
Using the Mapper Algorithm for Subgroup Mining	161
Stepping Through the Mapper Algorithm	162
Using TDAmapper to Find Cluster Structures in Data	163
Summary	166

**8
HOMOTOPY ALGORITHMS 167**

Introducing Homotopy	167
Introducing Homotopy-Based Regression	169
Comparing Results on a Sample Dataset	174
Summary	177

**9
FINAL PROJECT: ANALYZING TEXT DATA 179**

Building a Natural Language Processing Pipeline	180
The Project: Analyzing Language in Poetry	181
Tokenizing Text Data	183
Tagging Parts of Speech	183
Normalizing Vectors	184
Analyzing the Poem Dataset in R	184
Using Topology-Based NLP Tools	188
Summary	191

10	
MULTICORE AND QUANTUM COMPUTING	193
Multicore Approaches to Topological Data Analysis	194
Quantum Computing Approaches	195
Using the Qubit-Based Model	196
Using the Qumodes-Based Model	197
Using Quantum Network Algorithms	197
Speeding Up Algorithms with Quantum Computing	199
Using Image Classifiers on Quantum Computers	200
Summary	205
REFERENCES	207
INDEX	221