Contents

Preface vii

Reading Guide xi

1 Introduction 1
1.1 Statistical Learning 1
1.2 Support Vector Machines: An Overview 7
1.3 History of SVMs and Geometrical Interpretation 13
1.4 Alternatives to SVMs 19

2 Loss Functions and Their Risks 21
2.1 Loss Functions: Definition and Examples 21
2.2 Basic Properties of Loss Functions and Their Risks 28
2.3 Margin-Based Losses for Classification Problems 34
2.4 Distance-Based Losses for Regression Problems 38
2.5 Further Reading and Advanced Topics 45
2.6 Summary 46
2.7 Exercises 46

3 Surrogate Loss Functions (*) 49
3.1 Inner Risks and the Calibration Function 51
3.2 Asymptotic Theory of Surrogate Losses 60
3.3 Inequalities between Excess Risks 63
3.4 Surrogates for Unweighted Binary Classification 71
3.5 Surrogates for Weighted Binary Classification 76
3.6 Template Loss Functions 80
3.7 Surrogate Losses for Regression Problems 81
3.8 Surrogate Losses for the Density Level Problem 93
3.9 Self-Calibrated Loss Functions 97
3.10 Further Reading and Advanced Topics 105
3.11 Summary 106
3.12 Exercises 107

4 Kernels and Reproducing Kernel Hilbert Spaces 111
4.1 Basic Properties and Examples of Kernels 112
4.2 The Reproducing Kernel Hilbert Space of a Kernel 119
4.3 Properties of RKHSs 124
4.4 Gaussian Kernels and Their RKHSs 132
4.5 Mercer's Theorem (*) 149
4.6 Large Reproducing Kernel Hilbert Spaces 151
4.7 Further Reading and Advanced Topics 159
4.8 Summary 161
4.9 Exercises 162

5 Infinite-Sample Versions of Support Vector Machines 165
5.1 Existence and Uniqueness of SVM Solutions 166
5.2 A General Representer Theorem 169
5.3 Stability of Infinite-Sample SVMs 173
5.4 Behavior for Small Regularization Parameters 178
5.5 Approximation Error of RKHSs 187
5.6 Further Reading and Advanced Topics 197
5.7 Summary 200
5.8 Exercises 200

6 Basic Statistical Analysis of SVMs 203
6.1 Notions of Statistical Learning 204
6.2 Basic Concentration Inequalities 210
6.3 Statistical Analysis of Empirical Risk Minimization 218
6.4 Basic Oracle Inequalities for SVMs 223
6.5 Data-Dependent Parameter Selection for SVMs 229
6.6 Further Reading and Advanced Topics 234
6.7 Summary 235
6.8 Exercises 236

7 Advanced Statistical Analysis of SVMs (*) 239
7.1 Why Do We Need a Refined Analysis? 240
7.2 A Refined Oracle Inequality for ERM 242
7.3 Some Advanced Machinery 246
7.4 Refined Oracle Inequalities for SVMs 258
7.5 Some Bounds on Average Entropy Numbers 270
7.6 Further Reading and Advanced Topics 279
7.7 Summary 282
7.8 Exercises 283

8 Support Vector Machines for Classification 287
8.1 Basic Oracle Inequalities for Classifying with SVMs 288
8.2 Classifying with SVMs Using Gaussian Kernels 290
8.3 Advanced Concentration Results for SVMs (*) 307
8.4 Sparseness of SVMs Using the Hinge Loss 310
8.5 Classifying with other Margin-Based Losses (*) 314
8.6 Further Reading and Advanced Topics 326
8.7 Summary 329
8.8 Exercises 330

9 Support Vector Machines for Regression 333
9.1 Introduction 333
9.2 Consistency 335
9.3 SVMs for Quantile Regression 340
9.4 Numerical Results for Quantile Regression 344
9.5 Median Regression with the eps-Insensitive Loss (*) 348
9.6 Further Reading and Advanced Topics 352
9.7 Summary 353
9.8 Exercises 353

10 Robustness 355
10.1 Motivation 356
10.2 Approaches to Robust Statistics 362
10.3 Robustness of SVMs for Classification 368
10.4 Robustness of SVMs for Regression (*) 379
10.5 Robust Learning from Bites (*) 391
10.6 Further Reading and Advanced Topics 403
10.7 Summary 408
10.8 Exercises 409

11 Computational Aspects 411
11.1 SVMs, Convex Programs, and Duality 412
11.2 Implementation Techniques 420
11.3 Determination of Hyperparameters 443
11.4 Software Packages 448
11.5 Further Reading and Advanced Topics 450
11.6 Summary 452
11.7 Exercises 453

12 Data Mining 455
12.1 Introduction 456
12.2 CRISP-DM Strategy 457
12.3 Role of SVMs in Data Mining 467
12.4 Software Tools for Data Mining 467
12.5 Further Reading and Advanced Topics 468

12.6 Summary . . . 469
12.7 Exercises . . . 469

Appendix . . . 471
A.1 Basic Equations, Inequalities, and Functions . . . 471
A.2 Topology . . . 475
A.3 Measure and Integration Theory . . . 479
A.3.1 Some Basic Facts . . . 480
A.3.2 Measures on Topological Spaces . . . 486
A.3.3 Aumann's Measurable Selection Principle . . . 487
A.4 Probability Theory and Statistics . . . 489
A.4.1 Some Basic Facts . . . 489
A.4.2 Some Limit Theorems . . . 492
A.4.3 The Weak* Topology and Its Metrization . . . 494
A.5 Functional Analysis . . . 497
A.5.1 Essentials on Banach Spaces and Linear Operators . . . 497
A.5.2 Hilbert Spaces . . . 501
A.5.3 The Calculus in Normed Spaces . . . 507
A.5.4 Banach Space Valued Integration . . . 508
A.5.5 Some Important Banach Spaces . . . 511
A.5.6 Entropy Numbers . . . 516
A.6 Convex Analysis . . . 519
A.6.1 Basic Properties of Convex Functions . . . 520
A.6.2 Subdifferential Calculus for Convex Functions . . . 523
A.6.3 Some Further Notions of Convexity . . . 526
A.6.4 The Fenchel-Legendre Bi-conjugate . . . 529
A.6.5 Convex Programs and Lagrange Multipliers . . . 530
A.7 Complex Analysis . . . 534
A.8 Inequalities Involving Rademacher Sequences . . . 534
A.9 Talagrand's Inequality . . . 538

References . . . 553

Notation and Symbols . . . 579

Abbreviations . . . 583

Author Index . . . 585

Subject Index . . . 591