Cambridge University Press & Assessment
978-1-316-51282-1 — Machine Learning for Engineers
Osvaldo Simeone
Table of Contents
More Information

Contents

Preface *page* xi
Acknowledgements xvii
Notation xviii
Acronyms xxi

Part I Introduction and Background

1 When and How to Use Machine Learning 3
1.1 Overview 3
1.2 What is Machine Learning For? 3
1.3 Why Study Machine Learning Now? 4
1.4 What is Machine Learning? 5
1.5 Taxonomy of Machine Learning Methods 9
1.6 When to Use Machine Learning? 10
1.7 Summary 11
1.8 Recommended Resources 12
Bibliography 13

2 Background 14
2.1 Overview 14
2.2 Random Variables 14
2.3 Expectation 18
2.4 Variance 20
2.5 Vectors 21
2.6 Matrices 25
2.7 Random Vectors 29
2.8 Marginal Distributions 35
2.9 Conditional Distributions 36
2.10 Independence of Random Variables and Chain Rule of Probability 40
2.11 Bayes' Theorem 42
2.12 Law of Iterated Expectations 45
2.13 Summary 46
2.14 Recommended Resources 46
Problems 46
Bibliography 49

Part II Fundamental Concepts and Algorithms

3 Inference, or Model-Driven Prediction 53
3.1 Overview 53
3.2 Defining Inference 54

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE
Cambridge University Press & Assessment
978-1-316-51282-1 — Machine Learning for Engineers
Osvaldo Simeone
Table of Contents
More Information

3.3 Optimal Hard Prediction 56
3.4 Optimal Prediction for Jointly Gaussian Random Vectors 62
3.5 KL Divergence and Cross Entropy 65
3.6 Optimal Soft Prediction 70
3.7 Mutual Information 78
3.8 Log-Loss As a "Universal" Loss Function 82
3.9 Free Energy 82
3.10 Summary 85
3.11 Recommended Resources 86
Problems 86
Appendices 89
Bibliography 93

4 Supervised Learning: Getting Started 94
4.1 Overview 94
4.2 Defining Supervised Learning 95
4.3 Training Hard Predictors 100
4.4 Inductive Bias Selection and Validation 106
4.5 Bias and Estimation Error 113
4.6 Beyond the Bias vs. Estimation Error Trade-Off 116
4.7 Regularization 118
4.8 Testing 120
4.9 Training Soft Predictors 121
4.10 MAP Learning and Regularization 124
4.11 Kernel Methods 126
4.12 Summary 130
4.13 Recommended Resources 131
Problems 131
Appendices 134
Bibliography 135

5 Optimization for Machine Learning 136
5.1 Overview 136
5.2 Optimization for Training 136
5.3 Solving an Optimization Problem 137
5.4 First-Order Necessary Optimality Condition for Single-Variable Functions 139
5.5 Second-Order Optimality Conditions for Single-Variable Functions 142
5.6 Optimality Conditions for Convex Single-Variable Functions 144
5.7 Optimality Conditions for Multi-variable Functions 146
5.8 Gradient Descent 153
5.9 Convergence of Gradient Descent 157
5.10 Second-Order Methods 164
5.11 Stochastic Gradient Descent 165
5.12 Convergence of Stochastic Gradient Descent 167
5.13 Minimizing Population Loss vs. Minimizing Training Loss 171
5.14 Symbolic and Numerical Differentiation 172
5.15 Automatic Differentiation 173

© in this web service Cambridge University Press & Assessment www.cambridge.org

5.16 Summary 182
5.17 Recommended Resources 183
Problems 184
Appendices 186
Bibliography 193

6 Supervised Learning: Beyond Least Squares 194
6.1 Overview 194
6.2 Discriminative Linear Models for Binary Classification: Hard Predictors 194
6.3 Discriminative Linear Models for Binary Classification: Soft Predictors and Logistic Regression 203
6.4 Discriminative Non-linear Models for Binary Classification: Multi-layer Neural Networks 209
6.5 Generative Models for Binary Classification 224
6.6 Discriminative Linear Models for Multi-class Classification: Softmax Regression 231
6.7 Discriminative Non-linear Models for Multi-class Classification: Multi-layer Neural Networks 235
6.8 Generative Models for Multi-class Classification 237
6.9 Mixture Models 238
6.10 Beyond Feedforward Multi-layer Neural Networks 240
6.11 *K*-Nearest Neighbors Classification 243
6.12 Applications to Regression 243
6.13 Summary 244
6.14 Recommended Resources 245
Problems 245
Appendices 248
Bibliography 248

7 Unsupervised Learning 250
7.1 Overview 250
7.2 Unsupervised Learning Tasks 251
7.3 Density Estimation 251
7.4 Latent-Variable Models 258
7.5 Autoencoders: x → z → x 260
7.6 Discriminative Models: x → z 268
7.7 Undirected Generative Models: x ↔ z 271
7.8 Directed Generative Models: z → x 279
7.9 Training Directed Generative Models: *K*-Means Clustering 280
7.10 Training Directed Generative Models: Expectation Maximization 285
7.11 Summary 298
7.12 Recommended Resources 299
Problems 299
Appendices 301
Bibliography 306

© in this web service Cambridge University Press & Assessment

Cambridge University Press & Assessment
978-1-316-51282-1 — Machine Learning for Engineers
Osvaldo Simeone
Table of Contents
More Information

Part III Advanced Tools and Algorithms

8 Statistical Learning Theory 311
8.1 Overview 311
8.2 Benchmarks and Decomposition of the Optimality Error 312
8.3 Probably Approximately Correct Learning 315
8.4 PAC Learning for Finite Model Classes 319
8.5 PAC Learning for General Model Classes: VC Dimension 322
8.6 PAC Bayes and Information-Theoretic Bounds 326
8.7 Summary 329
8.8 Recommended Resources 330
Problems 330
Bibliography 331

9 Exponential Family of Distributions 332
9.1 Overview 332
9.2 Definitions and Examples 333
9.3 Exponential-Family Distributions As Maximum-Entropy Models 341
9.4 Gradient of the Log-Loss 342
9.5 ML Learning 344
9.6 Information-Theoretic Metrics 348
9.7 Fisher Information Matrix 349
9.8 Generalized Linear Models 354
9.9 Summary 356
9.10 Recommended Resources 357
Problems 357
Appendices 358
Bibliography 367

10 Variational Inference and Variational Expectation Maximization 368
10.1 Overview 368
10.2 Variational Inference, Amortized Variational Inference, and Variational EM 369
10.3 Exact Bayesian Inference 371
10.4 Laplace Approximation 377
10.5 Introducing Variational Inference 378
10.6 Mean-Field Variational Inference 380
10.7 Introducing Parametric Variational Inference 384
10.8 Black-Box Variational Inference 385
10.9 Reparametrization-Based Variational Inference 393
10.10 Combining Factorization and Parametrization for Variational Inference 401
10.11 Particle-Based Variational Inference and Stein Variational Gradient Descent 402
10.12 Amortized Variational Inference 404
10.13 Variational Expectation Maximization 407
10.14 Summary 411
10.15 Recommended Resources 412
Problems 413
Appendices 414
Bibliography 418

© in this web service Cambridge University Press & Assessment

CAMBRIDGE

Cambridge University Press & Assessment
978-1-316-51282-1 — Machine Learning for Engineers
Osvaldo Simeone
Table of Contents
More Information

11 Information-Theoretic Inference and Learning 419
11.1 Overview 419
11.2 I-Projection and M-Projection 420
11.3 Generalized Variational Inference and Generalized Variational Expectation Maximization 423
11.4 Maximum-Entropy Learning 427
11.5 InfoMax 428
11.6 Information Bottleneck 430
11.7 Rate-Distortion Encoding 431
11.8 Two-Sample Estimation of Information-Theoretic Metrics 433
11.9 Beyond the KL Divergence: f-Divergences and Two-Sample Estimators 437
11.10 Generative Adversarial Networks 440
11.11 Distributionally Robust Learning 444
11.12 Summary 445
11.13 Recommended Resources 445
Problems 445
Appendices 446
Bibliography 449

12 Bayesian Learning 451
12.1 Overview 451
12.2 Frequentist Learning and Calibration 452
12.3 Basics of Bayesian Learning 455
12.4 Why Bayesian Learning? 460
12.5 Exact Bayesian Learning 464
12.6 Laplace Approximation 467
12.7 MC Sampling-Based Bayesian Learning 469
12.8 Variational Bayesian Learning 476
12.9 Model Selection without Validation: Empirical Bayes 482
12.10 Bayesian Non-parametric Learning 485
12.11 Bayesian Learning with Local Latent Variables 490
12.12 Generalized Bayesian Learning 491
12.13 Summary 493
12.14 Recommended Resources 494
Problems 494
Appendices 495
Bibliography 498

Part IV Beyond Centralized Single-Task Learning

13 Transfer Learning, Multi-task Learning, Continual Learning, and Meta-learning 503
13.1 Overview 503
13.2 Transfer Learning 504
13.3 Multi-task Learning 512
13.4 Continual Learning 515
13.5 Meta-learning 522
13.6 Bayesian Perspective on Transfer, Multi-task, Continual, and Meta-learning 534
13.7 Summary 538

© in this web service Cambridge University Press & Assessment www.cambridge.org

CAMBRIDGE

Cambridge University Press & Assessment
978-1-316-51282-1 — Machine Learning for Engineers
Osvaldo Simeone
Table of Contents
More Information

13.8 Recommended Resources 539
Appendices 539
Bibliography 540

14 Federated Learning 541
14.1 Overview 541
14.2 Frequentist Federated Learning 542
14.3 Private Federated Learning 550
14.4 Bayesian Federated Learning 554
14.5 Summary 561
14.6 Recommended Resources 562
Bibliography 562

Part V Epilogue

15 Beyond This Book 565
15.1 Overview 565
15.2 Probabilistic Graphical Models 565
15.3 Adversarial Attacks 568
15.4 Causality 570
15.5 Quantum Machine Learning 571
15.6 Machine Unlearning 573
15.7 General AI? 573
15.8 Recommended Resources 574
Bibliography 575

Index 576

© in this web service Cambridge University Press & Assessment