Brief Contents

1 Introduction 1

I Foundations 29

2 Probability: Univariate Models 31

3 Probability: Multivariate Models 75

4 Statistics 103

5 Decision Theory 163

6 Information Theory 199

7 Linear Algebra 221

8 Optimization 269

II Linear Models 315

9 Linear Discriminant Analysis 317

10 Logistic Regression 333

11 Linear Regression 365

12 Generalized Linear Models * 409

III Deep Neural Networks 417

13 Neural Networks for Structured Data 419

14 Neural Networks for Images 461

15 Neural Networks for Sequences 497

IV Nonparametric Models 539

16 Exemplar-based Methods 541

17 Kernel Methods * 561

18 Trees, Forests, Bagging, and Boosting 597

V Beyond Supervised Learning 619
19 Learning with Fewer Labeled Examples 621
20 Dimensionality Reduction 651
21 Clustering 709
22 Recommender Systems 735
23 Graph Embeddings * 747
A Notation 767